



# Comparison of tree-based ensemble learning algorithms for landslide susceptibility mapping in Murgul (Artvin), Turkey

Ziya Usta<sup>1</sup> · Halil Akinci<sup>1</sup> · Alper Tunga Akın<sup>2</sup>

Received: 5 November 2023 / Accepted: 22 February 2024 / Published online: 4 March 2024  
© The Author(s) 2024

## Abstract

Turkey's Artvin province is prone to landslides due to its geological structure, rugged topography, and climatic characteristics with intense rainfall. In this study, landslide susceptibility maps (LSMs) of Murgul district in Artvin province were produced. The study employed tree-based ensemble learning algorithms, namely Random Forest (RF), Light Gradient Boosting Machine (LightGBM), Categorical Boosting (CatBoost), and eXtreme Gradient Boosting (XGBoost). LSM was performed using 13 factors, including altitude, aspect, distance to drainage, distance to faults, distance to roads, land cover, lithology, plan curvature, profile curvature, slope, slope length, topographic position index (TPI), and topographic wetness index (TWI). The study utilized a landslide inventory consisting of 54 landslide polygons. Landslide inventory dataset contained 92,446 pixels with a spatial resolution of 10 m. Consistent with the literature, the majority of landslide pixels (70% – 64,712 pixels) were used for model training, and the remaining portion (30% – 27,734 pixels) was used for model validation. Overall accuracy, precision, recall, *F1*-score, root mean square error (RMSE), and area under the receiver operating characteristic curve (AUC-ROC) were considered as validation metrics. LightGBM and XGBoost were found to have better performance in all validation metrics compared to other algorithms. Additionally, SHapley Additive exPlanations (SHAP) were utilized to explain and interpret the model outputs. As per the LightGBM algorithm, the most influential factors in the occurrence of landslide in the study area were determined to be altitude, lithology, distance to faults, and aspect, whereas TWI, plan and profile curvature were identified as the least influential factors. Finally, it was concluded that the produced LSMs would provide significant contributions to decision makers in reducing the damages caused by landslides in the study area.

**Keywords** Landslide · Landslide susceptibility · Machine learning · Ensemble learning algorithms · Murgul

## Introduction

Landslides are one of the most common natural disasters in areas with rugged topography. Landslides triggered by seismic activities, heavy rainfall, or human activities cause

significant economic damage, environmental destruction, and loss of life, especially in areas with steep slopes. Therefore, reducing and preventing landslide-related damages continue to be important research topics in disaster management (Zhang et al. 2022a). To reduce landslide-related damages and losses, landslide susceptibility (LS) assessment should be conducted in areas prone to landslides, and landslide susceptibility maps (LSMs) for these areas should be generated (Das et al. 2023). It is possible to identify areas with a high probability of future landslides through LS mapping (Ye et al. 2022).

Various statistical methods have been used to generate LSMs. Logistic regression (LR) (Yilmaz 2009; Kavzoglu et al. 2014; Dağ et al. 2020), frequency ratio (FR) method (Akgun et al. 2008; Demir 2019; Akinci and Yavuz Ozalp 2021), and weights of evidence (Lee and Choi 2004; Sifa et al. 2020; Wang et al. 2020a) are the most commonly used statistical methods. However, Wei et al. (2022) mentioned

---

Communicated by: H. Babaie

✉ Ziya Usta  
ziyausta@artvin.edu.tr

Halil Akinci  
halil.akinci@artvin.edu.tr

Alper Tunga Akın  
alpertunga@ktu.edu.tr

<sup>1</sup> Department of Geomatics Engineering, Artvin Coruh University, Artvin 08100, Turkey

<sup>2</sup> Department of Geomatics Engineering, Karadeniz Technical University, 61080 Trabzon, Turkey

that while statistical methods are relatively easy to use, they fall short in explaining the complex, inconsistent, and non-linear relationships between landslide events and the conditioning factors that influence them. Therefore, in recent years, machine learning (ML) and deep learning-based models have gained significant attention in this field.

In the past decade, individual or stand-alone ML algorithms such as artificial neural networks (Pradhan and Lee 2010; Aditian et al. 2018; Akinci 2022), decision trees (Tien Bui et al. 2012; Pradhan 2013), naive bayes (Chen et al. 2018; Pourghasemi et al. 2018), and support vector machines (SVM) (Goetz et al. 2015; Colkesen et al. 2016; Akinci and Zeybek 2021) have been widely used in LS mapping. On the other hand, several studies in the literature have reported that ensemble learning algorithms, such as adaptive boosting (AdaBoost), random forest (RF), gradient boosting machine (GBM), light gradient boosting machine (LightGBM), categorical boosting (CatBoost), and extreme gradient boosting (XGBoost), outperform individual ML algorithms in terms of prediction accuracy, stability, and robustness (Kavzoglu and Teke 2022; Liu et al. 2022; Wei et al. 2022). Ensemble learning algorithms generate multiple models using different samples of the same dataset to produce more accurate solutions and improve prediction performance by combining the predictions from these models. Sahin (2022) and Ye et al. (2022) emphasized that CatBoost, LightGBM, and XGBoost are rarely used and not extensively evaluated in LS mapping studies, mainly due to being relatively new methods compared to other ML models. Although tree-based ensemble learning algorithms demonstrate better performance than single models, it has not been determined yet which is the most suitable algorithm for LS mapping (Wei et al. 2022; Ye et al. 2022). Different studies highlight the prominence of different algorithms on the attributes of the research location, the conditioning factors used, and the values of the algorithms' hyperparameters. Therefore, more studies comparing the performance of different algorithms are needed for a comprehensive evaluation.

