

# Landslide susceptibility prediction mapping with advanced ensemble models: Son La province, Vietnam

QUYNH Bui (✉ [quynhbd@nuce.edu.vn](mailto:quynhbd@nuce.edu.vn))

Ha Noi University of Civil Engineering: National University of Civil Engineering <https://orcid.org/0000-0003-1489-8918>

**Hang Ha**

Ha Noi University of Civil Engineering: National University of Civil Engineering

**Dong Khuc Thanh**

Ha Noi University of Civil Engineering: National University of Civil Engineering

**Dinh Quoc Nguyen**

King Fahd University of Petroleum & Minerals College of Petroleum Engineering & Geosciences

**Jason von Meding**

University of Florida

**Lam Phuong Nguyen**

Ha Noi University of Civil Engineering: National University of Civil Engineering

**Chinh Luu**

Ha Noi University of Civil Engineering: National University of Civil Engineering

---

## Research Article

**Keywords:** Landslide susceptibility, hybrid machine learning models, landslide risk management, Son La province, Vietnam

**Posted Date:** July 21st, 2022

**DOI:** <https://doi.org/10.21203/rs.3.rs-1650275/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

Landslides are a serious geohazard in many mountainous areas of Vietnam during the rainy season. They directly threaten human lives and properties every year. Landslide susceptibility maps are useful tools for risk mitigation, land-use planning, and early warning systems for local areas. It is necessary to update these maps continuously because of the complexity of landslide events. This fact requires further extending the approach techniques with practical implications. Therefore, this study aimed to develop landslide susceptibility prediction maps based on advanced Machine Learning (ML) techniques. Five state-of-art hybrid ML models were developed: Bagging – MLP, Dagging – MLP, Decorate – MLP, Rotation Forest – MLP, Random SubSpace – MLP with Multi-Layer Perceptron (MLP) as a base classifier. Sixteen causative factors were collected to build landslide susceptibility maps based on the relationship between historical landslide locations and specific local geo-environmental conditions. The model performance was verified using various statistical indexes. Based on the Area Under ROC curve (AUC) analysis results of the testing dataset, the Rotation Forest – MLP model has the greatest predictive accuracy of AUC = 0.818. It is followed by the Decorate – MLP and Bagging – MLP (AUC=0.804), the Random SubSpace – MLP model (AUC=0.796), the Dagging – MLP (AUC=0.789), and the single MLP (AUC=0.698). The results of this study can be applied effectively to other mountainous regions to mitigate the risk of landslides.

## 1. Introduction

Landslides are among the most dangerous natural hazards in mountainous areas, generating more deaths and more economic damage than others (Wu et al., 1996; Zhu & Huang, 2006). Landslides can wreak havoc on buildings, roads, and road infrastructure, damage agriculture, and threaten lives (Guzzetti et al., 2004; Harp & Jibson, 1996). The factors of precipitation, topographic, geological, geomorphic, and geoenvironmental conditions are closely related to landslide occurrences (Hang et al., 2021). Landslides may also be influenced by human activities such as deforestation, land-use change, transportation development, and plant degradation (Skilodimou et al., 2018).

In Vietnam, landslides often occur in mountain areas during the rainy season (Nguyen et al., 2020). Landslide susceptibility mapping can provide useful information for government agencies to implement land use and mitigation measure planning (Pourghasemi et al., 2020). Moreover, landslide spatial forecasting can significantly help managers determine the trend for landslide occurrences in a specific region (Aleotti & Chowdhury, 1999). The susceptible landslide zoning is essential in identifying the areas where proper planning and infrastructure design could be developed and conducted considering landslide risks (Hang et al., 2021). In addition, the information on landslide-prone sites needs to be updated continuously because of their complex geological processes (Chen & Li, 2020; Van Westen et al., 2008).

Various methodologies and procedures have been utilized to map landslide susceptibility in hilly areas. Some statistical approaches used to construct landslide susceptibility maps, such as weights-of-evidence (Armaş, 2012; Mahdadi et al., 2018; Pradhan et al., 2010), frequency ratio (Juliev et al., 2019), multicriteria decision analysis (Feizizadeh & Blaschke, 2013), or a data-driven approach (Riaz et al., 2018). Recently,

Machine Learning (ML) techniques have been explored and applied as a potential approach in landslide prediction studies based on geo-environmental factors (Bui et al., 2019). For example, landslide susceptible areas were forecasted based on logistic regression (Hemasinghe et al., 2018; Hong et al., 2016), artificial neural networks, support vector machine (Huang & Zhao, 2018; Lee et al., 2017), random forest (Kim et al., 2018; Taalab et al., 2018), Naïve Bayes tree (Chen et al., 2018; Hu et al., 2021). Many advanced hybrid ML models have been constructed to improve the accuracy of predictive models (Achour & Pourghasemi, 2020). Along with data availability, the selection of efficient predictive models is very important to receive reliable landslide susceptibility maps (Achour & Pourghasemi, 2020; Pourghasemi et al., 2020).

Multi-Layer Perceptron (MLP) as a base classifier method is one of the most popular artificial neural network techniques since it can bring a higher predictive accuracy for predictive models (Adnan et al., 2020; Hong et al., 2020; Li et al., 2019; Meghanadh et al., 2022; Zare et al., 2013). MLP can be applied as a single model in creating a landslide susceptibility map (Adnan et al., 2020; Meghanadh et al., 2022; Zare et al., 2013). MLP can be also combined with other ML models to forecast the spatial distribution of landslides, such as the particle-swarm-optimized model (Li et al., 2019) and the stochastic gradient descent model (Hong et al., 2020). These studies concluded that the ensemble ML models with MLP have higher predictive accuracy than other ML models (Achour & Pourghasemi, 2020; Wang et al., 2020).

During the rainy season, large landslides frequently threaten human lives and properties in Son La province, which is located in Vietnam's northwest mountainous region (Ahlheim et al., 2008). However, it still lacks studies on evaluating landslide susceptibility for the area. This study aimed to create the landslide susceptibility map for the new case study of Son La province, using advanced ensemble ML models. We developed five hybrid models which combine complex ML models with the MLP model as a base classifier model. They are Bagging – MLP, Dagging – MLP, Decorate – MLP, Rotation Forest – MLP, and Random SubSpace – MLP. The landslide-prone areas are identified based on the research area's specific topographical, hydrological, geological, and environmental conditions. The resulting susceptibility map can supply useful information for local authorities in strengthening landslide disaster risk mitigation strategies at the regional scale.

## 2. Study Area

Son La is a mountainous province in Vietnam's northwest region. It has 250km of borderline with the Lao People's Democratic Republic. The province covers an area of 14,174 km<sup>2</sup> and is situated within 20°39' and 22°02' North latitude and between 103°11' and 105°02' East longitude (Fig. 1). The population of Son La was 1,252,650 people in 2019, of which 13.85% are in urban areas and 86.15% in rural areas. The Son La province has 12 ethnic groups, of which 53.2% Thai group, 17.6% Kinh group, 14.6% H'Mong group, 7.6% Muong group, and 7.2% other ethnic minorities groups. This province has many mountains, hills, and rivers and is surrounded by primitive forests. Especially the terrain of Son La province is strongly divided by the Da River, Ma River, and high mountain ranges (Thach & Canh, 2011).

In the study area, landslides are influenced by extreme weather, prolonged heavy rainfall, and human activities. (Van Hoang et al., 2019). The rainfall averages 300–700 mm per month during the rainy season, with maximum daily rainfall exceeding 100 mm (Hoang & Tien Bui, 2018). In addition, the other factors of topography, geology, fault density, land use, weathering crush, type of soil, and soil thickness also affect landslide susceptibility in this province (Thach & Canh, 2011). Landslide hazard has caused serious damage to humans, houses, road networks, agriculture, road infrastructure, and drainage system in the province (IFRC, 2021). For example, there were 10 killed people, 4 injured, 258 damaged houses by landslides in the Muong La district in 2017 (Dung et al., 2021).

### **3. Methodology**

#### **3.1 Methodology flowchart**

The methodological framework of this study is shown in Fig. 2. The framework consists of six main steps: (1) Data collection and preparation, (2) Checking the multicollinearity of the landslide-related variables, (3) Landslide inventory mapping, (4) Landslide modeling process using a hybrid ML ensemble framework, (5) Landslide susceptibility mapping, and (6) Model comparison and validation. First, the topographical, geological, hydrological, environmental data and historical landslide locations were collected and prepared for modeling samples. Second, these data were checked for multicollinearity to avoid computational instability in model assessment. Third, the landslide inventory map was created. Using the *Sample* tool in ArcGIS Pro, the sample data were randomly split into training (70%) and validating (30%) datasets. Fourth, ensemble ML techniques were developed for the modeling, including Bagging – MLP, Dagging – MLP, Decorate – MLP, Rotation Forest – MLP, and Random SubSpace – MLP. Fifth, the above hybrid ML models were used to create landslide susceptibility maps. Finally, the susceptibility models were compared and verified using the cross-validation approach.

#### **3.2 Landslide inventory map**

It is needed to rely on detailed information on previous landslide events in studies on the formation mechanism of landslides, landslide susceptibility mapping, and developing landslide risk mitigation strategies (Mirus et al., 2020). Thus, landslide inventory mapping is often the first and the most crucial step in data preparation for modeling (Bui et al., 2019). In this study, 1,689 landslide locations were collected from many different sources, in which 1,225 landslide positions were explored from the website of the Institute of Geosciences and Minerals of Vietnam (available at <http://canhbaotruotlo.vn/phanvungcactinh.html>), and 464 landslide locations were added through field survey combined with the interpretation of Google Earth images.

