Vietnam Journal of Earth Sciences, 46(3), 339-359, https://doi.org/10.15625/2615-9783/20706

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

# Assessing the relationship between landslide susceptibility and land cover change using machine learning

Huu Duy Nguyen[1*], Tung Cong Vu[2], Petre Bretcan[3], Alexandru-Ionut Petrisor[4,5,6,7]

*[1]Faculty of Geography, VNU University of Science, Vietnam National University, Hanoi, 334 Nguyen Trai, Thanh Xuan district, Hanoi, Vietnam*
*[2]Center for Educational Testing, Vietnam National University, Ha Noi, 5th floor, HT2 Building, Hoa Lac Campus, Thach That district, Hanoi, Vietnam*
*[3]Valahia University of Targoviste, Faculty of Humanities, Department of Geography, 130004 Targoviste, Romania*
*[4]Doctoral School of Urban Planning, Ion Mincu University of Architecture and Urbanism, Bucharest, Romania, 010014*
*[5]Department of Architecture, Faculty of Architecture and Urban Planning, Technical University of Moldova, 2004 Chisinau, Moldova*
*[6]National Institute for Research and Development in Constructions, Urbanism and Sustainable Spatial Development URBAN-INCERC, 21652 Bucharest, Romania*
*[7]National Institute for Research and Development in Tourism, 50741 Bucharest, Romania*

Received 29 January 2024; Received in revised form 08 March 2024; Accepted 02 May 2024

ABSTRACT

Landslides are natural disasters most frequent in the mountain region of Vietnam, producing critical damage to human lives and assets. Therefore, precisely identifying the landslide occurrence probability within the region is essential in supporting decision-makers or developers in establishing effective strategies for reducing the damage. This study is aimed at developing a methodology based on machine learning, namely Xgboost (XGB), lightGBM, K-Nearest Neighbors (KNN), and Bagging (BA) for assessing the connection of land cover change to landslide susceptibility in Da Lat City, Vietnam. 202 landslide locations and 13 potential drivers became input data for the model. Various statistical indices, namely the root mean square error (RMSE), the area under the curve (AUC), and mean absolute error (MAE), were used to evaluate the proposed models. Our findings indicate that the Xgboost model was better than other models, as shown by the AUC value of 0.94, followed by LightGBM (AUC=0.91), KNN (AUC=0.87), and Bagging (AUC=0.81). In addition, urban areas increased during 2017-2023 from 25 km² to 30 km² in very high landslide susceptibility areas. Our approach can be applied to test the other regions in Vietnam. Our findings might represent a necessary tool for land use planning strategies to reduce damage from natural disasters and landslides.

*Keywords:* machine learning, landslide susceptibility, Da Lat city, Vietnam.

## 1. Introduction

Landslide is a natural hazard that causes damage in the mountain regions, producing human losses and destruction of infrastructure and property, hence causing economic difficulties in countries (Chang, Catani et al., 2023; Merghadi et al., 2020; Nguyen et al.). Landslides occur due to several causes,

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

including heavy precipitation, earthquakes, and human activities. Global landslide data shows that more than 3,867 landslides occurred worldwide from 1995 to 2014, causing 11,689 injuries and 163,658 deaths. The mountain region of Vietnam is strongly affected by landslides, accounting for about 22% of landslides in the world (A. L. Achu et al., 2023). The report of the Meteorological and Hydrological Administration of Vietnam shows that there were 300 landslide and flash floods during 2000-2018, causing about 943 injuries and damages of billions of dollars (Bui et al., 2023; Nguyen, Dang, Bui, & Petrisor, 2023).

The increase in the frequency and magnitude of landslides linked to land use and climate change globally in recent years has been widely studied by researchers. Several studies have focused on evaluating the relationships between changes in land use and different characteristics of the landslide phenomenon, for example, an increased number of landslides due to the transformation of the surface from natural vegetation to agricultural and construction areas. These changes can weaken the roots that hold the soil together. This phenomenon increases the landslide risk. In addition, the construction of infrastructure on slopes can destroy the soil balance, causing the landslide to worsen. All these phenomena combine with more precipitation in the context of climate change, increasing the occurrence of severe landslides. So, good planning can reduce risks, while poor planning can increase the landslide risk. Therefore, assessing the relationship between land cover change and landslides is very necessary, as it can support decision-makers or planners in drafting effective strategies for reducing the effects on human life and the economy.

To carry out these tasks, it is necessary to classify landslide types, evaluate their trends, and construct a hazard zoning map. These maps can help decision-makers determine the regions with the probability of landslide occurrence to propose practical strategies for reducing the damage to human life and property (N. Sharma, Saharia, & Ramana, 2024; Su et al., 2023; Van Phong, Ly, Trinh, Prakash, & Btjvjoes, 2020).

Until now, several qualitative and quantitative studies have aimed to predict landslides. Although the qualitative approach is based mainly on expert knowledge, it has been proven effective with high precision (Ganesh, Vincent, Pathan, & Benitez, 2023; Huang, Peng, Li, Liu, & Zhou, 2023; Khaliq et al., 2023; Nhu, Thai, & Tien, 2023; Xuan et al., 2024). However, this approach is often applied locally and is very expensive and limited in hard-to-reach regions. Beginning in the 19$^{th}$ century, geologists began to recognize the connections of geology, topography, and likelihood of landslide occurrence. Then, they began using quantitative approaches for landslide zoning based on historical landslide events (Zeng et al., 2024). Advances in remote sensing and GIS have provided promising opportunities to determine regions with landslide probability, considering spatial relationships between natural and human factors (Nwazelibe et al., 2023; A. Sharma & Prakash, 2023). Although researchers have widely used remote sensing data, these data have been influenced by clouds and limited by spatial and temporal resolution. Limited spatial resolution can impede the detection of small landslides. Additionally, updating images less frequently can impede real-time tracking changes occurring in the field. Therefore, these methods must be replaced by more powerful ones, which can help decision-makers determine the exact regions at landslide risk to propose effective strategies.

Few data-driven approaches have been recently used to construct landslide susceptibility maps, including statistical engineering and machine learning. The statistical model analyzes the relation between landslide events and their drivers from a spatial perspective. Statistical relationships often used in landslide susceptibility include frequency ratio (Nwazelibe et al., 2023), bivariate statistical analysis (Poddar & Roy, 2024), weights of evidence (WOE) (B. Mandal, Mondal, & Mandal, 2023; Qazi, Singh, Vishwakarma, & Abdo, 2023). Although statistical analysis was shown effective in assessing landslide susceptibility, as mentioned above, the landslide occurs mainly due to slope and rainfall, and, under climate change settings, weather tends to be more erratic and rainy, making the prediction of landslides more difficult. At the same time, the statistical model has been limited in explaining the complex nonlinear connection between landslide events and their drivers. In recent years, machine learning has received attention from researchers to overcome these limitations. Machine learning algorithms include support vector machine (I. Mandal & Pal, 2020), random forest (Kim, Lee, Jung, & Lee, 2018; Taalab, Cheng, & Zhang, 2018), adaboost (Wu et al., 2020), Bagging (T. Zhang et al., 2022), decision tree (Saito, Nakayama, & Matsuyama, 2009; Yeon, Han, & Ryu, 2010). The main benefit of machine learning is its ability to reproduce and quantitatively analyze the impact of different drivers on the development of landslides and their potential to update continuously. (Cao et al., 2023) used five machine learning machines, namely Logistic Regression (LR), Support Vector Machines (SVM), Extreme Gradient Boosting (XGBoost), Random Forest (RF), and Linear Discriminant Analysis (LDA) to unearth the regions with the probability of occurrence of landslides in Western Henan Province. The findings indicated that the Xgboost model was better than the others. (Qasimi, Isazade, Enayat, Nadry, & Majidi, 2023) Three models, namely Maximum Entropy (ME), Generalized Linear Model (GLM), and Random Forest (RF), were applied to construct a landslide susceptibility map for Badakhshan province, Afghanistan. The results showed that RF outperformed the other two models. (Rai, Pandey, Sharma, & Sharma, 2024) compared seven machine learning models, namely Fisher discriminant analysis (FDA), boosted regression tree (BRT), multivariate adaptive regression splines (MARS), generalized linear model (GLM), random forest (RF), model-architect analysis (MDA), and the support vector machine (SVM) to construct a landslide susceptibility map for the Bhilangana Basin, Garhwal Himalaya. The findings show that the random forest model had a better accuracy than the others. The literature review shows no consensus concerning the most suitable type of model for landslide susceptibility. Even a slight improvement in the accuracy of landslide susceptibility can have a significant impact on constructing landslide susceptibility maps and change the characteristics of the landslide susceptibility level distribution. Therefore, the exploration and comparison of different models are needed to determine the most appropriate models to construct landslide susceptibility maps, which can felp decision-makers designing effective strategies for diminishing landslide-related damages.

This study aims to build a machine learning methodology using Bagging, KNN, LighGBM, and Xgboost to assess the connection between landslide susceptibility and land use change in Da Lat city, Lam Dong province, Vietnam. This study is the first to assess the relationships between land use change and landslide susceptibility in Da Lat city. In recent years, the rate of land use has changed; for example, the increase in construction areas has increased rapidly, increasing the flood risk. This study's results can help decision-makers make land use planning more sustainable. Moreover, the approach in this research can be replicated in other regions if data is available.