In Turkey, due to its geological, geomorphological, topographic, and climatic characteristics, regions such as the Black Sea Region, Eastern Anatolia, and Central Anatolia are frequently affected by landslides. Approximately 20% of landslides in Turkey occur in the Eastern Black Sea Region. The provinces with the highest happening of landslides in this region are Trabzon, Rize, Giresun, and Artvin, respectively. According to the statistical data from the Disaster and Emergency Management Presidency (AFAD) covering the years 1950–2019, Trabzon experienced 1517 landslides, Rize had 1319, Giresun had 913, and Artvin had 765 landslide events (AFAD 2020). In the study by Dalkes and Korkmaz (2023), LSMs of Akçaabat and Düzköy districts of Trabzon province were produced using Analytic Hierarchy Process (AHP) and FR methods. Upon comparison of

the LS maps produced by both methods, it was determined that the FR method yielded more accurate results than the AHP method in identifying the locations of the observed landslides in the study area. Yavuz Ozalp et al. (2023) produced LSMs of Ardeşen and Fındıklı districts of Rize province using tree-based ensemble learning algorithms such as RF, GBM, CatBoost and XGBoost. Using the ROC curve and AUC metric, the researchers found that CatBoost performed slightly better than other models. In the study by Kaya Topaçlı et al. (2024), LR and RF models were used to produce LSMs of the Bolaman river basin of Ordu province, located in the Eastern Black Sea Region in Turkey. The study area is one of the basins where landslides occur most frequently in Turkey, as in many places in the Eastern Black Sea Region. The validation results of the study showed that RF is superior to LR in terms of performance. Landslides that inflict harm upon structures and infrastructure, also economic losses and loss of life, can be observed in all districts of Artvin. However, a review of the literature reveals that LS assessments have been conducted in the districts of Arhavi, Hopa, Kemalpaşa, Merkez, Ardanuç, and Şavşat in Artvin (Akinci et al. 2020, 2021; Akinci and Yavuz Ozalp 2021; Akinci and Zeybek 2021), while no studies have been conducted in the districts of Borçka, Murgul, and Yusufeli.

The aim of this study was to produce susceptibility maps showing areas prone to landslides in Murgul district of Artvin province and to identify the main factors contributing to landslide susceptibility. Prediction models using RF, LightGBM, CatBoost and XGBoost ML algorithms were used for susceptibility mapping. The prediction capabilities or performances of the susceptibility models were evaluated using metrics such as overall accuracy, precision, recall, *F1*-score, root mean square error (RMSE), and AUC-ROC. Furthermore, the SHapley Additive exPlanations (SHAP) approach, which aims to explain the local behavior of black box ML models, was used to improve the interpretability of model predictions. When the current literature is reviewed, it is seen that there are limited number of studies using SHAP approach in LS modeling (Pradhan et al. 2023; Sun et al. 2023a; Teke and Kavzoglu 2023; Vega et al. 2023; Youssef et al. 2023; Zhang et al. 2023). Therefore, the innovative aspect and contribution of this study to the literature is the use of SHAP in LS modeling.

## Material and methods

The aim of this study is to produce LSMs of the study area using tree-based ensemble learning algorithms and to compare the performance of the ML models used. The study can be broadly divided into 6 steps: (1) prepare spatial data for landslide conditioning factors; (2) test the independence of conditioning factors using multicollinearity analysis; (3)

collection of landslide inventory data, creation of training and validation data sets; (4) produce LSMs using RF, GBM, CatBoost, and XGBoost models; (5) evaluate the performance of the models using validation metrics; (6) interpret the models using SHAP values and explain the formation mechanisms of landslides in the study area. This study was carried out using ArcGIS 10.5, SAGA GIS 7.9, Python and Scikit-learn library. A flowchart summarizing the methodology applied in this study, inspired by Yu et al. (2023), is shown in Fig. 1.

### Study area and geological structure

The study area consists of a bounding box covering the Murgul district. This area is located between 41° 7' 1.39"—41° 22' 24.86" north latitude and 41° 27' 55.41"—41° 41' 1.53" east longitude, with a total area of 48,938.23 hectares. The average elevation in the study area is 1347 m, and the elevation ranges from 100 to 3370 m (Fig. 2). The study area has a highly rugged topography, and the slope values change from 0° to 76.33°. The mean slope value in the study area is 29.09°. In the study area, 4.09% of the area has a slope below 10°, 15.78% has a slope between 10° and 20°, and 80.13% has a slope above 20°.

The study area includes 13 villages along with the town center of Murgul. Two of these villages, Civan and Akpınar,

are administratively affiliated with the Borçka district (Fig. 2). Based on the data from the Turkish Statistical Institute (TURKSTAT), the total population of Murgul district in 2021 is 6522 (TURKSTAT 2023). Out of this population, 5020 reside in the town center, while 1502 live in the villages. With the inclusion of the 2 villages affiliated with the Borçka district, the total population in the study area reaches 6896 people.

The study area is characterized by a Black Sea climate. Based on the meteorological measurements collected from the General Directorate of Meteorology for the years 2015–2022, the average temperature in Murgul district is 12.76 °C. The yearly mean rainfall in Murgul is 980.63 mm. The highest temperature recorded in the district was 31 °C in August 2022, while the lowest temperature was -3 °C in December 2016.