There are three main causes leading to landslide formation in Son La province, including construction structures on steep slopes, thick weathering crust, and high groundwater levels (Thach & Canh, 2011). In addition, heavy prolonged rainfall events are considered the primary cause of landslides in the study area

(Hoang & Tien Bui, 2018). The statistics have recorded many landslide developments in Quynh Nhai, Sop Cop, Song Ma, Moc Chau, Van Ho, and Thuan Chau districts (Hoang & Tien Bui, 2018; IFRC, 2021; Thach & Canh, 2011). The landslide curves stretched from 10 meters to 100 meters, and the landslide regions were bigger than 30m<sup>2</sup> (Thach & Canh, 2011). In this study, the landslide inventory data were randomly split into two samples: 70% training dataset (1,183 landslide sites) and 30% validating dataset (506 landslide sites) using the *Sample* tool of ArcGIS Pro software.

## 3.3 Landslide causative factors

Modeling work begins with the identification of landslide causative factors. The factors are referred to in previous studies and are based on the available data in the research area (Kavzoglu et al., 2019). Precipitation, topography, hydrology, geology, geomorphology, and geoenvironment are factors that significantly impact landslide formation in mountain areas (Bui et al., 2019; Meghanadh et al., 2022). In this study, 16 landslide causative factors were taken for the modeling in the following subsections (Fig. 3 and Table 1).

### 3.3.1 Elevation

In investigations of landslide susceptibility mapping, elevation is a crucial factor in studies on landslide susceptibility mapping (Goetz et al., 2015; Tehrany & Kumar, 2018). The higher elevation zones correspond to the higher landslide frequencies (Myronidis et al., 2016). The elevation map was constructed using the Digital Elevation Model (DEM). The DEM was created using an ALOS picture with a spatial resolution of 30 m obtained from <https://www.eorc.jaxa.jp/ALOS/en/aw3d30/> in March 2021. The research area's elevation ranges from 70 to 2884 meters and is separated into nine groups (Fig. 3 (a)).

### 3.3.2 Slope

The slope is a critical factor in landslide susceptibility evaluations since it can control the landslide creation and movement in tropical mountainous areas (Dai et al., 2002; Guns & Vanacker, 2013). The slope angle map was created using ArcGIS Pro software and a DEM with a spatial resolution of 30 m. The slope angle map is separated into six classes, ranging from 0<sup>0</sup> to 78.3<sup>0</sup>. (Fig. 3 (b)).

### 3.3.3 Aspect, curvature

Other topographical characteristics, such as aspect and curvature, are often used as primary input data in landslide prediction (Arabameri et al., 2020; Hang et al., 2021; Kavzoglu et al., 2019). These variables were estimated using ArcGIS Pro software and a DEM with a 30 m spatial resolution. The aspect map (Fig. 3 (c)) was divided into nine classes, and the curvature map (Fig. 3 (d)) was categorized into five levels.

### 3.3.4 Elevation difference

Elevation difference represents the altitude difference of all points on the Earth's surface (Corsini et al., 2005). This factor also reflects the exiguous separation of elevations where water redistribution is very significant to landslide formation (Van Westen et al., 2003). The elevation difference factor on the 1:50,000 topographical map was derived using relative altitude (meters) in each square grid ( $1 \text{ km}^2$ ). It was categorized into nine classes (Fig. 3 (e)).

### 3.3.5. TWI

The Topographic Wetness Index (TWI) measures the impact of topography on the location and amount of saturated runoff source zones (Pourghasemi et al., 2012). TWI was built from a 30 m DEM and was divided into seven categories in ArcGIS Pro software Fig. 3 (f). TWI is calculated as follows:

$$TWI = \ln \left( \frac{A_S}{\tan \beta} \right) \quad (1)$$

where  $A_S$  denotes the specific basin area ( $\text{m}^2/\text{m}$ ) and  $\beta$  denotes a sloped angle in degrees.

### 3.3.6 NDVI

The Normalized Difference Vegetation Index (NDVI) measures the development of vegetation on the Earth's surface (Jaafari et al., 2014). The NDVI explains the link between vegetation density and the occurrence and distribution of landslides (Chen et al., 2019). The NDVI is expressed as below:

$$NDVI = \frac{NIR - Red}{NIR + Red} \quad (2)$$

Where  $NIR$  denotes the infrared reflectance of the electromagnetic spectrum, and  $Red$  denotes the red reflectance of the electromagnetic spectrum.

The NDVI value in this research ranged from – 0.643 to 0.694 and was separated into five groups (Fig. 3(g)).

### 3.3.7 Rainfall

Heavy prolonged rainfall is considered the primary cause of landslides in mountain areas (Singh et al., 2021). It might trigger unexpected landslides depending on the topographical and geological characteristics of the ground/rock mass (Dung et al., 2021). In this study, the rainfall data were derived from 25 rain gauge stations in Son La province and neighboring provinces from 2010 to 2021. The

rainfall map was created by interpolating the rainfall distribution based on DEM using the geostatistical Kriging technique (Fig. 3 (h)).

### 3.3.8 Drainage density

Drainage density describes the drainage availability in the short term in response to changes in environmental conditions (Mezighi et al., 2011). Drainage density has a direct association with the landslide formation in mountain places (Arabameri et al., 2020). The drainage density is calculated by dividing the total drainage length (km) in each square grid by the number of square grids (Fig. 3 (i)).

### 3.3.9 Road density and distance to road

The road network is often associated with an increase in landslide events (Skilodimou et al., 2018). Meanwhile, road density is often used to measure the effect of development on landslide formation and distribution (Simon et al., 2015). Distance from road represents a negative relation with the landslide events (Akgün & Bulut, 2007). The shorter road distances are, the higher the landslide occurrences are (Skilodimou et al., 2018). The road density map was grouped into five levels (Fig. 3 (l)), and the distance to the road map was classified into six classes (Fig. 3 (k)).

### 3.3.10 Distance to rivers

Many previous studies have proved that landslide has a close relationship with distance to rivers (Arabameri et al., 2020; Bui et al., 2019). Landslides often happen along the sides of the valleys where the groundwater flows toward rivers and streams (Raja et al., 2017). In this study, the distance to rivers map was calculated in ArcGIS Pro software and consisted of seven classes (Fig. 3 (n)).

### 3.3.11 Hydrogeology

Hydrology and hydrogeology significantly affect landslide formation in the hilly areas (Kayastha et al., 2012). Many studies have looked at the importance and complexity of hydrogeology in landslide susceptibility evaluations (Frodella et al., 2021; Sujatha, 2021). The hydrogeology map of Son La province was obtained on a scale of 50,000 in 2020 from the Vietnamese Ministry of Natural Resources and Environment. The map is covered by four hydrogeological units (Fig. 3 (m)).

### 3.3.12 Geology and geomorphology

Although most of the geological and geomorphological factors change over relatively long periods, their characteristics have a substantial influence on the evolution of erosional and landslide processes in mountainous areas (Bui et al., 2019; Pisano et al., 2017). In 2020, the Vietnamese Ministry of Natural Resources and Environment released geological and geomorphological maps of Son La province on a

scale of 50,000. The geological map is covered by nine geological groups (Fig. 3 **(s)**). The detail of the nine geological groups is shown in Table 1. The geomorphological map is covered by fifteen geomorphological types (Fig. 3 **(p)**).

### 3.3.13 Land cover

Anthropogenic activities can change land cover due to the transformation of the natural landscape (Promper et al., 2014). Therefore this factor is often considered an important triggering factor for landslide occurrences in mountain regions (Van Westen et al., 2003). Land cover can affect the landslide frequency and distribution quickly (Hong et al., 2007). In this study, the land cover map was developed by ESRI in 2020 using Deep Learning models and satellite images (downloaded at <https://livingatlas.arcgis.com/landcover/>). The land cover map of the study area consists of eight groups, including bare ground (0.02%), built area (2.57%), crops (6.59%), flooded vegetation (0.01%), grass (0.53%), scrub/shrub (28.86%), trees (59.55%), and water bodies (1.87%) (Fig. 3 **(m)**).

Table 1  
Complex and formation types of geological in this study

No	Group	Name	Complex and Formation type
1	Group 1	Quaternary	Upper Holocene; Lower Middle Holocene; Middle-Upper Pleistocene; undifferentiated Quaternary;  HaiHung Formation; VinhPhuc Formation; HangMon Formation; NamPo Formation;  ThaiBinh Formation; HaNoi Formation; LeChi Formation
2	Group 2	PALEOGEN	ChoDon Complex; Ye Yen Sun Complex; Pu Sam Cap Complex; PuTra Formation;  NamXe_TamDuong Complex; NamBay Formation; CocPia Complex;
3	Group 3	JURA-CRETA-CRETACEOUS	YenChau Formation; NgoiThia Subcomplex; TuLe Subcapmplex; Phu Sa Phin Complex;  NamChien Complex; Middle Subformation; Lower Subformation; NgoiThia volcanic Subcomplex;  TuLe volcanic Subcomplex; SuoiBe Formation; YenChau Formation; NamMa Formation;  TuLe_NgoiThia Complex; NamPo Formation; BanMuong Complex; MongHinh Formation;  NamThep Formation; HaCoi Formation; TramTau Formation;
4	Group 4	TRIAS- TRIASIC	SuoiBang Formation; SongBoi Formation; NamTham Formation; NaKhuat Formation;  DongGiao Formation; KhonLang Formation; TamDao Formation; TanLac Formation;  VienNam Formation; Bavi Complex; CoNoi Formation; MongTrai Formation;  PhiaBioc Complex; PacMa Formation; BanXang Complex; SongMa Complex;  HoangMai Formation; DongTrau Formation; SuoiBang Formation;
5	Group 5	DEVON_DEVONIAN	BanCai Formation; BanPap Formation; BanNguon Formation; SongMua Formation;  NamPia Formation; BoHieng Formation; SinhVinh Formation; TayTrang Formation;  HuoiNhi Formation; NamSap Formation; MiaLe Formation;
6	Group 6	CAMBRI_ORDOVIC	BenKhe Formation; PoSen Complex; BanNgam Complex; MuongHum Complex;