## 2. Study area and material

### *2.1. Study Area*

Da Lat city is located on the Lam Vien plateau, northeast of Lam Dong province, with a natural area of 39,105 hectares (Fig. 1). Da Lat's terrain belongs to the plateau type, with an average elevation of 1,520 m above sea level, and is divided into three main categories, i.e., high mountains, low hills, and valleys, mainly characterized by a strong cleavage in the terrain. The north and northwest areas of Da Lat are divided by the Lang Biang mountain range, while the east and southeast gradually decrease in altitude to the Da Nhim valley. In the West and Southwest, the terrain gradually decreases in altitude to the Di Linh plateau.

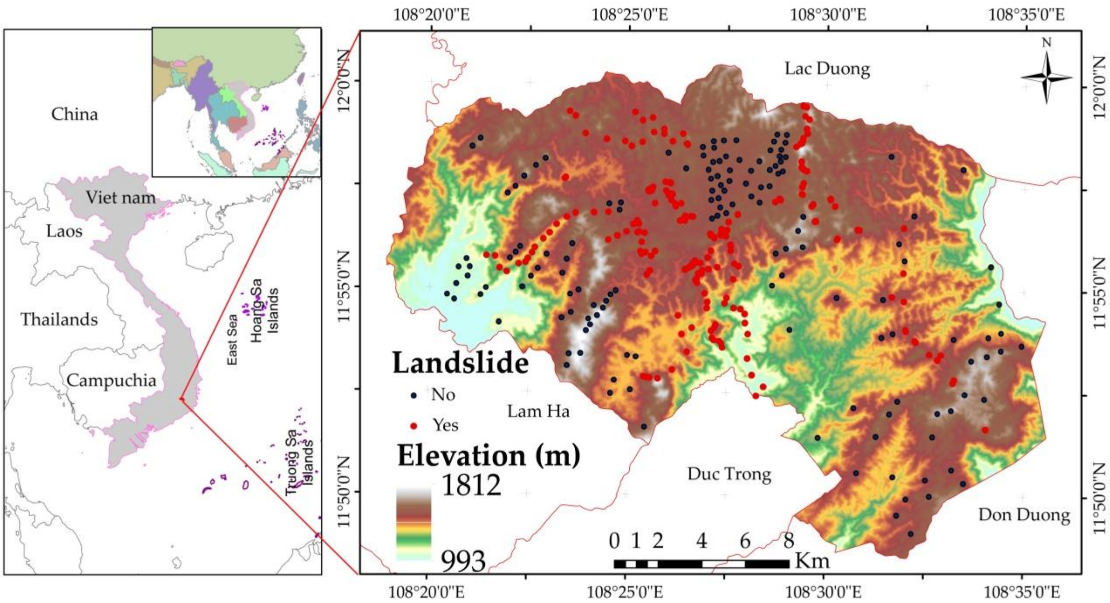


*Figure 1*. Location of Dat Lat city in Vietnam

Da Lat is located in the tropicalmonsoon area near the equator. However, its climate is mainly influenced by altitude and natural terrain. The climate in Da Lat has two main seasons: the rainy one spans from April to October every year, and the dry one from October of the previous year to April of the following year. The rainfall in Da Lat has an annual average of about 1800 mm, with a higher intensity in August and September every year. The dry season, where water becomes scarce, occurs in December, January, and February (Viet Nam Meteorological and Hydrological Administration).

Da Lat's geology comprises many raw materials, such as igneous rocks, sedimentary rocks, and metamorphic rocks. Typical soil types in Da Lat include red-yellow fertility soil (Fs), red-yellow fertility soil (Fa), grey-yellow humus soil (Fha), yellow-brown fertility soil (Fda), red-brown fertility soil developed in basalt (Fk), purple-brown fertility soil developed on metamorphic rocks (Ft), yellow-red soil developed on metamorphic rocks (Fj), alluvial soil (P), and conglomerate slope soil (Dt) (Lam Dong province Department of Natural Resources and Environment).

Da Lat city is one of Vietnam's landslide hotspots. Each year, landslides cause significant socioeconomic damage. For example, the landslide on 29 June 2023 left 2 people dead and 5 injured. Therefore, developing maps of landslide susceptibility can assist local authorities and policymakers in identifying areas highly susceptible to significant landslides and establish measures for adequate protection.

### *2.2. Materials*

#### *2.2.1. Landslide Inventory*

The landslide inventory plays an essential role in building the landslide susceptibility model. The landslide data set included the position, type, and size of the landslide event in history (Goetz, Brenning, Petschko, & Leopold, 2015; Kavzoglu, Colkesen, & Sahin, 2019). Ultimately, 202 landslide events from 2017 to 2020 were obtained from the Vietnam Institute of Geosciences and Mineral Resources, and 50 landslide events were collected from the field mission in 2021 and 2022.

In addition, this study used binary classification for the landslide susceptibility model. Therefore, non-landslide points must be collected. Several studies have pointed out that several non-landslides similar to the landslide points can increase the accuracy of models (Nguyen et al.; Viet Du et al., 2023). Therefore, 252 non-landslide points were selected randomly from regions never affected by the landslide, such as the low slope and elevation region.

In the end, 504 landslide and non-landslide points were assigned to two sets: the first 70% of the data was utilized to train the four landslide susceptibility models, and the other 30% to validate them.

#### *2.2.2. Conditioning factor*

The selection of drivers is an indispensable task when using machine learning to model landslide susceptibility because these factors have complex relationships with landslide events in history (Zhou et al., 2018). In this study, 13 drivers, i.e., elevation, aspect, slope, curvature, distance to the river, road distance, land use, NDVI (Normalized difference vegetation index), NDBI (Normalized Difference Built-up Index), NDMI (Normalized Difference Moisture Index), rainfall, soil type, lithology were chosen to construct the landslide susceptibility model. These drivers are similar to those used by previous studies (Fig. 2).

Elevation, aspect, curvature, and slope were obtained from a DEM (Digital Elevation Model) with a 10m resolution constructed from the topography map (available at the Ministry of Natural Resources and Environment) with a scale of 1/50,000. Distances to road and river were extracted from the topography map scaled 1/50,000. Soil type and lithology were obtained from the Ministry of Natural Resources and Environment. 2021 land use was downloaded from https://www.arcgis.com/apps/instant/media/index.html?appid=fc92d38533d440078f17678ebc20e8e2&fbclid=IwAR0V3ZEdUqhn79qN_JNPMtswxWfi2dE1_Gj-1ZD_XcN7oPyGMSn3- scE9KY. NDVI, NDBI, and NDMI were extracted from a Sentinel 2A image from September 2021 (available at https://dataspace.copernicus.eu). 2021 annual rainfall was retrieved from https://chrsdata.eng.uci.edu/.

Using machine learning, elevation is an essential driver for identifying regions with the probability of landslide occurrence. It is controlled by geology, lythology, and precipitation. Elevation influences the stability of slopes. It has been extensively used to study landslide susceptibility (J. Zhang et al., 2023).

The slope has an essential role in the landslide susceptibility model. The slope may significantly affect soil stability, increasing landslide likelihood in a specific region (A. Achu et al., 2023; Chang, Huang, et al., 2023; D. Sun, Q. Gu, et al., 2023).