In this study, a 1/100,000 scale geological map obtained from the General Directorate of Mineral Research and Exploration (GDMRE) was used (Keskin 2013a; 2013b). Based on this map, there are 18 different lithological units in the study area (Fig. 3). The study area comprises lithological units ranging in age from Late Cretaceous (Turonian-Coniacian) to Quaternary (Table 1). Four lithological units account for 78.52% of the study area: Kızılkaya Formation (Kk) covers 24.28%, Kabaköy Formation (Tek) covers 23.66%, Çatak Formation (Kç) covers 20.6%, and

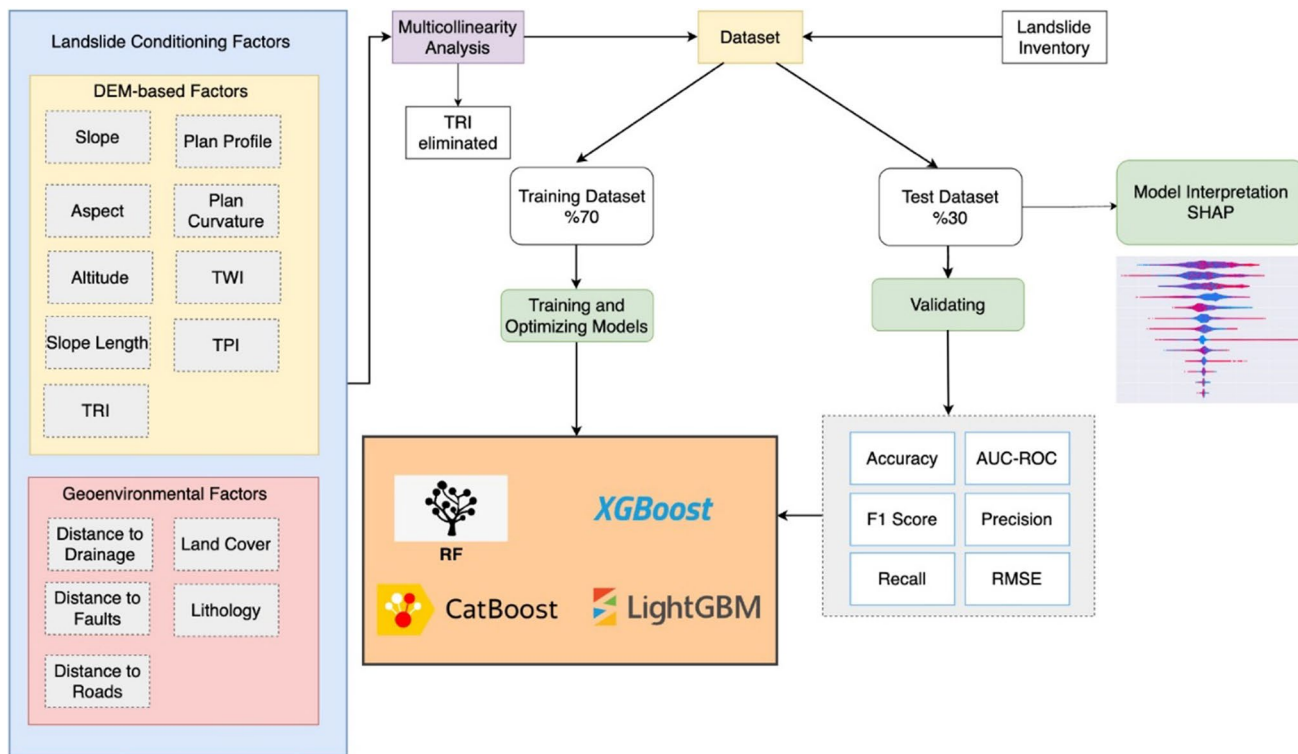
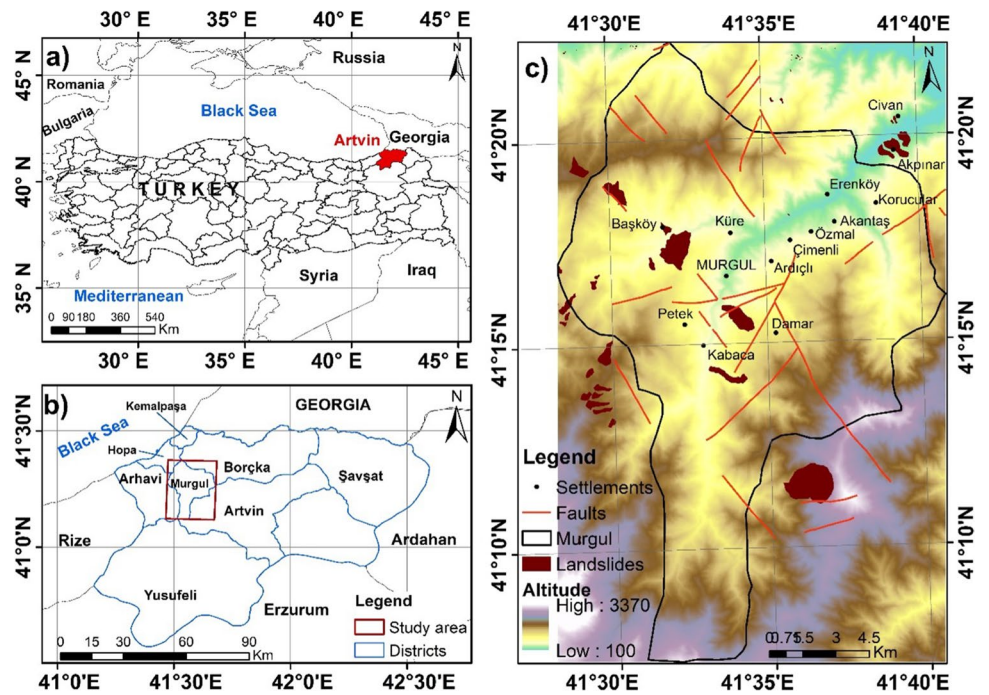
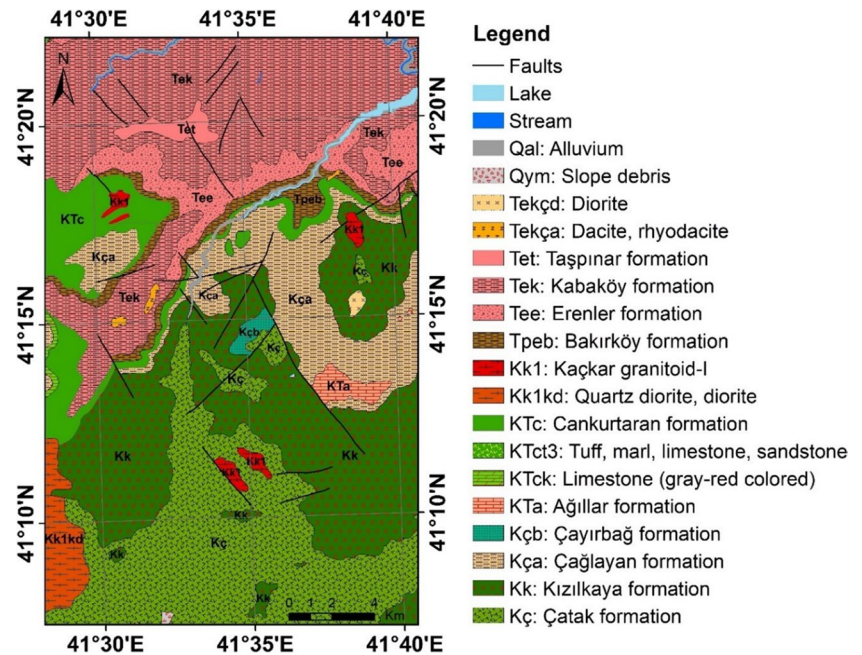