No	Group	Name	Complex and Formation type
			CamDuong Formation; NuiNa Complex; BoXinh Complex; ChiengKhong Complex;
			HamRong Formation; SongMa Formation; DongSon Formation; ChangPung Formation;
			HaGiang Formation
7	Group 7	NEOPROTEROZOIC CAMBRI	AnPhu Formation; ThacBa Formation; DaDinh Formation; ChaPa Formation;
			NamCo Formation; NamTy Formation; BoXinh Group; Nam Si Formation; HuoiHao Formation;
			SinhQuyen Formation;
8	Group 8	CARBON-PERMI	DaNieng Formation; BacSon Formation; SiPhay Formation; NaVang Formation;
			CamThuy Formation; YenDuyet Formation; BanDiet Group; DienBienPhu Complex;
			SongDa Formation; DienThuong Complex; PhuSiLung Complex; MuongLat Complex;
9	Group 9	Unknown	Unknown in age dykes and veins

Table 2  
Landslide influencing factors and their classes

Factor	Classes	Classification method
Elevation (m)	(1) 70–300, (2) 300–500, (3) 500–700, (4) 700–900, (5) 900–1100, (6) 1100–1300, (7) 1300–1500, (8) 1500–2000, (9) > 2000	Natural Breaks
Slope angle (degree)	(1) 0–10, (2) 10–20, (3) 20–30, (4) 30–40, (5) 40–50, (6) > 50	Natural Breaks
Aspect	(1) Flat (-1), (2) North (0-22.5), (3) Northeast (22.5–67.5), (4) East (67.5-112.5), (5) Southeast (112.5-157.5), (6) South (157.5-202.5), (7) Southwest (202.5-247.5), (8) West (247.5-292.5), (9) Northwest (292.5-337.5)	Azimuth
Curvature	(1) [(-9.786) - (-0.625)], (2) [(-0.625) - (-0.173)], (3) [(-0.173) - 0.208], (4) [(0.208-0.659], (5) [0.659-9.717],	Natural Breaks
Elevation difference	(1) [0-132.8], (2) [132.8-232.4], (3) [232.4-312.1], (4) [312.1-385.2], (5) [385.2-464.8], (6) [464.8-557.8], (7) [557.8-684.0], (8) [684.0-1035.9], (9) [1035.9-1700]	
TWI	(1) [2.283–4.542], (2) [4.542–5.432], (3) [5.432–6.459], (4) [6.459–7.759], (5) [7.759–9.334], (6) [9.334–11.251], (7) [11.251–19.808]	Natural Breaks
NDVI	(1) [(-0.643) - (-0.038)], (2) [(-0.038)-0.009], (3) [0.009–0.051], (4) [0.051–0.093], (5) [0.093–0.694],	Natural Breaks
Rainfall (mm)	(1) [972–1038], (2) [1038–1089], (3) [1089–1143], (4) [1143–1194], (5) [1194–1279]	Comile rainfall with DEM using geostatistical Kriging method
Drainage density (km/km <sup>2</sup> )	(1) [0-2.13], (2) [2.13–4.27], (3) [4.27–6.40], (4) [6.40–8.54], (5) [8.54–10.68], (6) [10.68–12.82], (7) [12.82–14.95], (8) [14.95–17.08], (9) [17.08–19.22]	Natural Breaks
Distance to road (m)	(1) [0–50], (2) [50–100], (3) [100–200], (4) [200–500], (5) [500–1000], (6) > 1000	Natural Breaks
Road density (km/km <sup>2</sup> )	(1) [0-2.077], (2) [2.077–3.676], (3) [3.676–5.274], (4) [5.274–7.671], (5) [7.671–13.638],	Natural Breaks
Distance to river (m)	(1) [0-200], (2) [200–500], (3) [500–1000], (4) [1000–1500], (5) [1500–2000], (6) [2000–2500], (7) > 2500	Natural Breaks
Geohydrology	(1) Rich layer of water, (2) Middle poor layer of ware, (3) Very poor layer of water, (4)Poor layer of water	Heohydrological categories
Geological	(1) Group1, (2) Group2, (3) Group3, (4) Group4, (5) Group5, (6) Group6, (7) Group7, (8) Group8, (9) Group9	Geological group

Factor	Classes	Classification method
Geomorphology	(1) Valley of invasion, (2) Cavitation plateaus develop on carbonate rocks, (3) Driftwood washing plateau grows on carbonate rock, (4) Erosion plateaus develop on carbonate rocks, (5) Erosion and erosion plateaus develop on carbonate rocks, (6) Cavitation mountain range growing on carbonate rock, (7) Massive and structural mountain ranges developed on non-carbonate rocks, (8) Erosion and erosion mountain ranges develop on rocks, (9) Erosion massif develops on carbonate rock, (10) Masses and eroded mountain ranges develop on rocks, (11) The valley erodes and accumulates, (12) Cavitation mountain range growing on non-carbonate rock, (13) The mountain range erodes the structure growing on the rock, (14) Karst Funnel, (15) Invasion valley	Geomorphological categories
Landcover	(1) Bare ground, (2) Built area, (3) Clouds, (4) Crops, (5) Flooded vegetation, (6) Grass, (7) Scrub/Shrub, (8) Tree, (9) Water	Landcover categories

## 3.4 Landslide susceptibility modeling

The base classifier model, Multilayer Perceptron (MLP), is used in this study. Five ML ensemble techniques were developed with Multilayer Perceptron (MLP) as a base classifier, including Bagging – MLP (BAMLP), Dagging – MLP (DAMLP), Decorate – MLP (DEMLP), Rotation Forest – MLP (RFMLP), Random SubSpace – MLP (RSSMLP).

### 3.4.1 Multilayer Perceptron (MLP)

The MLP method is one of the most common artificial neural network approaches and is widely applied in landslide susceptibility assessment (Gómez & Kavzoglu, 2005; Li et al., 2019). The backpropagation algorithm is the training rule of MLP (Gómez & Kavzoglu, 2005). The goal function's minimal value and the method's optimal weight values may be modified and calculated (Li et al., 2019). This model has three main components: input, hidden, and output layers. Landslide influencing factors are considered the input layers. The output layers are made up of the categorized findings that are used to classify landslides and non-landslides. The hidden layers are applied to transform inputs into outputs (Gómez & Kavzoglu, 2005). The number of neurons in the input layer of  $X = (x_1, x_2, \dots, x_{m0})$ , hidden layer, and output layer are designated as  $m_0$ ,  $m_1$ , and  $m_2$  respectively if the MLP model comprises multi-input variables and multi-output variables. The input and output of neurons in the hidden layer are calculated as follows:

$$h_j = \sum_{i=1}^{m_0} w_{ij} y_i + \theta_j \quad (3)$$

$$z_j = f(h_j) = (1 + e^{-h_j})^{-1} \quad (4)$$

where  $h_j$ ,  $\theta_j$  and  $z_j$  denote the input, the threshold, and the output of the  $j$ th neuron in the hidden layer;  $w_{ij}$  denotes the weight value of the  $i$ th input and the  $j$ th in the hidden layer neurons;  $f(h_j)$  denotes the activating function. Afterward, the inputs and outputs of output layer neurons are represented as below:

$$h_k = \sum_{j=1}^{m_1} w_{jk} y_j + \theta_k \quad (5)$$

$$z_k = h_k \quad (6)$$

where  $h_k$ ,  $\theta_k$  and  $z_k$  denote the  $k$ th input, the threshold, and the output in the output layer neurons;  $w_{jk}$  denotes the weight value between the  $j$ th hidden layer neuron and the  $k$ th output layer neuron.

### 3.4.2 Bagging (BA)

Bootstrap aggregating or Bagging algorithm was proposed by Breiman (1996). This algorithm was applied to achieve an aggregated predictor based on different bootstrap samples (Breiman, 1996). This algorithm uses a training dataset  $T(x_k, y_k)$ , where  $x_k \in Q$ ,  $y_k \in \{\text{landslide; non-landslide}\}$ ,  $k = 1 - M$  and  $M$  is the number of bootstrap samples. Next, a bootstrap sample  $T_k$  is generated from the initial training dataset based on the replacement method. Therefore, the model is formed through a base classifier  $B_k$  using the bootstrap sample  $T_k$  and a classifier  $B_k(x)$  is developed from each bootstrap sample  $T_k$ . Finally, the classifier  $B^*$  is synthesized from  $B_1, B_2, \dots, B_n$  and calculated as below (Bauer & Kohavi, 1999):

$$B^*(x) = \operatorname{argmax}_{y \in Y} \sum_{i=1}^n 1(B_i(x) = y) \quad (7)$$

where  $B_i(x)$  represents a classifier that is generated from each bootstrap sample  $T_k$ .