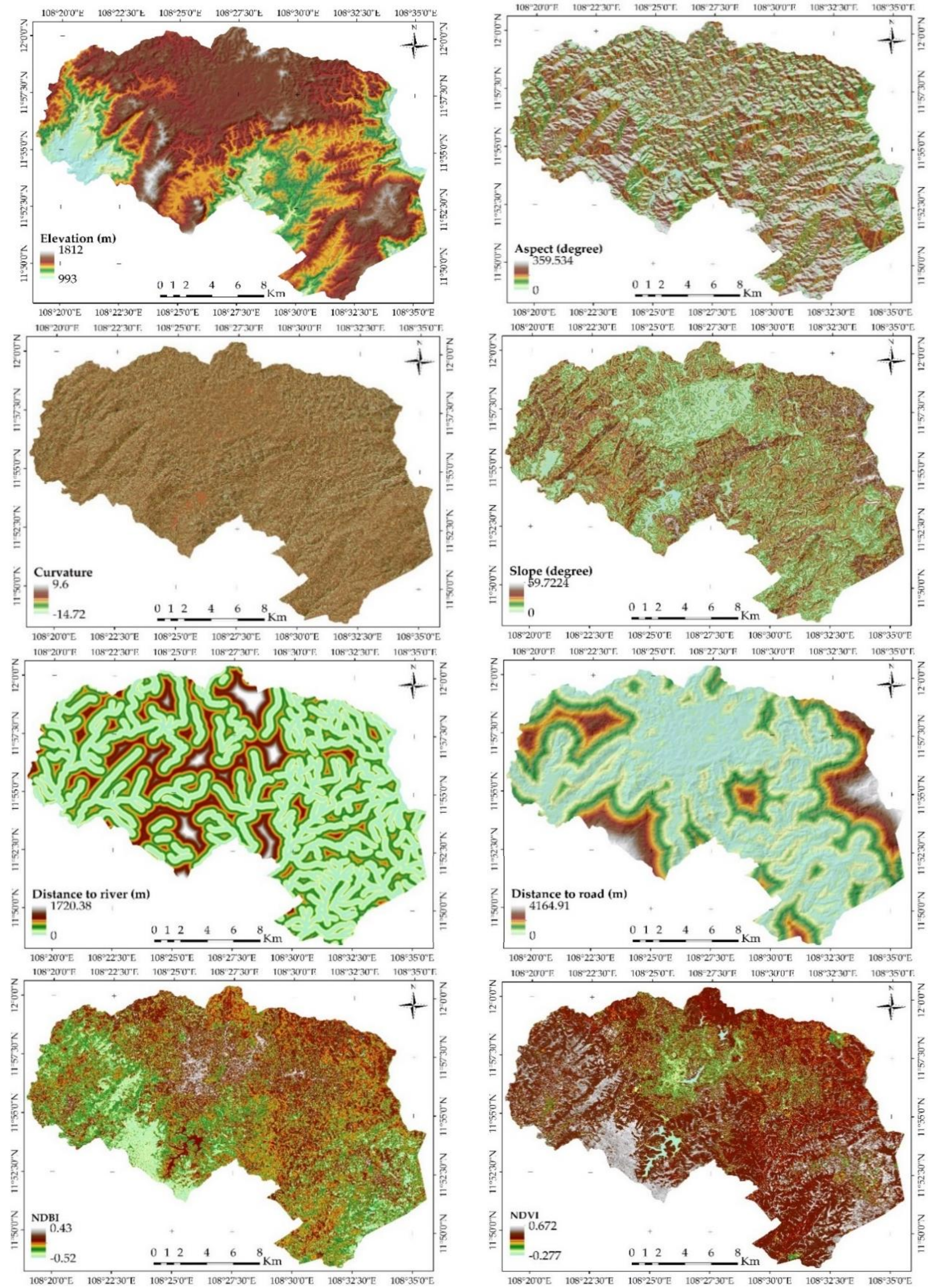


*Figure 2*. Conditioning factors for landslide susceptibility in the Da Lat city

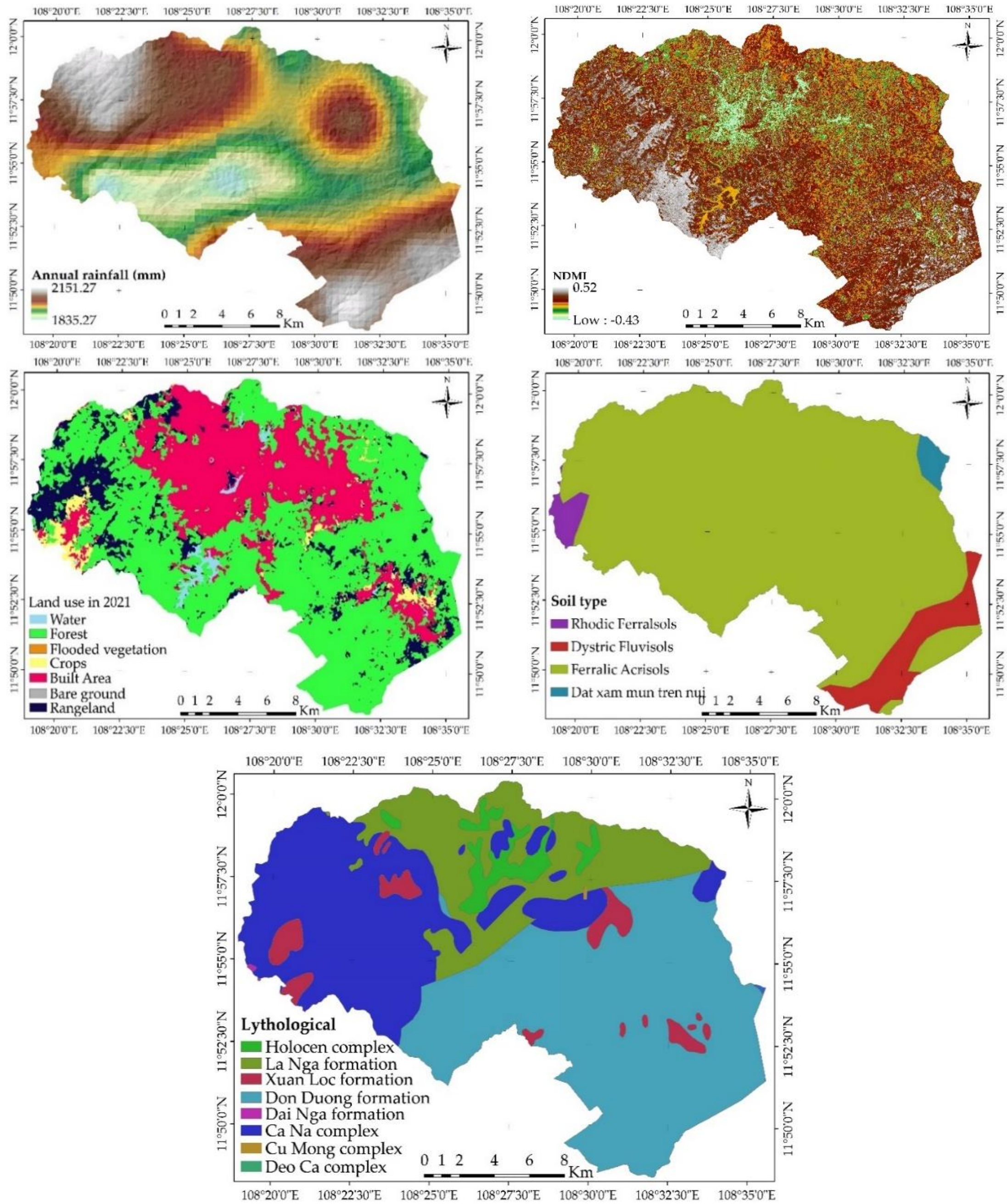


*Figure 2*. Cont.

Curvature is considered an indicator of the shape of the surface and influences landslide susceptibility in several ways. For example, the curvature has essential effects on the stability of the ground, which is linked to the modification of the distribution of stresses and shear stresses. Furthermore, the low curvature value can cause water to concentrate, significantly impacting soil stability (A. Achu et al., 2023).

Aspect, i.e., slope orientation, is essential in assessing landslide susceptibility because it affects precipitation, water flow direction, soil moisture, and vegetation distribution. The orientation of the slope exposed to prevailing wind and precipitation may increase precipitation and be more prone to moisture, increasing the likelihood of landslide occurrence. Moreover, the aspect influences the distribution of vegetation on the slope (H. Sun et al., 2023).

Land use/land cover is essential in analyzing the probability of landslide occurrence in a region as it is directly linked to human activities. Activities like urban expansion or infrastructure construction significantly trigger landslides. Such activities require cutting or excavating slopes to build infrastructure that modifies natural characteristics and disrupts soil stability (Tyagi, Tiwari, & James, 2023).

NDVI measures vegetation density in a region and is considered an indispensable driver in assessing landslide susceptibility. Vegetation is essential for soil stability because it reduces erosion and consolidates soils (Niraj, Singh, & Shukla, 2023). Generally, NDVI values range from – 1 to 1; the higher the vegetation value, the denser the curvature.

NDBI is the density of construction in a region. It is widely used to pinpoint regions with the probability of landslide occurrence because construction activities can cause soil instability, causing landslides (F. Huang et al., 2023).

NDMI is a surface soil moisture indicator. The state of soil moisture can change depending on several drivers, e.g., precipitation, topography, and vegetation density. Soil moisture can cause soil saturation. Saturated soil can lose cohesion, making it more susceptible to ground movement. Heavy precipitation on saturated soil can cause a landslide (Fatemi Aghda, Bagheri, & Razifard, 2018).

The soil type is another important factor in assessing the susceptibility because it is directly related to compaction. Areas containing high-density soils often experience fewer landslides than areas containing low-density soils (Lee & Min, 2001).

Rainfall is considered a factor triggering landslides in a region. Heavy precipitation causes the soil to become saturated, making it easier for the land to move. In addition, heavy precipitation can lead to increased runoff, eroding the soil. Additionally, precipitation can penetrate areas of weak soil, lubricating potential sliding surfaces and increasing the likelihood of landslides (Dahal et al., 2008).

The road distance is a driver for the landslide susceptibility model. Road construction must cut or excavate slopes, influencing the natural features of the slope and causing the probability of landslide occurrence to increase (Tien Bui, Ho, Revhaug, Pradhan, & Nguyen, 2014).

The distance from the river plays an essential role in assessing landslide susceptibility, as the river is an important factor in the movement of slope masses causing landslides (X. Sun et al., 2020).

Lithology plays a vital role in the probability of landslide occurrence as it is directly linked to rocks' hardness and weather resistance. More complicated rock areas have less risk of landslides than softer rock areas (Vakhshoori & Zare, 2016).

## 3. Methodology

The methodology used to identify the region with the probability of landslide occurrence in our research had four main steps: (i) preparing the input data of the machine learning model; (ii) establishing the machine learning model; (iii) evaluating the accuracy of the model and (iv) analyzing the landslide susceptibility map (Fig. 3).

(i) The input data of the machine learning model includes the landslide inventory map and its drivers. This study used 252 landslide and 252 non-landslide points and 13 drivers as input data for the model. It must be stressed that drivers were measured with unit differences; therefore, we normalized them to a range from 0 to 1 before using them. In this study, we used the max-min normalization.

(ii) Establishing the machine learning model: in this research, four models, namely Xgb, LightGBM, KNN, and BA, were used to construct the landslide susceptibility map. Their accuracy depends on optimizing the parameters. In this research, the trial-and-error method was utilized to adjust the parameters. The results showed that the parameters were: for the XGB model, n_estimators = 100, max_depth = 3, learning_rate = 0.05, subsample = 0.8, colsample_bytree = 0.8; for LightGBM, random_state = None, max_depth = 5, learning_rate =0.05, subsample=0.8, colsample_bytree=0.8; for LightGBM, random_state=None, max_depth=5, learning_rate=0.02, feature_fraction=0.8; for KNN, n_neighbors=4, weights="uniform," algorithm="kd_tree"; and for BA, n_estimators=100, max_depth=2, learning_rate=0.01, random_state=None. The process of constructing the machine learning model was implemented in the Python platform.

(iii) Evaluating the precision of the machine learning model was based on several statistical indices, i.e., AUC, RMSE, MAE, and $R^2$.