Fig. 1 The flowchart representing the methodology followed in the study

**Fig. 2** Study area: (a) location of Artvin, (b) location of Murgul, (c) landslides in Murgul



**Fig. 3** Lithological map of the study area



Çağlayan Formation (Kça) covers 9.98%. The Santonian-aged Kızılkaya Formation (Kk) consists of rhyodacitic, dacitic lavas, and pyroclastics.

The Kabaköy Formation (Tek), which is of Middle Eocene age, starts with clastic and carbonates and includes andesitic, basaltic lavas, and pyroclastics, as well as conglomerates, sandy limestone, sandstone, marl, and tuff. The Late Cretaceous (Turonian-Coniacian)

Çatak Formation (Kç) comprises basaltic andesitic lavas and pyroclastics, along with argillaceous limestone, marl, siltstone, and shale. The Campanian–Maastrichtian-aged Çağlayan Formation (Kça) is composed of basaltic andesitic lavas and pyroclastics, as well as mudstone and sandstone (Keskin 2013a; 2013b). The rocks contained in the other lithological units in the study area are given in Table 1.

**Table 1** Lithological units of the study area

Pixel value	Symbol	Age / Formation name	The content of the formation
1	Qal	Quaternary / -	Alluvium
2	Qym	Quaternary / -	Slope debris
3	Tekçd	Middle-Upper Eocene / -	Diorite
4	Tekça	Middle-Upper Eocene / -	Dacite, rhyodacite
5	Tet	Middle Eocene / Taşınar formation	Andesitic and dacitic volcanics, volcano-clastic turbiditic rocks
6	Tek	Middle Eocene / Kabaköy formation	Andesite, basaltic lava and pyroclastics, sandy limestone, sandstone, marl, tuff
7	Tee	Middle Eocene / Erenler formation	Mudstone, claystone, sandstone alternation
8	Tpeb	Paleocene-Lower Eocene / Bakırköy formation	Siltstone, claystone, sandstone, clayey limestone, marl
9	Kk1	Upper Cretaceous-Paleocene / Kaçkar granitoid-I	Granite, granodiorite, quartz diorite, adamellite, gabbro, diabase
10	Kk1kd	Upper Cretaceous-Paleocene / -	Quartz diorite, diorite
11	KTc	Maastrichtian-Danian / Cankurtaran formation	Sandy limestone, micritic limestone, tuff, marl, volcanics sandstone, agglomerate
12	KTct3	Maastrichtian-Danian / -	Tuff, marl, limestone, sandstone
13	KTck	Maastrichtian-Danian / -	Limestone (gray-red colored)
14	KTa	Maastrichtian-Danian / Ağıllar formation	Reefal limestone, sandy limestone
15	Kçb	Maastrichtian / Çayırbağ formation	Dacite, rhyolite, rhyodacitic lava and pyroclastics
16	Kça	Campanian–Maastrichtian / Çağlayan formation	Basaltic, andesitic lava and pyroclastics, mudstone, sandstone
17	Kk	Santonian / Kızılkaya formation	Rhyodacitic, dacitic lava and pyroclastics
18	Kç	Turonian-Coniacian / Çatak formation	Basalt, andesitic lava and pyroclastics, clayey limestone, marl, siltstone, claystone

## Landslide inventory

Landslide susceptibility (LS) mapping studies require landslide inventory data both for model training and validation stages. For models to generate reliable results, the inventory data needs to be up-to-date, accurate, and complete. The landslide inventory map (LIM) used in the study contains 54 landslide polygons (Fig. 2).