### 3.4.3 Dagging (DA)

Dagging was first proposed by Ting & Witten (1997). This method determined the final prediction based on the majority vote. In this technique, the training dataset was divided into many separate classified parts, and each part of the data corresponded with a basic learner (Ting & Witten, 1997). If the input

training dataset D has K samples, the dagging algorithm creates N datasets from the input training dataset. Each dataset consists of k samples ( $kN < k$ ), and other datasets do not contain a similar sample. After that, each dataset would be trained by a basic classifier to build a classification model. Thus, N classification models can be formed from the N original datasets. These models make their prediction classes based on the given query samples. Finally, the prediction class of the dagging method has the most votes (Chen & Li, 2020).

### 3.4.4 Decorate (DE)

The Decorate was developed by Melville & Mooney (2003). Decorate is an ensemble meta-learner to create Diverse Ensemble Creation by Oppositional Relabeling of Artificial Training Examples (DECORATE). This algorithm is applied to create a new classifier based on the combination of original training datasets. The content of the Decorate algorithm can be understood as follows (Melville & Mooney, 2003):

1. Input a training dataset  $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_p, y_p)\}$  with  $x_i \in R^b, y_i \in Y = \{l_1, l_2, \dots, l_q\}, i = 1, 2, \dots, p; j = 1, 2, \dots, q$ , in which  $p$  denotes the number of training sub-datasets ( $p > 1$ ),  $b$  denotes the number of attributes in each sub-dataset ( $b > 1$ ),  $m$  denotes the number of class labels in a classification model ( $q > 1$ ).
2. The base learning algorithm – *BaseLearn* - is used to train a classifier  $C_1$  from the original input dataset D, and the first ensemble  $C^* = (C_1)$  is got.
3. Decorate Algorithm Generates Classifiers Iteratively.

### 3.4.5 Rotation Forest (RF)

Rotation Forest (RF) was an ensemble learning method proposed by Rodríguez et al. (2006). This method developed an ensemble of decision trees based on the random subspace and bagging methodology with principal component analysis (PCA). In this technique, the input training datasets were divided into many sub-datasets to create the classifiers, and PCA was used for each sub-dataset (Kuncheva & Rodríguez, 2007).

The input training dataset is E, the class label of the input training dataset is F, and the feature set is G. If the quantity of training times is T with t features, then E is  $T \cdot t$  matrix. Assuming that class labels F are formed based on the class set of  $H = (h_1, h_2, \dots, h_j)$ ; the feature set is allocated into M sub-datasets, P-decision tree in a forest rotation of  $D = (D_1, D_2, \dots, D_P)$ . According to that, two indices – M and P – need to be pre-calculated. The procedure to determine the input training dataset consists of four steps (Sahin et al., 2020):

- (1) Divide the feature set G into M feature subset, and the attribute of each feature subset of  $N = n/M$ ;

(2)  $G_{ij}$  is the attribute subset to train the classifier  $D_i$  and  $Y_{ij}$  is the dataset in  $G_{ij}$ ;

(3) Construct a rotation matrix  $R_i$  with the nominated ratio in the matrix  $F_{ij}$ . The coefficients of the matrix  $F_{ij}$  include:  $f_{ij}^{(1)}, f_{ij}^{(2)}, \dots, f_{ij}^{(Nj)}$ , that is determined based on a linear transformation;

(4) The rotation matrix  $R_i$  can be calculated by:

$$R_i = \begin{bmatrix} f_{i1}^{(1)}, \dots, f_{i1}^{(N_1)} & 0 & \dots & 0 \\ 0 & f_{i2}^{(1)}, \dots, f_{i2}^{(N_2)} & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & f_{iP}^{(1)}, \dots, f_{iP}^{(N_P)} \end{bmatrix} \quad (8)$$

In addition, the classification step can be evaluated for a given case  $y$ , then the classifier  $Q_i = q_{ij}(yR_i^b)$  will be used to classify the probability of this case. In this way, the confidence of a class can be determined as the following equation:

$$\sigma_j = \frac{1}{m} \sum_{i=1}^m q_{ij}(yR_i^b), j = 1, 2, \dots, d \quad (9)$$

Finally,  $y$  is assigned to a class which has the largest confidence determined.

### 3.4.6 Random SubSpace (RSS)

The Random SubSpace (RSS) was first proposed by Ho (1998). The training dataset was created in the modified feature space in the RSS method to build a higher number of training variables (Ho, 1998). The RSS algorithm can be expressed as follows: Input a training dataset  $D(x_i, y_i)$ ; where  $x_i \in T$  and is a  $m$ -dimensional vector  $x_i = (x_{i1}, x_{i2}, \dots, x_{im})$ ,  $y_i \in \{\text{landslide; non-landslide}\}$ . First,  $m*$  features are randomly selected from the training dataset, where  $m* < m$ . In this way, the  $m*$  dimensional random subspace of the original  $m$ -dimensional feature space is generated. Second, the modified training dataset  $D^*$  consists of  $m*$ -dimensional training features  $x_i = (x_{i1}, x_{i2}, \dots, x_{im*})$ . Finally, a final classifier is developed based on the combination of primary classifiers according to a voting scheme (Lai et al., 2006).

## 3.5 Model validation and comparison

It is required to validate the prediction models to assess the applicable ability. We used a variety of statistical metrics to assess the performance of five suggested ensemble models in this research. They are Positive Predictive Value (PPV), Negative Predictive Value (NPV), Sensitivity, Specificity, Accuracy (ACC), F-measure, Jaccard, Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Reciever Operating Characteristic (ROC) curve, and Area Under the ROC Curve (AUC). Sensitivity and Specificity denote the true positive rate and true negative rate of landslide and non-landslide locations. Accuracy is the mathematics average of Sensitivity and Specificity or the scale between the number of landslide and non-landslide pixels. F-measure is expressed as a weighted harmonious average of the accuracy and revocation in binary classification. Jaccard is the Jaccard sameness coefficient factor. The ACC, Kappa, MSE, and RMSE statistical indicators are determined as in the following equations:

$ACC = \frac{TN + TP}{TN + FN + TP + FP}$	(10)
$Kappa = \frac{ACC - ACC\_EXP}{1 - ACC\_EXP}$	(11)
$MSE = \sum_{i=1}^m \frac{(X_{Act.} - X_{Pred.})^2}{N}$	(12)
$RMSE = \sqrt{\sum_{i=1}^m \frac{(X_{Act.} - X_{Pred.})^2}{N}}$	(13)

in which TP denotes true positive, TN denotes true negative, FP denotes false positive, and FN denotes false negative.

RMSE represents the level of dispersion of predictive values from actual values. The less the RMSE value is, the higher the accuracy of the prediction model is. Meanwhile, the AUC indicator is considered the main value in assessing the performance of the predictive models (Ye et al., 2016).

## 4. Results

### 4.1 Multicollinearity and factor selection

Variance Inflation Factors (VIF) and Tolerance parameters were used for checking the multicollinearity of the landslide causative factors in order to select optimal input factors. The tolerance and VIF determine the transformation in the standard errors of the attended factors (Bui et al., 2019). ). The multicollinearity problem of used factors will happen if VIF is greater than 10 or tolerance is less than 0.1 (Lin et al., 2017). Figure 4 shows the results of VIF and tolerance calculated for sixteen landslide influencing factors. All factors have  $VIF > 10$  and  $\text{tolerance} < 0.1$ , therefore, they can be applied to build landslide susceptibility modelings in this research.

## **4.2 Validation and comparison of the susceptibility models**

In this study, five advanced hybrid ML models (BAMLP, DAMLP, DEMLP, RFMLP, RSSMLP) and the single model (MLP) were applied to generate landslide susceptibility models. The cross-validation method was applied to validate these used models. Several standard quantitative indexes were tested on training and validating datasets (Table 3 and Fig. 5).

Table 3  
Indexes for the model's performance assessment.

<b>Sample</b>	<b>Parameters</b>	<b>MLP</b>	<b>BAMLP</b>	<b>DAMLP</b>	<b>DEMLP</b>	<b>RFMLP</b>	<b>RSSMLP</b>
Training	TP	1109	1121	920	1147	1133	1122
	TN	963	1088	816	1120	1119	1081
	FP	67	55	256	29	43	54
	FN	226	101	373	69	70	108
	PPV(%)	94.30	95.32	78.23	97.53	96.34	95.41
	NPV(%)	80.99	91.51	68.63	94.20	94.11	90.92
	SST(%)	83.07	91.73	71.15	94.33	94.18	91.22
	SPF(%)	93.50	95.19	76.12	97.48	96.30	95.24
	ACC (%)	99.95	96.14	81.38	98.12	96.38	96.56
	F-Measure (%)	88.33	93.49	74.52	95.90	95.25	93.27
	Jaccard (%)	79.10	87.78	59.39	92.13	90.93	87.38
	MCC (%)	78.11	87.67	60.78	92.05	90.90	87.23
	RMSE	0.3122	0.2593	0.4175	0.327	0.2175	0.3192
Validating	TP	340	380	377	403	374	395
	TN	384	341	342	327	363	347
	FP	167	126	129	103	132	111
	FN	122	166	165	180	144	160
	PPV(%)	67.06	75.10	74.51	79.64	73.91	78.06
	NPV(%)	75.89	67.26	67.46	64.50	71.60	68.44
	SST(%)	73.59	69.60	69.56	69.13	72.20	71.17
	SPF(%)	69.69	73.02	72.61	76.05	73.33	75.76
	ACC (%)	64.33	78.03	77.10	86.02	74.65	81.87
	F-Measure (%)	70.18	72.24	71.95	74.01	73.05	74.46
	Jaccard (%)	54.05	56.55	56.18	58.75	57.54	59.31
	MCC (%)	59.06	58.84	58.70	59.28	60.34	60.63
	RMSE	0.5031	0.422	0.4334	0.4387	0.4182	0.4342

For the training dataset, the DEMLP model has the highest Sensitivity (94.33%), Specificity (97.48%), Accuracy (98.12%), F-Measure (95.90%), Jaccard (92.13%), and MCC (92.05%) values in comparison to the remaining models (BAMLP, DAMLP, RFMLP, RSSMLP, and MLP). The RFMLP model has the best value of RMSE (0.218). The ROC curve analysis also reflects that the DEMLP model has the best value of AUC (0.992), followed by RFMLP (0.985), RSSMLP (0.982), BAMLP (0.981), MLP (0.915), and DAMLP (0.827). The training dataset results indicate that the DEMLP and RFMLP models have the best performance.