(iv) After evaluating the models, these models were utilized to construct the landslide susceptibility map. The output value of the model is the landslide susceptibility index, ranging from 0 to 1. These indices were assigned to five classes (ranging from very low to very high) through the natural break method.

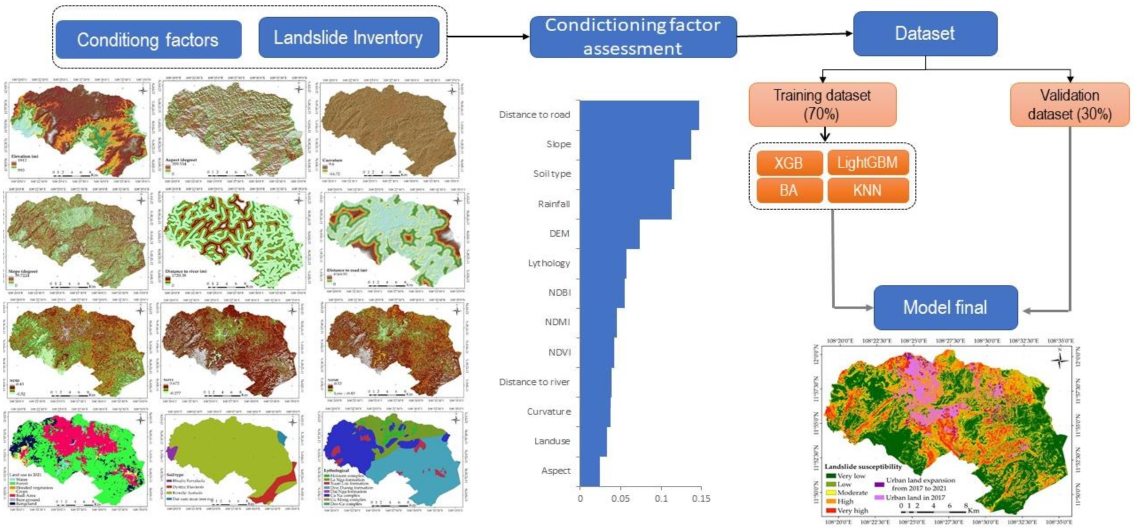


*Figure 3*. Methodology used for the landslide susceptibility in the Da Lat city

### *3.1. Bagging*

Bagging is considered a method of assembly using the Bootstrap technique, first presented in (Breiman, 1996). The bootstrap approach randomly selects a replaced sample to generate many samples, creating the training data set. Each generated subset is utilized to build decision trees, which are merged into the final model. This model reduces the classification accuracy by decreasing the variance of classification error (Hong et al., 2018; T. Zhang et al., 2022). Bagging functions require four main steps: i) construction of the original algorithms and training data set; ii) the precision of original algorithms is low; iii) original algorithms are repeated several times to obtain a prediction sequence; iv) final model accuracies are improved by using results from multiple models (Wu et al., 2020). Details of the LighGBM equation are in (Pham, Tien Bui, & Prakash, 2018). In our research, the bagging model was utilized to build the landslide susceptibility model for Da Lat City.

### *3.2. K-Nearest Neighbors*

KNN is a non-parametric supervised learning discriminant that utilizes proximity to make classifications or predictions on the clustering of an individual data point. This algorithm can solve the classification and regression problem (Abu El-Magd, Ali, & Pham, 2021). KNN presumes that similar points can be found next to each other with four main steps (Liu, 2023):

(i) Construction of the set of already classified data (called dataset), a distance d and an integer k.

(ii) KNN computes the distance between all already classified and newly entered data.

(iii) KNN then extracts the k already classified data "closest" to the new data entered; that is to say, the data already classified has the minor distance d with the new data entered.

(iv) The algorithm finally chooses which family the new data belongs to by searching for the majority family among the k-identified data.

Details of the LighGBM equation are in (Abu El-Magd et al., 2021).

### *3.3. LighGBM*

LightGBM is an ensemble gradient optimization method relying on gradient boosting and decision trees. LightGBM can be utilized to solve the classification and regression problem. It builds decision trees (D. Sun, Wu, Wen, & Gu, 2023). Subsequent trees were constructed by adjusting the residuals from the previous tree to improve the model. The last added tree aggregates the results of each step, producing a robust model. Decision trees in LightGBM use leaf growth strategies, which means that given a specific condition, a single leaf is split, depending on the gain. By identifying leaves containing considerable decomposition information throughout all current leaves, the classification accuracy of the approach is increased by decomposition. Leaf-aware trees can sometimes cause overfitting, especially with smaller datasets. Limiting shaft depth can help prevent overfitting (Iban & Bilgilioglu, 2023; Zeng et al., 2024).

The LightGBM algorithm utilizes a histogram-based method that splits data into groups using their histogram. Instead of each data point, these groups are used to iterate, calculate the gain, and split the data. Another distinctive feature of the LightGBM algorithm is the clustering of exclusive entities. This technique combines exclusive features to perform dimensional reduction, making the algorithm faster and more efficient (Saber et al., 2022; Ye, Yu, Chen, Liu, & Ye, 2022). Details of the LighGBM equation are in (D. Sun, D. Chen, et al., 2023).

### *3.4. XgBoost*

XGBoost is a supervised machine training method for classification and regression. This method relies on decision trees and improves the gradient-boosting algorithm (Sahin, 2020). In the activity procedure, XGBoost is an assembly of decision trees (weak learners), which predicts residuals and corrects the errors of previous decision trees. Like the gradient boosting and adaptive models, XGBoost uses this " pruning " method (Nguyen, Van, et al., 2023). More precisely, weak learners are corrected until they entirely play their role. When these trees do not perform well, they are deleted. Details of the LighGBM equation are in (Al-Najjar et al., 2021).

### *3.5. Model assessment*

In our research, several statistical indices were utilized to evaluate the accuracy of prediction models, i.e., RMSE, MAE, ROC, and AUC. These indices have been used in many previous studies (Luu et al., 2022; Nguyen et al., 2024; Yang, Liu, Huang, Huang, & Wang, 2023).

ROC was computed by the actual positive rate on the Y axis and the false positive rate on the X axis. The areas under the ROC curve represent the model accuracy of the AUC value, ranging from 0 to 1.

$$AUC = \frac{\sum TP + \sum TN}{P + N}$$

TP and TN are the rates of pixels correctly classified as landslide/non-landslide, while P and N are the pixel rates of landslide/non-landslide.

RMSE and MAE are the errors between the model prediction value and the observation one.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_{predicted} - Y_{observed})^2}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|Y_{predicted} - Y_{observed}|$$

## 4. Results

### *4.1. Landslide conditioning factor assessment*

To increase the accuracy of the landslide susceptibility model, it is crucial to identify landslide points and their causes accurately. In our research, a few field missions collected and verified landslide points. While the conditioning factors were similar to previous studies, we used Random Forest to find the critical factors and eliminate non-useful ones. Given that the model is exposed to redundant data, models can modify its parameters in ways that are too specific for individual examples instead of generalizing from the underlying samples. Additionally, redundancy data does not provide information for the model, which can reduce diversity in the data set. A good model must be able to generalize from several different situations, and data redundancy can reduce this ability. The Fig. 4 shows the importance of drivers. It can be seen that all drivers help identify the region with the probability of landslide occurrence. Among them, the most critical factors are distance to the road, slope, soil type, and rainfall. In recent years, urban expansion has taken place more and more rapidly in the province of Dat Lat, which has led to the construction of several infrastructures, e.g., roads. The construction of roads leads to cutting and evacuating the slope, causing modification of the original state and instability of slopes. The slope was second because, as mentioned in the study area section, the city of Da Lat is situated in the south of Vietnam's central highlands, so the slope is very high. This is considered an essential cause for producing the landslide. In addition, the landslide frequency also depends on the compactness of the soil.

The ferralit soil group accounts for over 90% of the Da Lat city area. This group of soils often has poor drainage capacity. The ground cannot absorb water in heavy rain, which easily causes landslides and floods. Additionally, feralite soil has a weak structure, so it becomes unstable and easily eroded in direct contact with water. In the study region, the average precipitation in Da Lat city is approximately 1,800 mm/year, combined with its geographical location in high mountains and hilly areas. The hilly terrain is often less drained than flat terrain, and in the event of heavy rain, the ground can quickly become saturated with water, increasing the landslide risk. Additionally, in recent years, cutting virgin forests to make room for urban development resulted in the soil losing its inherent compactness and becoming, in heavy rains, less stable and susceptible to landslides.

In this study, land use, aspect, and curvature are less critical in the probability of a landslide occurrence. This r uses machine learning, so the statistical relationship is essential. Landslide occurrence in Dat Lat town depends more on the distance to the road, slope, soil type, and rainfall and less on aspect, land use, and curvature (Fig. 4).