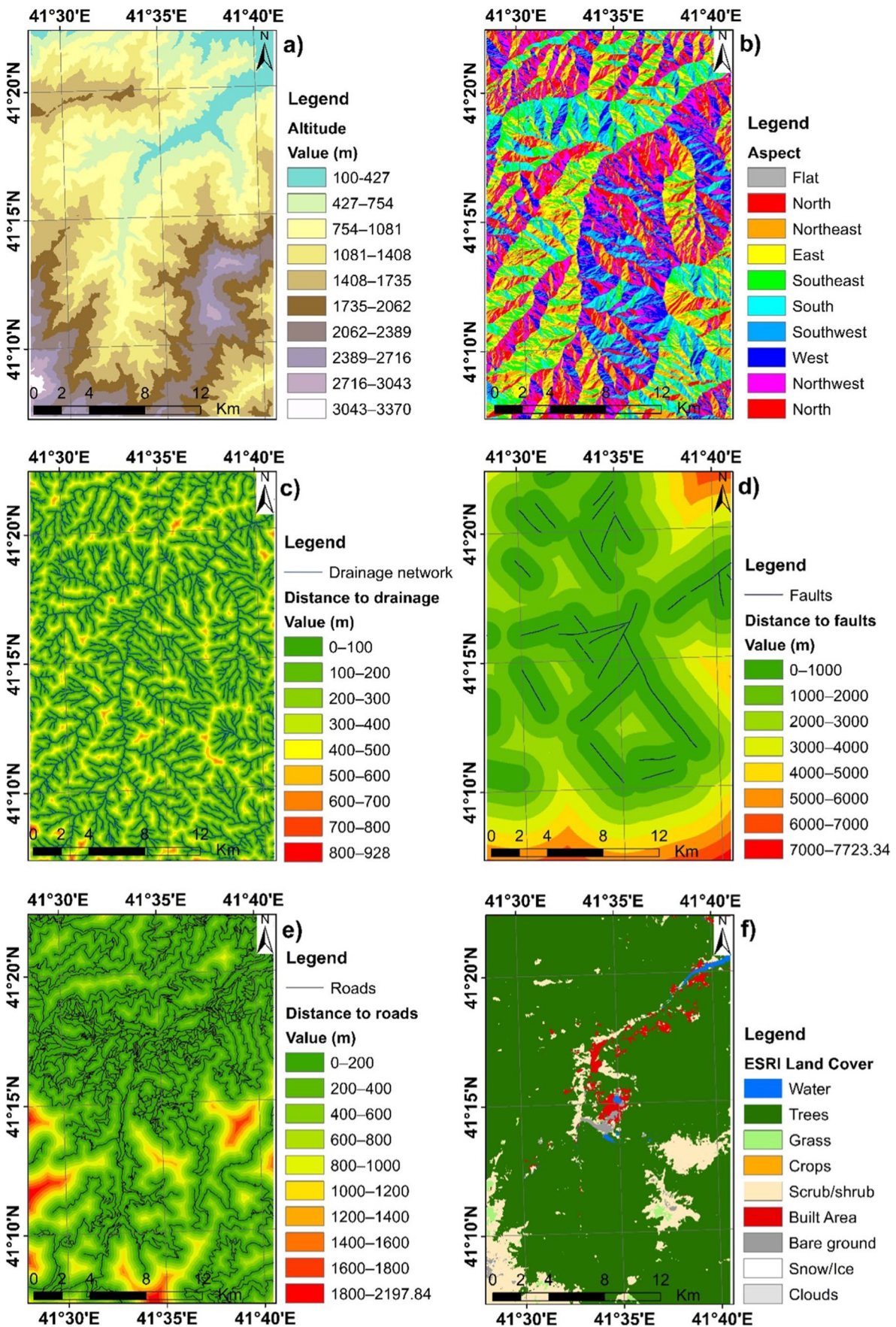
Among the landslide polygons, 11 of them were obtained from the 1/25,000 scale LIM generated by the GDMRE. In this inventory, landslide polygons are classified as Type 1, Type 2, and Type 4. Type 1 represents inactive landslides, Type 2 represents active landslides, and Type 4 represents active flows (Erener et al. 2016; Duman and Çan 2023). The remaining 43 landslide polygons used in the study were obtained from the map indicating the inventory of landslides produced by the Artvin Provincial Disaster and Emergency Directorate. When examining landslides and related data in Turkey, it is known that predominantly translational and rotational landslides, as well as complex landslides combining multiple types, occur (AFAD 2020). According to Varnes' (1978) slope movement classification, out of the 43 landslides, 1 is categorized as rotational slide, 8 categorized as translational slide, 24 as flow, 6 as debris flow, and 4 as complex landslide. The total area covered by the landslide

polygons is 9,241,303.99 m<sup>2</sup>. The landslide polygons cover approximately 2% of the study area. The smallest and the largest landslide polygons in the study area have areas of 38.32 m<sup>2</sup> and 2,809,890 m<sup>2</sup>, respectively.