For the validating dataset, the DEMLP model has the highest value of Specificity (76.05%), Accuracy (86.02%) in comparison to the remaining models (BAMLP, DAMLP, RFMLP, RSSMLP, and MLP). The MLP model has the highest value of Sensitivity (73.59%); meanwhile, the RSSMLP has the best F-Measure (74.46%), Jaccard (59.31%), and MCC (60.63%). The RFMLP model has the best value of RMSE (0.418). According to the ROC curve analysis, the RFMLP model had the best value of AUC (0.818), followed by the DEMLP and BAMLP models (0.804), the RSSMLP model (0.796), the DAMLP model (0.789), and the MLP model (0.698). The testing dataset results indicate that the DEMLP and RFMLP models have the best predictive ability in predicting landslide susceptibility. It can be seen that five advanced ensemble models (BAMLP, DAMLP, DEMLP, RFMLP, and RSSMLP) had more accurate predictions than a single MLP model. To sum up, the RFMLP is the best predictive model among used models (BAMLP, DAMLP, DEMLP, RSSMLP, and MLP) (Fig. 6).

## 4.3 Landslide susceptibility mapping

Five advanced hybrid ML models and a single MLP model were used to create the landslide susceptibility maps based on the training dataset. These resulting maps were divided into five susceptibility classes: very low, low, moderate, high, and very high susceptibility, based on the Natural Break technique in ArcGIS Pro (Fig. 7).

Based on the statistical indicators in Table 3, the RFMLP model is the best prediction model compared to the other models (DEMLP, BAMLP, DAMLP, RSSMLP, and MLP). As a result, the RFMLP model was chosen to create the most accurate landslide susceptibility map for the research region. The final results represented 217,377 ha (15.45%) in very high susceptibility areas; 303,376 ha (21.56%) in high susceptibility areas, 271,515 ha (19.30%) in moderate susceptibility areas, 215,139 ha (15.29%) in low susceptibility areas, and 399,679 ha (28.40%) in very low susceptibility areas (Fig. 8).

## 5. Discussion

Landslides are dangerous and devastating geological procedures in mountainous regions (Šilhán, 2020). They have caused thousands of deaths and injuries annually and are the main natural hazards that cause loss of life in Asia (Highland & Bobrowsky, 2008). Therefore, the information on landslide susceptible areas needs to be updated continuously because of their complex geological processes (Chen & Li, 2020; Van Westen et al., 2008). The present study provided a feasible approach for

susceptible landslide evaluation using advanced ensemble ML models. The five advanced ensemble ML models were also developed to estimate predictive landslide susceptibility maps.

ML approaches have been considered an effective and potential method in natural hazard assessment because of their flexibility and predictive accuracy (Bui et al., 2019). Hybrid ML models, in particular, do not rely on statistical assumptions and may quantify the relevance and effect of landslide-related factors (Achour & Pourghasemi, 2020). Previously, single models, such as the MLP model, were often applied to build landslide susceptibility maps (Adnan et al., 2020; Meghanadh et al., 2022; Zare et al., 2013). The results in Table 3 and Fig. 6 also showed that ensemble ML models had better predictive accuracy than a single ML model. Several studies have currently developed hybrid models that combine MLP and other ML models; these hybrid models showed better predictive accuracy. Li et al. (2019) employed an ensemble of PSO-MLP to create a landslide susceptibility map for Shicheng County in China with the AUC = 0.881. In another study, Hong et al. (2020) applied an ensemble of MLP-SGD to predict the landslide susceptible areas for Yanshan County in Jiangxi province in China with the AUC = 0.822. In this study, we explored advanced hybrid ML models, namely BAMPL, DAMLP, DEMLP, RFMLP, and RSSMLP, with the MLP as a base classifier model. The validation results showed that the RFMLP model has the highest predictive accuracy of AUC = 0.818. Compared with the previous studies, the RFMLP ensemble model with the case study of Son La province was found to be appropriate. This study also provided a new and efficient hybrid model for assessing landslide susceptibility in mountainous areas.

Son La is situated in the center of the Northwest area of Vietnam, and extensive landslides often destroy many regions during the rainy season (Ahlheim et al., 2008). The increasing of landslide events has threatened human lives and properties in the study area (IFRC, 2021). The landslide susceptibility maps can supply relevant information for government agencies and local authorities to implement planning, land uses sustainable management, or build early warning systems (Bălteanu et al., 2020). The result of the susceptibility map may supply helpful information for analyzing and evaluating the landslide risk with practical meanings at the local scale. This present study generally added a new approach to assessing landslide susceptibility based on ML techniques.

## 6. Conclusion

We used state-of-the-art hybrid ML models to create landslide susceptibility prediction maps in this work. Historical landslide placements, as well as precise geo-environmental, meteorological, hydrogeological, geological, and topographical data, are included in the input data. Several statistical indices were utilized to verify and compare the forecast models that were employed. It can be seen that the predictive accuracy values of the used models are good. The RFMLP model has the best AUC (0.818), followed by the DEMLP and BAMPL (0.804), the RSSMLP (0.796), the DAMLP (0.789), and the MLP (0.698). The RFMLP ensemble model was selected to build the landslide susceptibility prediction map for the research region. This resulting map can supply useful information for policy-makers and decision-makers in preventing landslide risks in the future. This research also supports the use of ML approach to analyze and manage natural risks in mountainous locations.

## Declarations

**Funding:** This research is funded by the Hanoi University of Civil Engineering (HUCE) under grant number 19-2021/KHxD-TĐ.

**Conflicts of Interest:** The authors declare no conflict of interest.

**Availability of data and material:** The corresponding author will provide data supporting the findings of this study upon reasonable request.

**Code availability (software application or custom code):** Not applicable

## References

1. Achour Y, Pourghasemi HR (2020) How do machine learning techniques help in increasing accuracy of landslide susceptibility maps? *Geosci Front.* <https://doi.org/10.1016/j.gsf.2019.10.001>
2. Adnan MSG, Rahman MS, Ahmed N, Ahmed B, Rabbi MF, Rahman RM (2020) Improving spatial agreement in machine learning-based landslide susceptibility mapping. *Remote Sens.* <https://doi.org/10.3390/rs12203347>
3. Ahlheim M, Frör O, Heinke A, Keil A, Nguyen MD, Pham VD, Saint-Macary C, Zeller M (2008) *Landslides in mountainous regions of Northern Vietnam: causes, protection strategies and the assessment of economic losses*
4. Akgün A, Bulut F (2007) GIS-based landslide susceptibility for Arsin-Yomra (Trabzon, North Turkey) region. *Environmental Geology*. <https://doi.org/10.1007/s00254-006-0435-6>
5. Aleotti P, Chowdhury R (1999) Landslide hazard assessment: Summary review and new perspectives. *Bull Eng Geol Environ* 58(1):21–44. <https://doi.org/10.1007/s100640050066>
6. Arabameri A, Saha S, Roy J, Chen W, Blaschke T, Bui DT (2020) Landslide susceptibility evaluation and management using different machine learning methods in the Gallicash River Watershed, Iran. *Remote Sens.* <https://doi.org/10.3390/rs12030475>
7. Armaş I (2012) Weights of evidence method for landslide susceptibility mapping. Prahova Subcarpathians, Romania. *Natural Hazards*. <https://doi.org/10.1007/s11069-011-9879-4>
8. Bălteanu D, Micu M, Jurchescu M, Malet JP, Sima M, Kucsicsa G, Dumitrică C, Petrea D, Mărgărint MC, Bilaşco Ş, Dobrescu CF, Călăraşu EA, Olinic E, Boţi I, Senzaconi F (2020) National-scale landslide susceptibility map of Romania in a European methodological framework. *Geomorphology*. <https://doi.org/10.1016/j.geomorph.2020.107432>
9. Bauer E, Kohavi R (1999) Empirical comparison of voting classification algorithms: bagging, boosting, and variants. *Mach Learn* 36(1):105–139. <https://doi.org/10.1023/a:1007515423169>
10. Bragagnolo L, da Silva RV, Grzybowski JMV (2020) Landslide susceptibility mapping with r.landslide: A free open-source GIS-integrated tool based on Artificial Neural Networks. *Environ Model Softw.* <https://doi.org/10.1016/j.envsoft.2019.104565>