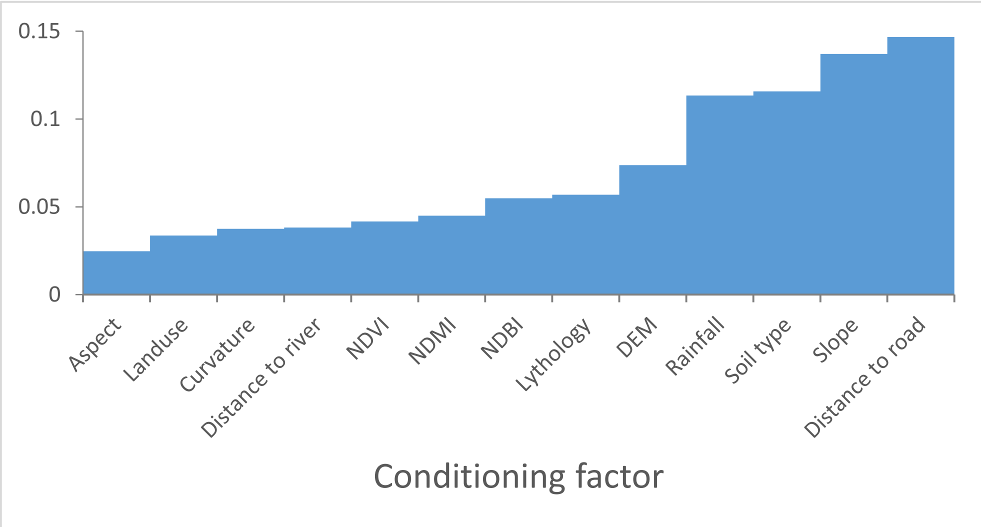


*Figure 4*. Importance of landslide conditioning factor using Random forest

### *4.2. Assessment and comparison of model performance*

Landslide susceptibility model performance was evaluated using the ROC curve on all training and validation data (Fig. 5). It can be seen that all proposed models are performing well in evaluating landslides in Da Lat city. Among them, the XGB model was better than other models with the AUC value of 0.99, followed by lightGBM with 0.97, BA with 0.94, and KNN with 0.92, respectively, for training data. The XGB model performed even better for validation data, with an AUC value of 0.94, followed by LightGBM with 0.91, KNN with A0.87, and BA with 0.81.

The XGB model was generally better than the other models in terms of training and validation data.

Apart from the ROC index, this study used the MAE and RMSE indices to assess the performance of the landslide susceptibility model. For the RMSE index and training dataset, the XGB model is better than other models, with RMSE=0.25, followed by BA (0.29), LighGBM (0.3), and KNN (0.32). For the validation dataset, the XGB is the best, with RMSE=0.32, followed by LighGBM (0.35), KNN (0.37), and BA (0.43). For the MAE index, the XGB model was the most accurate, with MAE=0.2 for the training dataset and 0.27 for the validation one, followed by LightGBM (MAE=0.26 for the training data and 0.31 for the validation data), BA (MAE=0.27 for training data and 0.35 for validation data), KNN (MAE=0.32 for training data, and 0.36) (Table 1).

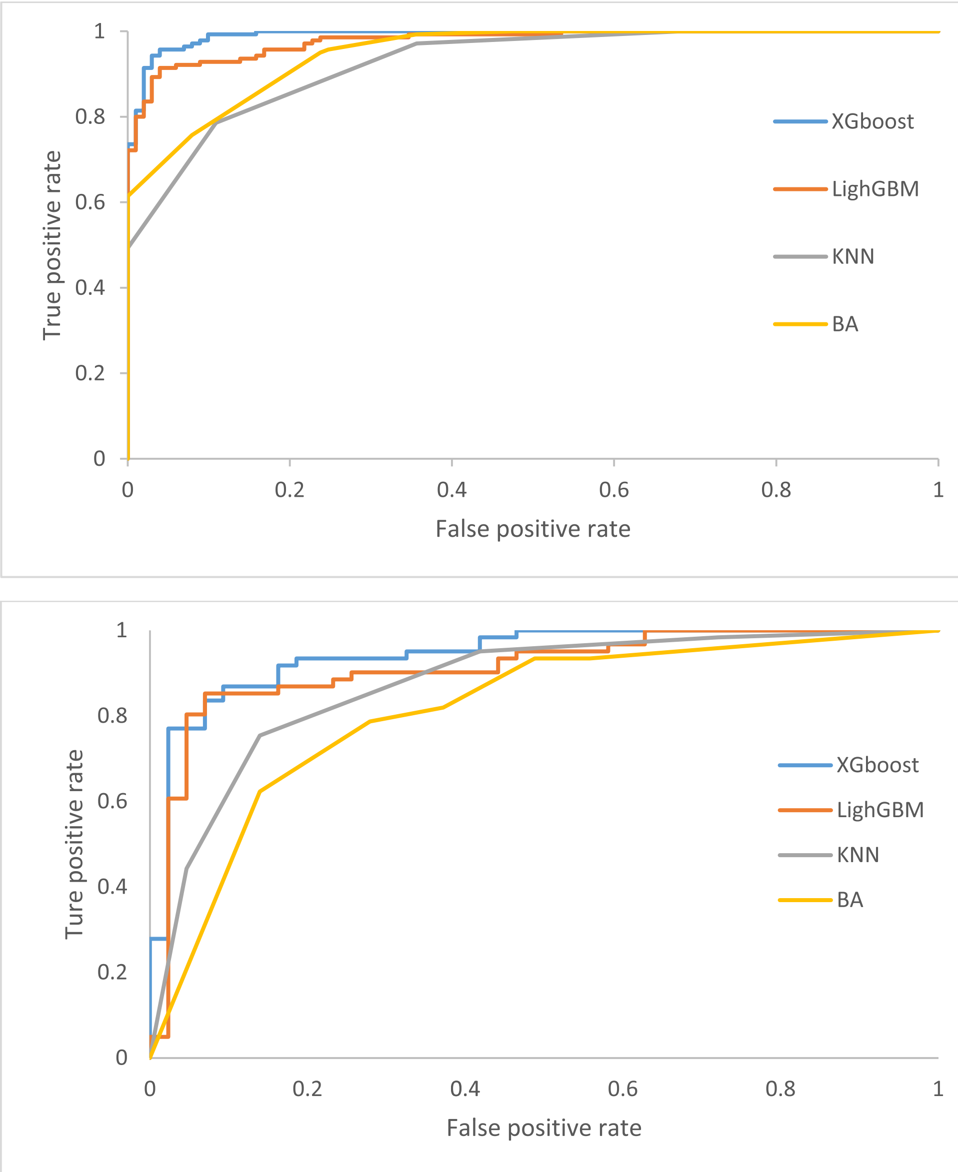


*Figure 5*. Performance of the models proposed for landslide susceptibility for training dataset (top) and validation dataset (dow) using AUC

*Table 1*. Performance and comparison of XGB, LightGBM, KNN, and BA models

| | Training dataset | | | Validating dataset | | |
|---|---|---|---|---|---|---|
| | RMSE | MAE | AUC | RMSE | MAE | AUC |
| XGB | 0.25 | 0.208 | 0.991 | 0.325 | 0.272 | 0.943 |
| LighGBM | 0.3 | 0.26 | 0.977 | 0.355 | 0.311 | 0.911 |
| KNN | 0.32 | 0.329 | 0.925 | 0.37 | 0.364 | 0.873 |
| BA | 0.29 | 0.27 | 0.94 | 0.43 | 0.35 | 0.81 |

### *4.3. Landslide susceptibility map*

After validation, the four models generated the landslide susceptibility map in Da Lat City. The landslide susceptibility map construction process provided the models with pixels covering the entire study area. Each pixel was associated with 14 drivers. The output corresponds to the landslide susceptibility index, varying from 0 to 1. The value of landslide susceptibility was divided into five classes (shallow, low, moderate, high, and very high) via the natural break method. Figure 6 and Table 2 present the landslide susceptibility map produced by Xgboost, LightGBM, Bagging, and KNN.

*Table 2.* Distribution of landslide susceptibility class in Da Lat City, Vietnam

| | Very low ($Km^2$) | Low ($Km^2$) | Moderate ($Km^2$) | High ($Km^2$) | Very high ($Km^2$) |
|---|---|---|---|---|---|
| XGB | 137.6675 | 47.3962 | 36.2654 | 115.4446 | 55.7419 |
| LightGBM | 90.7016 | 87.0316 | 59.2656 | 127.9615 | 27.5518 |
| BA | 47.5067 | 48.0201 | 136.8979 | 138.6492 | 21.4417 |
| KNN | 69.3972 | 74.0694 | 92.5696 | 93.7374 | 62.742 |

For the Xgboost model, approximately 137 km² were found in the shallow category, 47 km² in the low, 36 km² in the moderate, 115 km² in the high, and 55 km² in the very high. For the LightGBM model, approximately 90 km² of the study area was found in the shallow landslide susceptibility zone, 87 km² in the low landslide one, 59 km² in the moderate one, 127 km² in the high one, and 27 km² in the very high one. For the Bagging model, 47 km² of the surface of the study area were found in the shallow landslide susceptibility zone, 48 km² in the low one, 136 km² in the moderate one, 138 km² in the high one, and 21 km² in the very high one. For the KNN model, approximately 39 km² of the study area lies in the shallow landslide susceptibility zone, 74 km² in the low, 92 km² in the moderate, 93 km² in the high, and 62 km² in the very high.