A total of 54 landslide polygons in the study area were transformed into raster format, utilizing a spatial resolution of 10 m. As a result of this conversion, the “1” value was assigned to 92,446 pixels, which are referred to as positive examples. To create negative examples, a matching quantity of pixels without landslides were randomly selected in the implementation of ML models in Python, and these pixels were assigned a value of “0”. Subsequently, the total of 184,892 pixels, consisting of landslide and non-landslide examples, was divided into two datasets in a 70/30 ratio, following the literature (Sahin 2022; Youssef and Pourghasemi 2021; Akinci 2022; Liu et al. 2022; Wei et al. 2022). These datasets were employed for the training and validation of the model.

## Landslide conditioning factors

In this study, LSMs were generated using 13 factors, including land cover, aspect, slope, lithology, elevation, plan curvature, profile curvature, distance to drainage networks, distance to faults, distance to roads, slope length, topographic



**Fig. 4** Maps that represent conditioning factors: **a)** altitude, **b)** aspect, **c)** distance to drainage, **d)** distance to faults, **e)** distance to roads, **f)** land cover

position index (TPI), and topographic wetness index (TWI). These factors were determined based on the availability or producibility of spatial data, the geological and environmental characteristics of the study area, and relevant literature. Raster-based factor maps were produced at a spatial resolution of 10 m using ArcGIS 10.5 and SAGA GIS 7.9 software (Fig. 4 and Fig. 5). Digital topographic maps of the study area were obtained from the General Directorate of Mapping. The digital elevation model (DEM) of the study area with a resolution of 10 m was produced by using the contour lines in these 1/25,000 scale topographic maps. Aspect, slope, elevation, plan and profile curvature were prepared in ArcGIS 10.5 software using this DEM. Additionally, slope length, TPI and TWI were produced using the same DEM in SAGA GIS software. As explained in Section "Study Area and Geological Structure", the lithological units and fault lines were obtained from the 1:100,000 geological map provided by GDRME. The map displaying the distance to faults in the study area was generated using the "Euclidean Distance" function in ArcGIS software. The study area's drainage network was generated using SAGA GIS software based on the DEM. Additionally, the "Euclidean Distance" function of ArcGIS software was used to produce the distance map to the drainage networks. A comprehensive digital road network dataset including highways, village roads and forest roads in the study area was obtained from Artvin Regional Directorate of Forestry. The map representing the distance to the roads was produced using the "Euclidean Distance" function of ArcGIS software, as in other distance maps. On the other hand, the 10 m resolution land cover dataset of the study area was obtained from ESRI (<https://livingatlas.arcgis.com/landcover/>). The relationship between conditioning factors and landslide occurrences has been well explained in numerous studies (Gomez and Kavzoglu 2005; Kavzoglu et al. 2014; Pourghasemi and Rahmati 2018; Zhao et al. 2019; Dağ et al. 2020; Youssef and Pourghasemi 2021; Ye et al. 2022), hence detailed discussions on this topic are not provided in this article. Instead, basic statistical data related to conditioning factors are presented (Table 2).