11. Breiman L (1996) Bagging predictors. *Mach Learn* 24(2):123–140.  
<https://doi.org/10.1007/bf00058655>
12. Bui DT, Shahabi H, Omidvar E, Shirzadi A, Geertsema M, Clague JJ, Khosravi K, Pradhan B, Pham BT, Chapi K, Barati Z, Ahmad B, Rahmani B, Gróf H, Lee S (2019) Shallow landslide prediction using a novel hybrid functional machine learning algorithm. *Remote Sens* 11(8):931–953.  
<https://doi.org/10.3390/rs11080952>
13. Chen W, Li Y (2020) GIS-based evaluation of landslide susceptibility using hybrid computational intelligence models. *CATENA*. <https://doi.org/10.1016/j.catena.2020.104777>
14. Chen W, Zhang S, Li R, Shahabi H (2018) Performance evaluation of the GIS-based data mining techniques of best-first decision tree, random forest, and naïve Bayes tree for landslide susceptibility modeling. *Sci Total Environ*. <https://doi.org/10.1016/j.scitotenv.2018.06.389>
15. Chen W, Zhao X, Shahabi H, Shirzadi A, Khosravi K, Chai H, Zhang S, Zhang L, Ma J, Chen Y, Wang X, Ahmad B, Li R (2019) Spatial prediction of landslide susceptibility by combining evidential belief function, logistic regression and logistic model tree. *Geocarto Int*.  
<https://doi.org/10.1080/10106049.2019.1588393>
16. Corsini A, Pasuto A, Soldati M, Zannoni A (2005) Field monitoring of the Corvara landslide (Dolomites, Italy) and its relevance for hazard assessment. *Geomorphology*.  
<https://doi.org/10.1016/j.geomorph.2004.09.012>
17. Dai FC, Lee CF, Ngai YY (2002) Landslide risk assessment and management: An overview. *Eng Geol*.  
[https://doi.org/10.1016/S0013-7952\(01\)00093-X](https://doi.org/10.1016/S0013-7952(01)00093-X)
18. Dung N, Van, Hieu N, Phong T, Van, Amiri M, Costache R, Al-Ansari N, Prakash I, Le H, Van, Nguyen HBT, Pham BT (2021) Exploring novel hybrid soft computing models for landslide susceptibility mapping in Son La hydropower reservoir basin. *Geomatics Nat Hazards Risk*.  
<https://doi.org/10.1080/19475705.2021.1943544>
19. Feizizadeh B, Blaschke T (2013) GIS-multicriteria decision analysis for landslide susceptibility mapping: Comparing three methods for the Urmia lake basin, Iran. *Natural Hazards*.  
<https://doi.org/10.1007/s11069-012-0463-3>
20. Frodella W, Spizzichino D, Ciampalini A, Margottini C, Casagli N (2021) Hydrography and geomorphology of Antananarivo High City (Madagascar). *J Maps* 17(4):215–226.  
<https://doi.org/10.1080/17445647.2020.1721343>
21. Goetz JN, Brenning A, Petschko H, Leopold P (2015) Evaluating machine learning and statistical prediction techniques for landslide susceptibility modeling. *Comput Geosci* 81:1–11.  
<https://doi.org/10.1016/j.cageo.2015.04.007>
22. Gómez H, Kavzoglu T (2005) Assessment of shallow landslide susceptibility using artificial neural networks in Jabonosa River Basin, Venezuela. *Eng Geol*.  
<https://doi.org/10.1016/j.enggeo.2004.10.004>
23. Guns M, Vanacker V (2013) Forest cover change trajectories and their impact on landslide occurrence in the tropical Andes. *Environ Earth Sci*. <https://doi.org/10.1007/s12665-013-2352-9>

24. Guzzetti F, Cardinali M, Reichenbach P, Cipolla F, Sebastiani C, Galli M, Salvati P (2004) Landslides triggered by the 23 November 2000 rainfall event in the Imperia Province, Western Liguria, Italy. *Engineering Geology*. <https://doi.org/10.1016/j.enggeo.2004.01.006>
25. Ha H, Luu C, Bui QD, Pham D-H, Hoang T, Nguyen V-P, Vu MT, Pham BT (2021) Flash flood susceptibility prediction mapping for a road network using hybrid machine learning models. *Nat Hazards* 109(1):1247–1270. <https://doi.org/10.1007/s11069-021-04877-5>
26. Hang HT, Tung H, Hoa PD, Phuong NV, Phong T, Van, Costache R, Nguyen HD, Amiri M, Le H-A, Le H, Van, Prakash I, Pham BT (2021) Spatial prediction of landslides along National Highway-6, Hoa Binhh province, Vietnam using novel hybrid models. *Geocarto Int* 1–26. <https://doi.org/10.1080/10106049.2021.1912195>
27. Harp EL, Jibson RW (1996) Landslides triggered by the 1994 Northridge, California, earthquake. *Bulletin of the Seismological Society of America*
28. Hemasinghe H, Rangali RSS, Deshapriya NL, Samarakoon L (2018) Landslide susceptibility mapping using logistic regression model (a case study in Badulla District, Sri Lanka). *Procedia Eng*. <https://doi.org/10.1016/j.proeng.2018.01.135>
29. Highland LM, Bobrowsky P (2008) The landslide Handbook - A guide to understanding landslides. US Geological Survey Circular
30. Ho TK (1998) The random subspace method for constructing decision forests. *IEEE Trans Pattern Anal Mach Intell* 20(8):832–844. <https://doi.org/10.1109/34.709601>
31. Hoang ND, Tien Bui D (2018) Spatial prediction of rainfall-induced shallow landslides using gene expression programming integrated with GIS: a case study in Vietnam. *Nat Hazards*. <https://doi.org/10.1007/s11069-018-3286-z>
32. Hong H, Pourghasemi HR, Pourtaghi ZS (2016) Landslide susceptibility assessment in Lianhua County (China): A comparison between a random forest data mining technique and bivariate and multivariate statistical models. *Geomorphology*. <https://doi.org/10.1016/j.geomorph.2016.02.012>
33. Hong H, Tsangaratos P, Ilia I, Loupasakis C, Wang Y (2020) Introducing a novel multi-layer perceptron network based on stochastic gradient descent optimized by a meta-heuristic algorithm for landslide susceptibility mapping. *Sci Total Environ*. <https://doi.org/10.1016/j.scitotenv.2020.140549>
34. Hong Y, Adler R, Huffman G (2007) Use of satellite remote sensing data in the mapping of global landslide susceptibility. *Nat Hazards*. <https://doi.org/10.1007/s11069-006-9104-z>
35. Hu X, Huang C, Mei H, Zhang H (2021) Landslide susceptibility mapping using an ensemble model of Bagging scheme and random subspace-based naïve Bayes tree in Zigui County of the Three Gorges Reservoir Area, China. *Bulletin of Engineering Geology and the Environment*. <https://doi.org/10.1007/s10064-021-02275-6>
36. Huang Y, Zhao L (2018) Review on landslide susceptibility mapping using support vector machines. In *Catena*. <https://doi.org/10.1016/j.catena.2018.03.003>
37. IFRC (2021) *Viet Nam, Flooding, Landslide and Whirlwinds in Son La Province* (24 Aug 2021). <https://reliefweb.int/report/viet-nam/viet-nam-flooding-landslide-and-whirlwinds-son-la-province-24->

aug-2021

38. Jaafari A, Najafi A, Pourghasemi HR, Rezaeian J, Sattarian A (2014) GIS-based frequency ratio and index of entropy models for landslide susceptibility assessment in the Caspian forest, northern Iran. *Int J Environ Sci Technol.* <https://doi.org/10.1007/s13762-013-0464-0>
39. Juliev M, Mergili M, Mondal I, Nurtaev B, Pulatov A, Hübl J (2019) Comparative analysis of statistical methods for landslide susceptibility mapping in the Bostanlik District, Uzbekistan. *Sci Total Environ.* <https://doi.org/10.1016/j.scitotenv.2018.10.431>
40. Kavzoglu T, Colkesen I, Sahin EK(2019) Machine learning techniques in landslide susceptibility mapping: A survey and a case study. In *Advances in Natural and Technological Hazards Research.* [https://doi.org/10.1007/978-3-319-77377-3\\_13](https://doi.org/10.1007/978-3-319-77377-3_13)
41. Kayastha P, Dhital MR, De Smedt F(2012) Landslide susceptibility mapping using the weight of evidence method in the Tinau watershed, Nepal. *Natural Hazards.* <https://doi.org/10.1007/s11069-012-0163-z>
42. Kim JC, Lee S, Jung HS, Lee S (2018) Landslide susceptibility mapping using random forest and boosted tree models in Pyeong-Chang, Korea. *Geocarto Int.* <https://doi.org/10.1080/10106049.2017.1323964>
43. Kuncheva LI, Rodríguez JJ(2007) An experimental study on rotation forest ensembles. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 4472 LNCS, 459–468.* [https://doi.org/10.1007/978-3-540-72523-7\\_46](https://doi.org/10.1007/978-3-540-72523-7_46)
44. Lai C, Reinders MJT, Wessels L (2006) Random subspace method for multivariate feature selection. *Pattern Recognit Lett.* <https://doi.org/10.1016/j.patrec.2005.12.018>
45. Lee S, Hong SM, Jung HS(2017) A support vector machine for landslide susceptibility mapping in Gangwon Province, Korea. *Sustainability (Switzerland).* <https://doi.org/10.3390/su9010048>
46. Li D, Huang F, Yan L, Cao Z, Chen J, Ye Z (2019) Landslide susceptibility prediction using particle-swarm-optimized multilayer perceptron: Comparisons with multilayer-perceptron-only, BP neural network, and information value models. *Appl Sci (Switzerland).* <https://doi.org/10.3390/app9183664>
47. Lin GF, Chang MJ, Huang YC, Ho JY (2017) Assessment of susceptibility to rainfall-induced landslides using improved self-organizing linear output map, support vector machine, and logistic regression. *Eng Geol.* <https://doi.org/10.1016/j.enggeo.2017.05.009>
48. Mahdadi F, Boumezbeur A, Hadji R, Kanungo DP, Zahri F (2018) GIS-based landslide susceptibility assessment using statistical models: a case study from Souk Ahras province, N-E Algeria. *Arab J Geosci.* <https://doi.org/10.1007/s12517-018-3770-5>
49. Meghanadh D, Kumar Maurya V, Tiwari A, Dwivedi R(2022) A multi-criteria landslide susceptibility mapping using deep multi-layer perceptron network: A case study of Srinagar-Rudraprayag region (India). *Advances in Space Research.* <https://doi.org/10.1016/j.asr.2021.10.021>
50. Melville P, Mooney RJ(2003) Constructing diverse classifier ensembles using artificial training examples. *IJCAI International Joint Conference on Artificial Intelligence, August,* 505–510