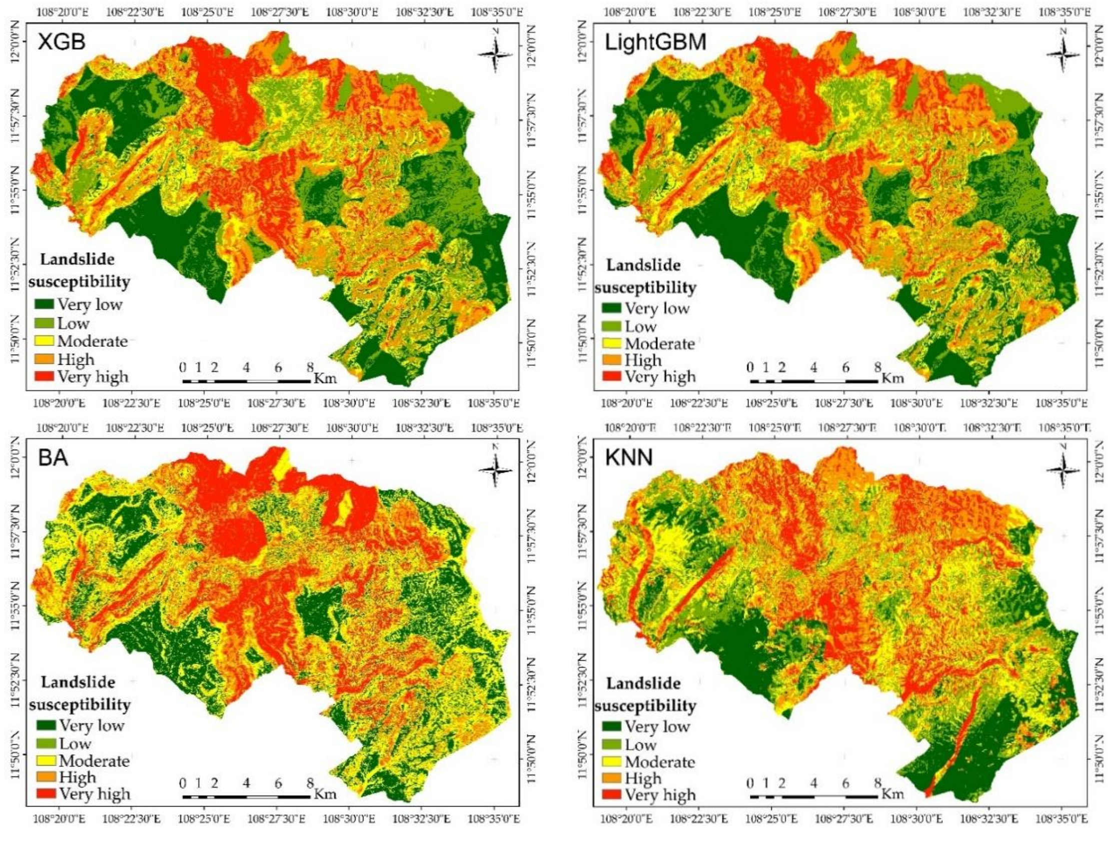


*Figure 6.* Landslide susceptibility in the Da Lat city product by XGB, LightGBM, BA, and KNN

### *4.4. Urban planning in the landslide area*

Among all the models proposed in this research, the XGB model exhibited higher accuracy than others. Therefore, this model was utilized to assess the urban development in the landslide-prone area. The results revealed that approximately 25 km² of the urban area was found in the very high landslide zone in 2017, increasing to 30 km² in 2022. An essential part of the forest has been replaced by urban areas for developing tourism, leading to slope instability and significant landslides in the city of Da Lat (Fig. 7).

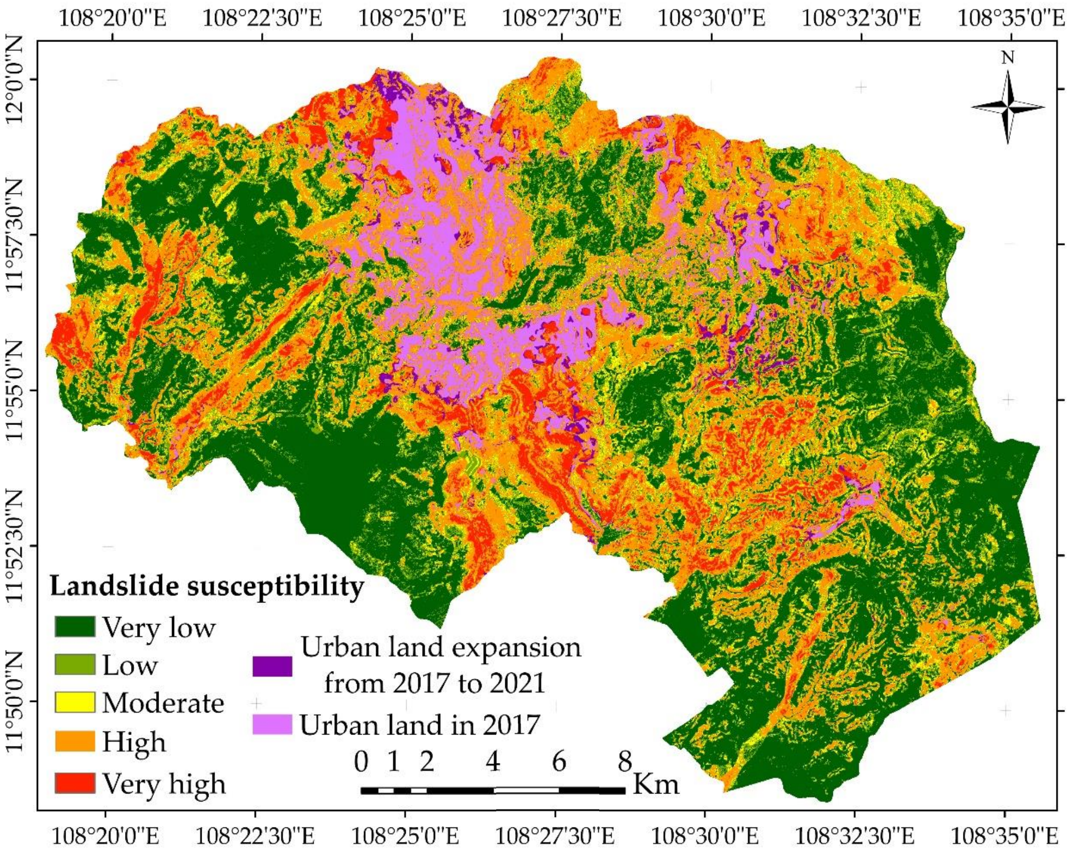


*Figure 7*. Urban land in the very high area of Landslide susceptibility

### 5. Discussions

Landslides are among the most dangerous natural disasters, causing significant damage to human life and property. Thus, appropriate measures must be immediately proposed to decrease the damage (Hong, 2024; Nhu, Bui, My, Vuong, & Duc, 2022; D. Sun, Q. Gu, et al., 2023). Therefore, applying practical methods to identify areas of the probability of landslide occurrence is a crucial factor for landslide risk assessment and management (Ghasemain et al., 2020; N. Sharma et al., 2024). Through spatial mapping, regions of high and very high landslide occurrence can be identified, and appropriate measures can be applied to reduce or prevent damage caused by landslides. Various models and methods were developed to predict landslides' local and regional spatial distribution. Constructing the landslide susceptibility map with accuracy has received attention from researchers. Recently, few studies have utilized machine learning to construct the landslide susceptibility map. Therefore, this study aimed to develop a machine learning-based approach using XGB, LightGBM, BA, and KNN to construct the landslide map in Da Lat City. Our research findings are considered an important tool to help planners or decision-makers determine the appropriate regions for sustainable land development.

Machine learning has several advantages for building the landslide susceptibility map. Machine learning algorithms can solve geospatial issues and integrate information on topography, geology, and land use. In addition, the ability of machine learning models to detect complex and nonlinear connections between different causes can increase the model's performance. The flexibility of machine learning allows the model to be adjusted and improved when new data is added, ensuring the ability to adapt to environmental changes (Chang, Huang, et al., 2023; N. Sharma et al., 2024). To build the landslide susceptibility map, our research used four machine learning models, i.e., Xgb, LightGBM, KNN, and BA. These models are popular algorithms that are widely used in different fields. Among them, the XGB model performed better than others because it combines several weak learners using the sequential method to improve the observation ability repetitively. This method helps reduce high biases.

Additionally, the XGB model can easily be understood and interpreted. It does not need any preprocessing because it has built-in routines allowing for missing data to be processed (Hajek, Abedin, & Sivarajah, 2023; W. Zhang, He, Wang, Liu, & Meng, 2023). The lightGBM model was second, with an AUC value of 0.92, because it can process categorical data without requiring additional preprocessing, simplifying the workflow. Additionally, this model performs well on the large dataset (Qiu, Wang, & Li, 2023; T. Zhang et al., 2023). The KNN and Bagging model had lower performance because the algorithm becomes slower as the number of observations and independent variables increases. Additionally, model accuracy depends on data quality (Rezapour, Jamali, & Bahmanyar, 2023; Ukey et al., 2023). Although the BA model is straightforward, its computing time was high. In addition, this algorithm is difficult to interpret and improve (Le Minh et al., 2023).

Several previous studies have highlighted the dependence of landslide susceptibility models on the conditioning factors utilized to construct the model. Selecting these factors accurately can improve model performance. However, there are no universal guidelines for selecting these drivers. This study's drivers were related to environmental, hydrology, climate, and anthropogenic activities. 12 conditioning factors were selected to use as machine learning model input data. This selection depends on the availability of data. Ultimately, all these factors were rated as necessary using the RF technique.

Our research's landslide susceptibility map in Da Lat City aligns with previous ones. Therefore, the landslide susceptibility map in our research can be used to evaluate the relationship between land use change and flood susceptibility in Da Lat City.

Landslide susceptibility and land use planning are closely tied and received attention from researchers and local authorities, particularly those dealing with urban planning and natural disaster management. Land use planning includes the construction of infrastructure urbanization, which can strongly influence slope stability, increasing the landslide probability. For example, rapid urban growth can lead to the transformation of forest surface agriculture to construction surface, which weakens the resistance of the soil to mass movements (Petrişor, 2015). Consequently, land use planners must account for the risks of natural hazard susceptibility, such as landslide susceptibility, to ensure food security for residents. Therefore, the theoretical framework in this study can help decision-makers and local authorities integrate landslide risk into land use planning.