## Machine learning algorithms used in the study

### Random forest (RF)

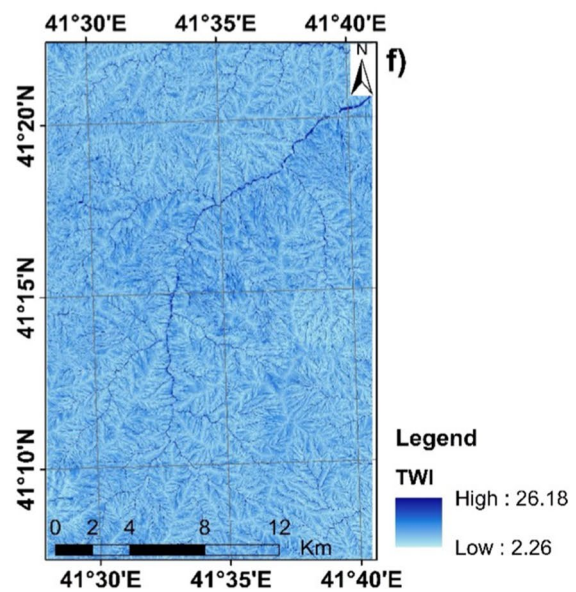
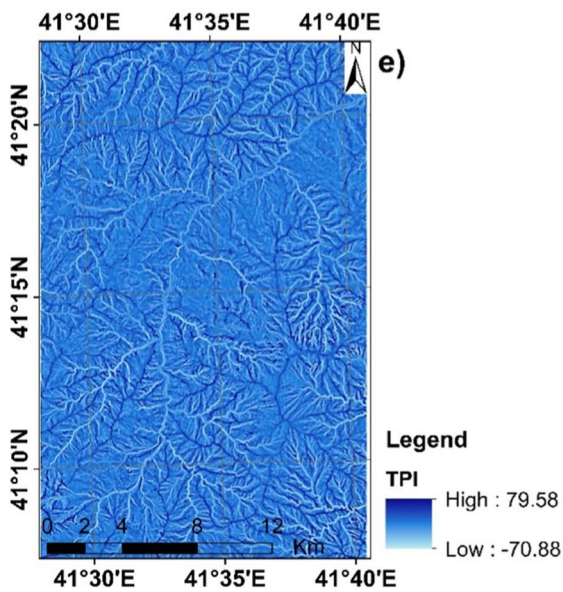
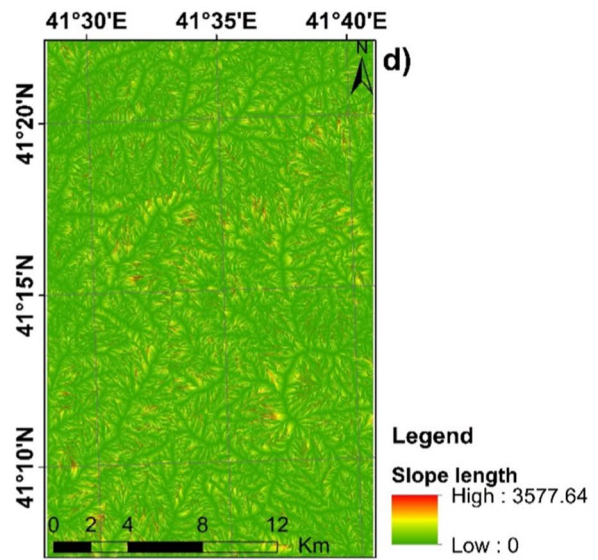
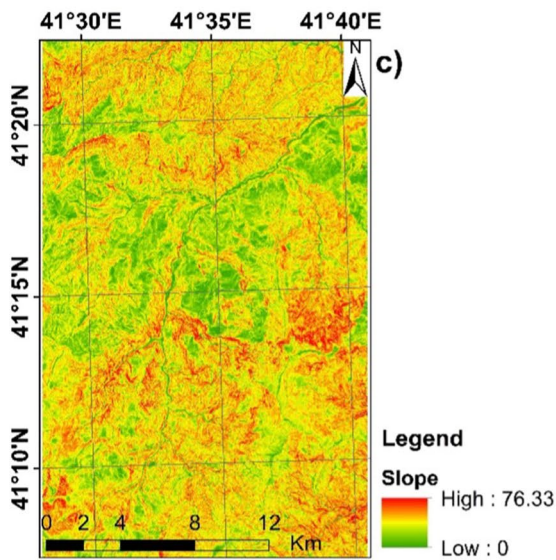
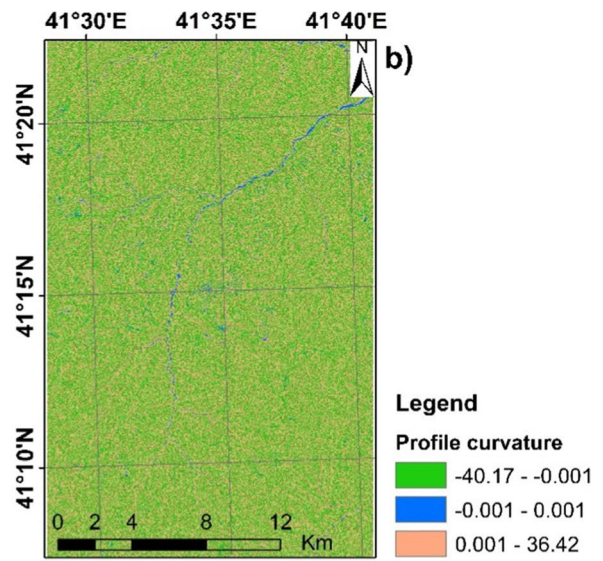
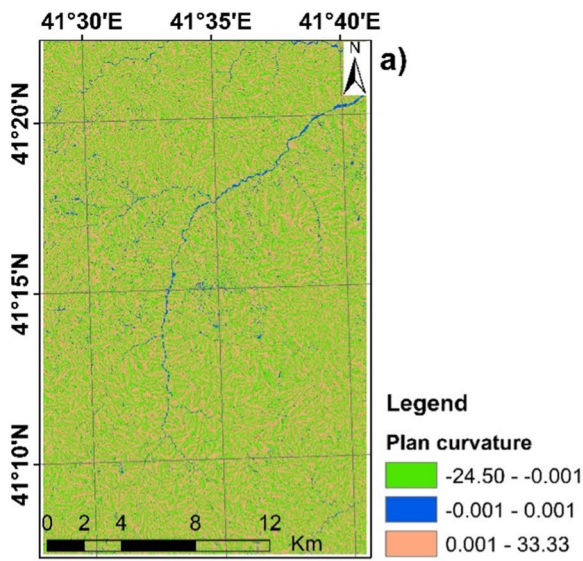
The Random Forest (RF) algorithm, originally proposed by Breiman (2001), is a type of ML algorithm designed for nonparametric multivariate classification. It has been widely adopted in LS mapping studies, and has been discussed in detail by Catani et al. (2013). The Random Forest (RF)

algorithm is a well-known technique in the field of ensemble learning. It is frequently utilized in classification and regression tasks. In contrast to a single decision tree, which is susceptible to overfitting and may exhibit high variance or bias (Taalab et al. 2018; Park and Kim 2019), RF generates multiple instances of decision trees and aggregates their predictions to arrive at a final classification (Youssef et al. 2016). This approach allows the algorithm to mitigate the weaknesses of individual trees and improve predictive performance, making it a widely-used tool in the field of ML. Decision trees are created using randomly selected subsets of the training data. The final prediction produced by RF is achieved by aggregating the predictions of all the decision trees (Akinci et al. 2020).