51. Mezughi TH, Akhir JM, Rafeq AG, Abdullah I (2011) Landslide susceptibility assessment using frequency ratio model applied to an area along the E-W highway (Gerik-Jeli). *Am J Environ Sci.* <https://doi.org/10.3844/ajessp.2011.43.50>
52. Mirus BB, Jones ES, Baum RL, Godt JW, Slaughter S, Crawford MM, Lancaster J, Stanley T, Kirschbaum DB, Burns WJ, Schmitt RG, Lindsey KO, McCoy KM(2020) Landslides across the USA: occurrence, susceptibility, and data limitations. In *Landslides*. <https://doi.org/10.1007/s10346-020-01424-4>
53. Myronidis D, Papageorgiou C, Theophanous S(2016) Landslide susceptibility mapping based on landslide history and analytic hierarchy process (AHP). *Natural Hazards*. <https://doi.org/10.1007/s11069-015-2075-1>
54. Nguyen LC, Tien P, Van, Do TN (2020) Deep-seated rainfall-induced landslides on a new expressway: a case study in Vietnam. *Landslides* 17(2):395–407. <https://doi.org/10.1007/s10346-019-01293-6>
55. Pham BT, Bui T, Prakash D, I., & Dholakia MB(2017) Hybrid integration of Multilayer Perceptron Neural Networks and machine learning ensembles for landslide susceptibility assessment at Himalayan area (India) using GIS. *Catena*. <https://doi.org/10.1016/j.catena.2016.09.007>
56. Pisano L, Zumpano V, Malek, Rosskopf CM, Parise M (2017) Variations in the susceptibility to landslides, as a consequence of land cover changes: A look to the past, and another towards the future. *Sci Total Environ.* <https://doi.org/10.1016/j.scitotenv.2017.05.231>
57. Pourghasemi HR, Kariminejad N, Amiri M, Edalat M, Zarafshar M, Blaschke T, Cerda A (2020) Assessing and mapping multi-hazard risk susceptibility using a machine learning technique. *Sci Rep* 10(1):1–11. <https://doi.org/10.1038/s41598-020-60191-3>
58. Pourghasemi HR, Mohammady M, Pradhan B(2012) Landslide susceptibility mapping using index of entropy and conditional probability models in GIS: Safarood Basin, Iran. *Catena*. <https://doi.org/10.1016/j.catena.2012.05.005>
59. Pradhan B, Oh HJ, Buchroithner M (2010) Weights-of-evidence model applied to landslide susceptibility mapping in a tropical hilly area. *Geomatics Nat Hazards Risk.* <https://doi.org/10.1080/19475705.2010.498151>
60. Promper C, Puissant A, Malet JP, Glade T (2014) Analysis of land cover changes in the past and the future as contribution to landslide risk scenarios. *Appl Geogr.* <https://doi.org/10.1016/j.apgeog.2014.05.020>
61. Raja NB, Çiçek I, Türkoğlu N, Aydin O, Kawasaki A (2017) Landslide susceptibility mapping of the Sera River Basin using logistic regression model. *Nat Hazards*. <https://doi.org/10.1007/s11069-016-2591-7>
62. Riaz MT, Basharat M, Hameed N, Shafique M, Luo J (2018) A Data-Driven Approach to Landslide-Susceptibility Mapping in Mountainous Terrain: Case Study from the Northwest Himalayas, Pakistan. *Nat Hazards Rev.* [https://doi.org/10.1061/\(asce\)nh.1527-6996.0000302](https://doi.org/10.1061/(asce)nh.1527-6996.0000302)
63. Rodríguez JJ, Kuncheva LI, Alonso CJ (2006) Rotation forest: A New classifier ensemble method. *IEEE Trans Pattern Anal Mach Intell.* <https://doi.org/10.1109/TPAMI.2006.211>

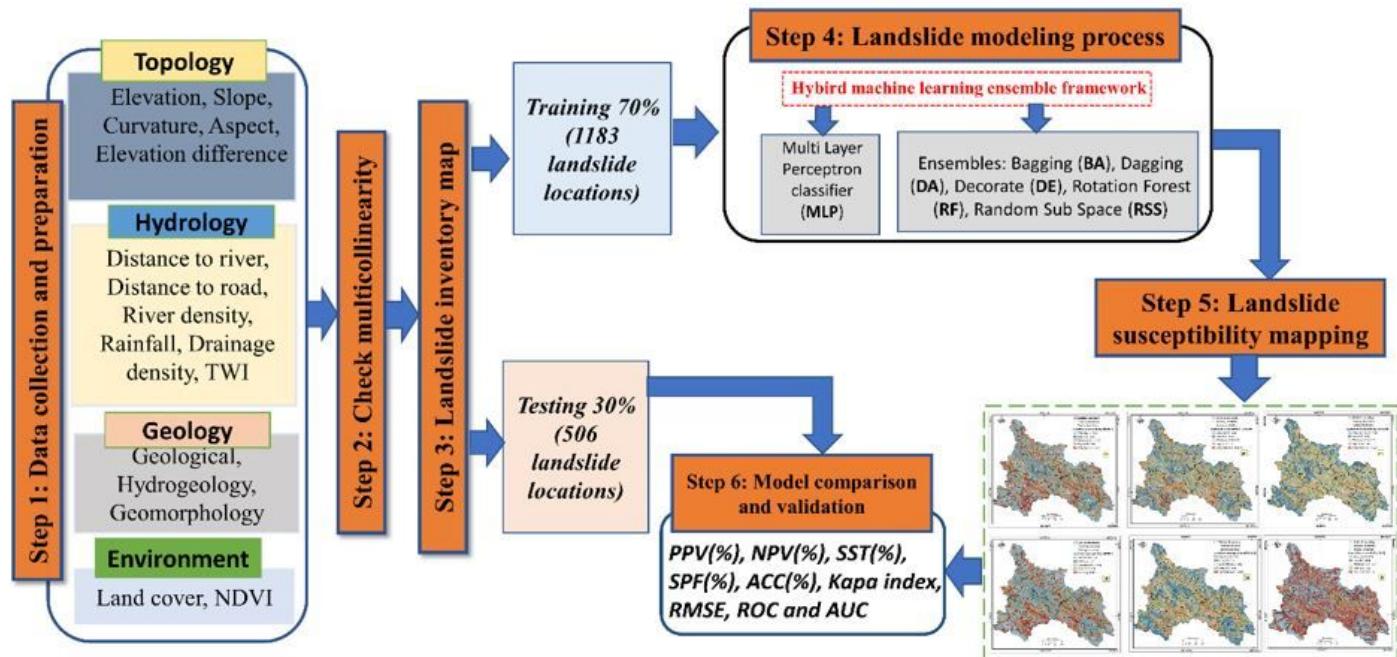
64. Sahin EK, Colkesen I, Kavzoglu T (2020) A comparative assessment of canonical correlation forest, random forest, rotation forest and logistic regression methods for landslide susceptibility mapping. *Geocarto Int* 35(4):341–363. <https://doi.org/10.1080/10106049.2018.1516248>
65. Šilhán K (2020) Dendrogeomorphology of landslides: principles, results and perspectives. *Landslides*. <https://doi.org/10.1007/s10346-020-01397-4>
66. Simon N, Crozier M, de Roiste M, Rafek AG, Roslee R (2015) Time series assessment on landslide occurrences in an area undergoing development. *Singap J Trop Geogr*. <https://doi.org/10.1111/sjtg.12096>
67. Singh P, Sharma A, Sur U, Rai PK (2021) Comparative landslide susceptibility assessment using statistical information value and index of entropy model in Bhanupali-Beri region, Himachal Pradesh, India. *Environment, Development and Sustainability*. <https://doi.org/10.1007/s10668-020-00811-0>
68. Skilodimou HD, Bathrellos GD, Koskeridou E, Soukis K, Rozos D (2018) Physical and anthropogenic factors related to landslide activity in the northern Peloponnese. Greece Land. <https://doi.org/10.3390/land7030085>
69. Sujatha ER (2021) An integrated landslide susceptibility model to assess landslides along linear infrastructure for environmental management. *Environ Earth Sci*. <https://doi.org/10.1007/s12665-021-09747-8>
70. Taalab K, Cheng T, Zhang Y (2018) Mapping landslide susceptibility and types using Random Forest. *Big Earth Data*. <https://doi.org/10.1080/20964471.2018.1472392>
71. Tehrany MS, Kumar L (2018) The application of a Dempster–Shafer-based evidential belief function in flood susceptibility mapping and comparison with frequency ratio and logistic regression methods. *Environ Earth Sci* 77(13):1–24. <https://doi.org/10.1007/s12665-018-7667-0>
72. Thach NN, Canh PX (2011) Using remote sensing and geographical information system to establish the landslide sensitivity map for Son La city area. *VNU J Science: Earth Environ Sci* 27(4):219–228
73. Ting KM, Witten IH (1997) Stacking bagged and dagged models. *Proc. of ICML'97*
74. Van Hoang T, Chou TY, Nguyen NT, Fang YM, Yeh ML, Nguyen QH, Nguyen XL (2019) A robust early warning system for preventing flash floods in mountainous area in Vietnam. *ISPRS Int J Geo-Information*. <https://doi.org/10.3390/ijgi8050228>
75. Van Phong T, Dam ND, Trinh PT, Van Dung N, Hieu N, Tran CQ, Van TD, Nguyen QC, Prakash I, Pham BT (2022) GIS-based Logistic Regression application for landslide susceptibility mapping in Son La Hydropower Reservoir Basin. *Lecture Notes in Civil Engineering* [https://doi.org/10.1007/978-981-16-7160-9\\_186](https://doi.org/10.1007/978-981-16-7160-9_186)
76. Van Westen CJ, Rengers N, Soeters R (2003) Use of geomorphological information in indirect landslide susceptibility assessment. *Nat Hazards*. <https://doi.org/10.1023/B:NHAZ.0000007097.42735.9e>
77. van Westen CJ, Castellanos E, Kuriakose SL (2008) Spatial data for landslide susceptibility, hazard, and vulnerability assessment: An overview. *Engineering Geology*. <https://doi.org/10.1016/j.enggeo.2008.03.010>