In a nutshell, this research highlighted that Da Lat, particularly in the mountain region of Vietnam, often faces significant landslide-related challenges. However, risk management policies and technical state-of-the-art solutions for climate change and urban development are lacking. Therefore, our study can help local authorities or developers establish effective measures to lower the effects of severe landslides that are predicted to occur soon.

## 6. Conclusions

A landslide is a natural disaster with the highest frequency in mountain regions worldwide, particularly in Vietnam. This study aimed to develop a methodology based on machine learning, i.e., Xgb, lightGBM, KNN, and BA, to identify the region with the probability of landslide occurrence in the city of Da Lat in Vietnam. The following points can summarize the results:

All models proposed in our research performed well in building the landslide susceptibility map. Among them, the XGB model was the best, with an AUC value of 0.94, followed by LightGBM with 0.91, KNN with 0.87, and Bagging with 0.81, respectively. Our findings underline the potential of machine learning to identify areas with the probability of landslide occurrence in mountain regions.

In Da Lat city, about 150-170 km² were in the high and very high landslide susceptibility zone. Identifying the regions of high and very high landslide susceptibility can help decision-makers and developers establish effective measures to decrease damage to human life and property.

Although this study successfully constructed the landslide susceptibility map, it still has a few limitations. For example, several studies have highlighted that landslide is strongly connected to climate change and urban expansion, so future research must evaluate their impacts on landslide susceptibility. Moreover, a DEM with a resolution of 10 m was constructed utilizing the topography map scaled 1:50,000. However, this DEM only presents the terrain surface, while the DEM is produced in UAV, which can present the most surface detail, such as vegetation and infrastructure. However, this technique is costly, and its application is limited to vast regions. Finally, non-landslide points were randomly selected from within the study region. Inaccurate selection of points other than landslides also affects the model performance. However, there are currently no specific guidelines on selecting non-landslide points.

The city of Da Lat is often affected by landslides. This phenomenon is increasing seriously under climate and land use change settings. Therefore, our findings are significant in helping decision-makers and planners implement effective measures for sustainable land use planning to decrease damage to human lives and property.

## References

Abu El-Magd S.A., Ali S.A., Pham Q.B., 2021. Spatial modeling and susceptibility zonation of landslides using random forest, naïve bayes and K-nearest neighbor in a complicated terrain. Earth Science Informatics, 14(3), 1227–1243.

Achu A., Aju C., Di Napoli M., Prakash P., Gopinath G., Shaji E., Chandra V., 2023. Machine-learning based landslide susceptibility modelling with emphasis on uncertainty analysis. Geoscience Frontiers, 14(6), 101657.

Achu A.L., Aju C.D., Di Napoli M., Prakash P., Gopinath G., Shaji E., Chandra V., 2023. Machine-learning based landslide susceptibility modelling with emphasis on uncertainty analysis. Geoscience Frontiers, 14(6), 101657. https://doi.org/10.1016/j.gsf.2023.101657.

Al-Najjar H.A., Pradhan B., Kalantar B., Sameen M.I., Santosh M., Alamri A., 2021. Landslide susceptibility modeling: an integrated novel method based on machine learning feature transformation. Remote Sensing, 13(16), 3281.

Breiman L., 1996. Bagging predictors. Machine learning, 24, 123–140.

Bui Q.D., Ha H., Khuc D.T., Nguyen D.Q., von Meding J., Nguyen L.P., Luu C., 2023. Landslide susceptibility prediction mapping with advanced ensemble models: Son La province, Vietnam. Natural Hazards, 116(2), 2283–2309.

Cao W.G., Fu Y., Dong Q.Y., Wang H.G., Ren Y., Li Z.Y., Du Y.Y., 2023. Landslide susceptibility assessment in Western Henan Province based on a comparison of conventional and ensemble machine learning. China Geology, 6(3), 409–419. https://doi.org/10.31035/cg2023013.

Chang Z., Catani F., Huang F., Liu G., Meena S.R., Huang J., Zhou C., 2023. Landslide susceptibility prediction using slope unit-based machine learning models considering the heterogeneity of conditioning factors. Journal of Rock Mechanics and Geotechnical Engineering, 15(5), 1127–1143.

Chang Z., Huang J., Huang F., Bhuyan K., Meena S.R., Catani F., 2023. Uncertainty analysis of non-landslide sample selection in landslide susceptibility prediction using slope unit-based machine learning models. Gondwana Research, 117, 307–320.

Dahal R.K., Hasegawa S., Nonomura A., Yamanaka M., Masuda T., Nishino K., 2008. GIS-based weights-of-evidence modelling of rainfall-induced landslides in small catchments for landslide susceptibility mapping. Environmental Geology, 54, 311–324.

Fatemi Aghda S., Bagheri V., Razifard M., 2018. Landslide susceptibility mapping using fuzzy logic system and its influences on mainlines in lashgarak region, Tehran, Iran. Geotechnical and Geological Engineering, 36, 915–937.

Ganesh B., Vincent S., Pathan S., Benitez S.R.G., 2023. Machine learning based landslide susceptibility mapping models and GB-SAR based landslide deformation monitoring systems: Growth and evolution. Remote Sensing Applications: Society and Environment, 29, 100905.

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

Goetz J., Brenning A., Petschko H., Leopold P., 2015. Evaluating machine learning and statistical prediction techniques for landslide susceptibility modeling. Computers & Geosciences, 81, 1–11.

Hajek P., Abedin M.Z., Sivarajah U., 2023. Fraud detection in mobile payment systems using an XGBoost-based framework. Information Systems Frontiers, 25(5), 1985–2003.

Hong H., 2024. Landslide susceptibility assessment using locally weighted learning integrated with machine learning algorithms. Expert systems with Applications, 237, 121678.

Hong H., Liu J., Bui D.T, Pradhan B., Acharya T.D., Pham B.T., Zhu X.Z., Chen W., Ahmad B.B., 2018. Landslide susceptibility mapping using J48 Decision Tree with AdaBoost, Bagging and Rotation Forest ensembles in the Guangchang area (China). Catena, 163, 399–413. https://doi.org/10.1016/j.catena.2018.01.005.

Huang F., Xiong H., Yao C., Catani F., Zhou C., Huang J., 2023. Uncertainties of landslide susceptibility prediction considering different landslide types. Journal of Rock Mechanics and Geotechnical Engineering, 15(3), 2954-2972. https://doi.org/10.1016/j.jrmge.2023.03.001.

Huang Z., Peng L., Li S., Liu Y., Zhou S., 2023. GIS-based landslide susceptibility mapping in the Longmen Mountain area (China) using three different machine learning algorithms and their comparison. Environmental Science and Pollution Research, 30(38), 88612–88626.

Iban M.C., Bilgilioglu S.S., 2023. Snow avalanche susceptibility mapping using novel tree-based machine learning algorithms (XGBoost, NGBoost, and LightGBM) with eXplainable Artificial Intelligence (XAI) approach. Stochastic Environmental Research and Risk Assessment, 37(6), 2243–2270.

Kavzoglu T., Colkesen I., Sahin E.K., 2019. Machine learning techniques in landslide susceptibility mapping: a survey and a case study. Landslides: Theory, practice and modelling, 283–301.

Khaliq A.H., Basharat M., Riaz M.T., Wani S., Al-Ansari N., Long B.L., Linh N.T.T., 2023. Spatiotemporal landslide susceptibility mapping using machine learning models: A case study from district Hattian Bala, NW Himalaya, Pakistan. Ain Shams Engineering Journal, 14(3), 101907.

Kim J.-C., Lee S., Jung H.-S., Lee S., 2018. Landslide susceptibility mapping using random forest and boosted tree models in Pyeong-Chang, Korea. Geocarto International, 33(9), 1000–1015.

Le Minh N., Truyen P.T., Van Phong T., Jaafari A., Amiri M., Van Duong N., Nguyen V.B., Dao M.D., Prakash I., Pham B.T., 2023. Ensemble models based on radial basis function network for landslide susceptibility mapping. Environmental Science and Pollution Research, 30(44), 99380–99398.

Lee S., Min K., 2001. Statistical analysis of landslide susceptibility at Yongin, Korea. Environmental Geology, 40, 1095–1113.

Liu C., 2023. Optimization of negative sample selection for landslide susceptibility mapping based on machine learning using K-means-KNN algorithm. Earth Science Informatics, 16(4), 4131–4152.

Luu C., Nguyen D., Amiri M., Van Phong T., Bui Q., Prakash I., Pham B., 2022. Flood susceptibility modeling using Radial Basis Function Classifier and Fisher's linear discriminant function. Vietnam J. Earth Sci., 44(1), 55–75. https://doi. org/10.15625/2615-9783/16626.

Mandal B., Mondal S., Mandal S., 2023. GIS-based landslide susceptibility zonation (LSZ) mapping of Darjeeling Himalaya, India using weights of evidence (WoE) model. Arabian Journal of Geosciences, 16(7), 1–20.

Mandal I., Pal S., 2020. Modelling human health vulnerability using different machine learning algorithms in stone quarrying and crushing areas of Dwarka river Basin, Eastern India. Advances in Space Research, 66(6), 1351–1371.

Merghadi A., Yunus A.P., Dou J., Whiteley J., Thai Pham B., Bui D.T., Ram A., Abderrahmane, B. (2020). Machine learning methods for landslide susceptibility studies: A comparative overview of algorithm performance. Earth-Science Reviews, 207, 103225.

Nguyen H.D., Dang D.K., Bui Q.T., Petrisor A.I., 2023. Multi-hazard assessment using machine learning and remote sensing in the North Central region of Vietnam. Transactions in GIS, 27(5), 1614–1640.