78. Wang Y, Fang Z, Wang M, Peng L, Hong H (2020) Comparative study of landslide susceptibility mapping with different recurrent neural networks. *Comput Geosci*. <https://doi.org/10.1016/j.cageo.2020.104445>
79. Wu TH, Tang WH, Einstein HH (1996) Landslide hazard and risk assessment. *Special Report - National Research Council, Transportation Research Board*
80. Ye F, Zhang L, Zhang D, Fujita H, Gong Z (2016) A novel forecasting method based on multi-order fuzzy time series and technical analysis. *Inf Sci*. <https://doi.org/10.1016/j.ins.2016.05.038>
81. Zare M, Pourghasemi HR, Vafakhah M, Pradhan B (2013) Landslide susceptibility mapping at Vaz Watershed (Iran) using an artificial neural network model: A comparison between multilayer perceptron (MLP) and radial basic function (RBF) algorithms. *Arab J Geosci*. <https://doi.org/10.1007/s12517-012-0610-x>
82. Zhu L, Huang JF (2006) GIS-based logistic regression method for landslide susceptibility mapping in regional scale. *J Zhejiang University: Sci*. <https://doi.org/10.1631/jzus.2006.A2007>

## Figures

**Figure 1**

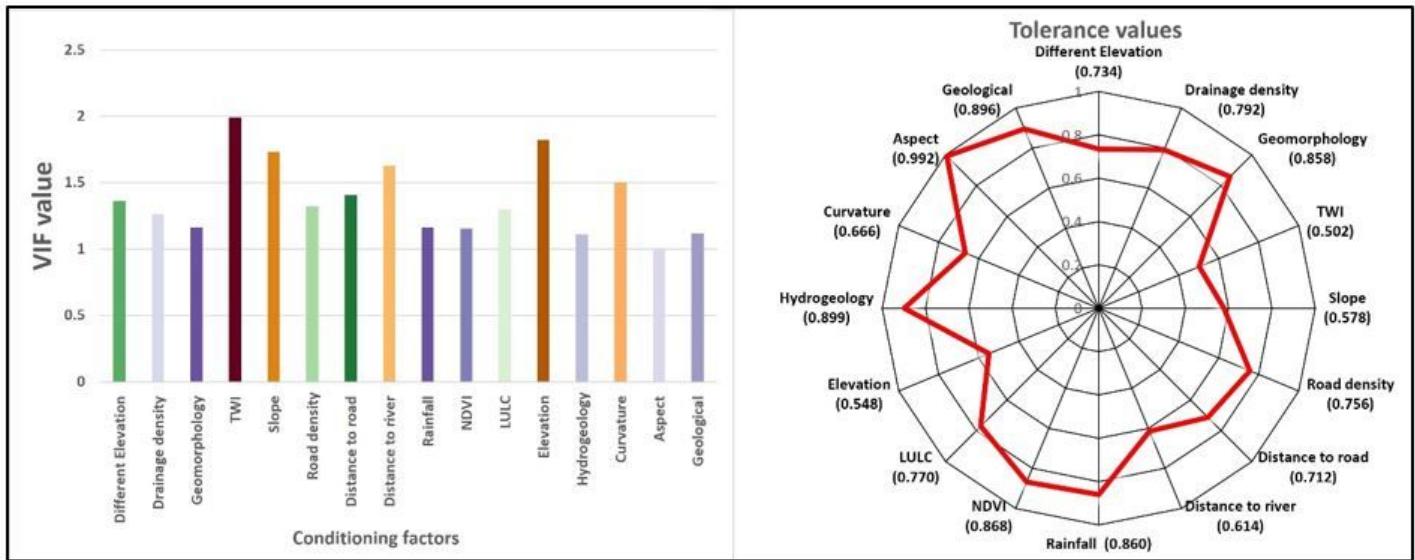
The Son La province and 1,689 landslide locations.



**Figure 2**

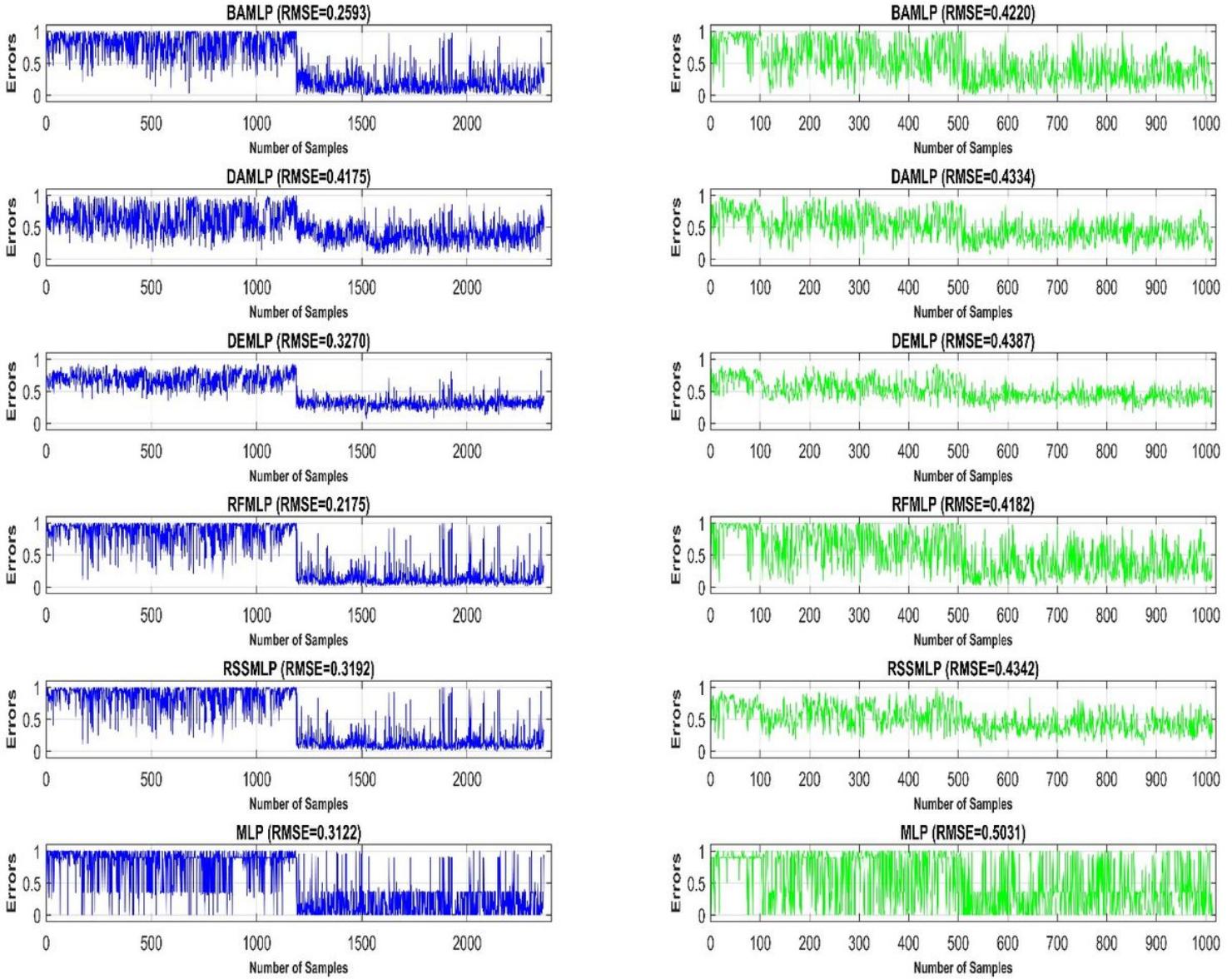
**Figure 3**

Thematic maps of landslide causative factors.



**Figure 4**

VIF and Tolerance values for multicollinearity analysis



**Figure 5**

RMSE of models

**Figure 6**

ROC curve and AUC of (a) the training dataset and (b) the testing dataset.

## **Figure 7**

Landslide susceptibility maps in Son La province: (a) BAMPLP, (b) DAMLP, (c) DEMLP, (d) RFMLP, (e) RSSMLP and (f) MLP.

## **Figure 8**

Areas of landslide susceptibility and percentages by classification.