Nguyen H.D., Nguyen Q.H., Du Q.V.V., Pham V.T., Pham L.T., Van Hoang T., Quang-Hai T., Quang-Thanh B., Petrisor A.I., 2023. Landslide susceptibility prediction using machine learning and remote sensing: Case study in Thua Thien Hue province, Vietnam. Geological Journal, 59(2), 636–658.

Nguyen H.D., Nguyen V.H., Du Q.V.V., Nguyen C.T., Dang D.K., Truong Q.H., Ngo D.B.T., Tran Q.T., Nguyen Q.H., Bui Q.-T., 2024. Application of hybrid model-based machine learning for groundwater potential prediction in the north central of Vietnam. Earth Science Informatics, 1–21.

Nguyen H.D., Van C.P., Nguyen T.G., Dang D.K., Pham T.T.N., Nguyen Q.-H., Bui Q.T., 2023. Soil salinity prediction using hybrid machine learning and remote sensing in Ben Tre province on Vietnam's Mekong River Delta. Environmental Science and Pollution Research, 1–18.

Nhu V.-H., Bui T.T., My L.N., Vuong H., Duc H.N., 2022. A new approach based on integration of random subspace and C4. 5 decision tree learning method for spatial prediction of shallow landslides. Vietnam J. Earth Sci., 44(3), 327–342. https://doi.org/10.15625/2615-9783/16929.

Nhu V.H., Thai B.P., Tien D.B., 2023. A novel swarm intelligence optimized extreme learning machine for predicting soil shear strength: A case study at Hoa Vuong new urban project (Vietnam). Vietnam J. Earth Sci., 45(2), 219–237. https://doi.org/10.15625/2615-9783/18338.

Niraj K., Singh A., Shukla D.P., 2023. Effect of the normalized difference vegetation index (NDVI) on

GIS-enabled bivariate and multivariate statistical models for landslide susceptibility mapping. Journal of the Indian Society of Remote Sensing, 51(8), 1739–1756.

Nwazelibe V.E., Egbueri J.C., Unigwe C.O., Agbasi J.C., Ayejoto D.A., Abba S.I., 2023. GIS-based landslide susceptibility mapping of Western Rwanda: an integrated artificial neural network, frequency ratio, and Shannon entropy approach. Environmental Earth Sciences, 82(19), 439.

Petrişor A.I., 2015. Using CORINE data to look at deforestation in Romania: Distribution & possible consequences. Urbanism. Arhitectură. Construcţii, 6(1), 83–90.

Pham B.T., Tien Bui D., Prakash I., 2018. Bagging based support vector machines for spatial prediction of landslides. Environmental Earth Sciences, 77, 1–17.

Poddar I., Roy R., 2024. Application of GIS-based data-driven bivariate statistical models for landslide prediction: a case study of highly affected landslide prone areas of Teesta River basin. Quaternary Science Advances, 13, 100150.

Qasimi A.B., Isazade V., Enayat E., Nadry Z., Majidi A.H., 2023. Landslide susceptibility mapping in Badakhshan province, Afghanistan: A comparative study of machine learning algorithms. Geocarto International, 38(1), 2248082.

Qazi A., Singh K., Vishwakarma D.K., Abdo H.G., 2023. GIS based landslide susceptibility zonation mapping using frequency ratio, information value and weight of evidence: a case study in Kinnaur District HP India. Bulletin of Engineering Geology and the Environment, 82(8), 332.

Qiu Y., Wang J., Li Z., 2023. Personalized HRTF prediction based on LightGBM using anthropometric data. China Communications, 20(6), 166–177.

Rai S.C., Pandey V.K., Sharma K.K., Sharma S., 2024. Landslide susceptibility analysis in the Bhilangana Basin (India) using GIS-based machine learning methods. Geosystems and Geoenvironment, 100253. https://doi.org/10.1016/j.geogeo.2024.100253.

Rezapour H., Jamali S., Bahmanyar A., 2023. Review on Artificial Intelligence-Based Fault Location Methods in Power Distribution Networks. Energies, 16(12), 4636.

Mohamed S., Tayeb B., Icon, Mawloud G., Karim I.A., Sameh A.K., Tetsuya S., Hamouda B., Daisuke N., Emad M., 2022. Examining LightGBM and CatBoost models for wadi flash flood susceptibility prediction. Geocarto International, 37(25), 7462–7487.

Sahin E.K., 2020. Assessing the predictive capability of ensemble tree methods for landslide susceptibility mapping using XGBoost, gradient boosting machine, and random forest. SN Applied Sciences, 2(7), 1308.

Saito H., Nakayama D., Matsuyama H., 2009. Comparison of landslide susceptibility based on a decision-tree model and actual landslide occurrence: the Akaishi Mountains, Japan. Geomorphology, 109(3–4), 108–121.

Sharma A., Prakash C., 2023. Impact assessment of road construction on landslide susceptibility in mountainous region using GIS-based statistical modelling. Journal of the Geological Society of India, 99(8), 1131–1140.

Sharma N., Saharia M., Ramana G., 2024. High resolution landslide susceptibility mapping using ensemble machine learning and geospatial big data. Catena, 235, 107653.

Taalab K., Cheng T., Zhang Y., 2018. Mapping landslide susceptibility and types using Random Forest. Big Earth Data, 2(2), 159–178.

Tien Bui D., Ho T.C., Revhaug I., Pradhan B., Nguyen D.B., 2014. Landslide susceptibility mapping along the national road 32 of Vietnam using GIS-based J48 decision tree classifier and its ensembles. Paper presented at the Cartography from pole to pole: selected contributions to the XXVI$^{th}$ international conference of the ICA, Dresden 2013.

Tyagi A., Tiwari R.K., James N., 2023. Mapping the landslide susceptibility considering future land-use land-cover scenario. Landslides, 20(1), 65–76.

Ukey N., Yang Z., Li B., Zhang G., Hu Y., Zhang W., 2023. Survey on exact knn queries over high-dimensional data space. Sensors, 23(2), 629.

Vakhshoori V., Zare M., 2016. Landslide susceptibility mapping by comparing weight of evidence, fuzzy logic, and frequency ratio methods. Geomatics, Natural Hazards and Risk, 7(5), 1731–1752.

Van Phong T., Ly H.-B., Trinh P.T., Prakash I., Dao T.H., 2020. Landslide susceptibility mapping using Forest by Penalizing Attributes (FPA) algorithm based machine learning approach. Vietnam J. Earth Sci., 42(3), 237–246. https://doi.org/10.15625/0866-7187/42/3/15047.

Viet Du Q.V., Nguyen H.D., Pham V.T., Nguyen C.H., Nguyen Q.-H., Bui Q.-T., Doan T.T., Tran A.T., Petrisor A.-I., 2023. Deep learning to assess the effects of land use/land cover and climate change on landslide susceptibility in the Tra Khuc river basin of Vietnam. Geocarto International, 2172218.

Wu Y., Ke Y., Chen Z., Liang S., Zhao H., Hong H., 2020. Application of alternating decision tree with AdaBoost and bagging ensembles for landslide susceptibility mapping. Catena, 187, 104396.

Xuan B.T., Pham The T., Luu Thuy D., Tran Van P., Vuong Hong N., Van Le H., Duc Nguyen D., Prakash I., Pham Thanh T., Binh Thai B., 2024. Groundwater potential zoning using Logistics Model Trees based novel ensemble machine learning model. Vietnam J. Earth Sci., 46(2), 272–281. https://doi.org/10.15625/2615-9783/20316.

Yang C., Liu L.-L., Huang F., Huang L., Wang X.-M., 2023. Machine learning-based landslide susceptibility assessment with optimized ratio of landslide to non-landslide samples. Gondwana Research, 123, 198–216.

Ye P., Yu B., Chen W., Liu K., Ye L., 2022. Rainfall-induced landslide susceptibility mapping using machine learning algorithms and comparison of their performance in Hilly area of Fujian Province, China. Natural Hazards, 113(2), 965–995.

Yeon Y.-K., Han J.-G., Ryu K.H., 2010. Landslide susceptibility mapping in Injae, Korea, using a decision tree. Engineering Geology, 116(3-4), 274–283.

Zeng T., Jin B., Glade T., Xie Y., Li Y., Zhu Y., Yin K., 2024. Assessing the imperative of conditioning factor grading in machine learning-based landslide susceptibility modeling: A critical inquiry. Catena, 236, 107732.

Zhang J., Ma X., Zhang J., Sun D., Zhou X., Mi C., Wen H., 2023. Insights into geospatial heterogeneity of landslide susceptibility based on the SHAP-XGBoost model. Journal of Environmental Management, 332, 117357.

Zhang T., Fu Q., Wang H., Liu F., Wang H., Han L., 2022. Bagging-based machine learning algorithms for landslide susceptibility modeling. Natural Hazards, 110(2), 823–846.

Zhang T., Li S., Jing X., Song J., Shi L., He X., 2023. Negative comment recognition model based on lightGBM. Proc. SPIE 12597, Second International Conference on Statistics, Applied Mathematics, and Computing Science (CSAMCS 2022), 125973Z (28 March 2023). https://doi.org/10.1117/12.2672779.

Zhang W., He Y., Wang L., Liu S., Meng X., 2023. Landslide Susceptibility mapping using random forest and extreme gradient boosting: A case study of Fengjie, Chongqing. Geological Journal, 58(6), 2372-2387.

Zhou C., Yin K., Cao Y., Ahmed B., Li Y., Catani F., Pourghasemi H.R., 2018. Landslide susceptibility modeling applying machine learning methods: A case study from Longju in the Three Gorges Reservoir area, China. Computers & Geosciences, 112, 23–37.