Two parameters need to be defined when creating an RF, the number of decision trees (*n<sub>tree</sub>*) and the number of variables or factors used at each node of the decision tree (*m<sub>try</sub>*). Although there is no definitive rule for selecting the number of trees in RF, augmenting the number of trees does not guarantee an enhancement in the model's accuracy (Taalab et al. 2018). Conversely, the variable numbers utilized at each node of the decision tree should be equal to the square root of the total number of variables (Chen et al. 2020). The out-of-bag (OOB) error, which is the percentage of misclassifications over all out-of-bag factors, is used to estimate the generalization error and evaluate the importance of variables (Achour and Pourghasemi 2020; Cao et al. 2020). In this study, RF method implemented using scikit-learn Random Forest regressor with the *n<sub>tree</sub>* and *m<sub>try</sub>* parameters set to 100 and 13, respectively. A tenfold cross-validation approach was used to validate the consistency of the model's results.

### Extreme gradient boosting (XGBoost)

XGBoost, introduced by Chen and Guestrin (2016) is a combination of the gradient boosting algorithm and the decision tree models (Wei et al. 2022; Cao et al. 2020). The XGBoost has gained popularity in LS mapping studies (Kavzoglu and Teke 2022). The main advantage of the XGBoost is its performance in the terms of runtime speed and accuracy (Wei et al. 2022). The XGBoost algorithm utilizes a gradient boosting technique, which constructs a tree by splitting features and recursively adding trees (Zhang et al. 2020a, b). For each time a new tree is added, a new function is yielded by fitting the residual value of the previous predictions. The tree is constructed by training the model, hence, the leaf node of the tree stores a score, and the sample's predicted value is the sum of the scores of all nodes (Ye et al. 2022). The aim of the model is to minimize the difference between predicted value and true value by minimizing the loss function of the training data as shown in Eq. 1 (Wang et al. 2020b).





**Fig. 5** Maps that represent conditioning factors: **a)** plan curvature, **b)** profile curvature, **c)** slope, **d)** slope length, **e)** TPI, **f)** TWI

$$\varphi^t = \sum_{i=1}^n l\left(y_i, \left(\hat{y}_i^{(t-1)} + f_t(x_i)\right)\right) + \Omega(f_t) \quad (1)$$

where  $i$  stands for the number of a given predicted value  $\hat{y}$  ( $i = 1, 2, 3, \dots, n$ );  $n$  stands for the total number of  $y$  values;  $t$  stands for the iteration number;  $l(y_i, \hat{y}_i)$  stands for the loss function between actual value  $y_i$  and the predicted value  $\hat{y}_i$ ;  $X_i$  stands for the features of the  $i$ 'th sample;  $f_t(X_i)$  stands for the base learner added to the  $t$ th iteration;  $\Omega(f_t)$  stands for regularization and finally  $t$  stands for objective function.

ML algorithms have some parameters that need to be tuned during the training phase. These parameters, called hyperparameters, significantly affect the accuracy, performance and processing speed of the model (Yavuz Ozalp et al. 2023). XGBoost includes several hyperparameters that can be tuned to improve the performance of the model. These include “*n\_estimators*” (the maximum number of iterations or trees), “*max\_depth*” (the maximum depth of the trees), “*eta*” (the learning rate), “*gamma*” (the regularization parameter), “*colsample\_bytree*” (the number of features or variables supplied to a tree), “*min\_child\_weight*” (the minimum sum of instance weight needed in a child), and “*subsample*” (the number of samples or observations supplied to a tree).

### Light gradient boosting machine (LightGBM)

Light gradient boosting (LightGBM), firstly proposed by Ke et al. (2017), is a gradient boosting approach. LightGBM is designed to overcome performance issues encountered by gradient boosting decision trees in data intensive applications (Zhang et al. 2022b). Thanks to its support for efficient concurrent training, faster training speed, reduced memory usage and distributed computation capabilities, it can handle big data efficiently and can be considered an improvement to the XGBoost (Dai et al. 2021).

The efficiency of LightGBM comes from two novel optimization techniques namely Gradient-based One-Side Sampling (GOSS) that diminishes the number of data instances and Exclusive Feature Bundling (EFB) which decreases the number of features (Zhang et al. 2022b). GOSS down-samples the data instances by discarding a portion of instances with small gradients and retaining instances with large gradients to evaluate information gain. More detailed information can be found in Ke et al. (2017). EFB reduces the number of features by grouping related features together. To be able to bundle features without compromising accuracy, conflict rate is used to determine whether a feature should be bundled or not (Fang et al. 2021).

When creating the LightGBM model, some hyperparameters need to be tuned. These basic hyperparameters are “*boosting\_type*”, “*num\_leaves*”, “*max\_depth*”, “*num\_iterations*”, and “*learning\_rate*” (Zhou et al. 2022; Omotehinwa et al. 2023). The “*boosting\_type*” parameter defines the gradient boosting method to be run. Valid values are “*gbdt*”, “*rf*”, “*dart*”, and “*goss*”, but the default is “*gbdt*”. The parameter “*num\_leaves*” refers to the maximum number of leaves in a tree. The “*max\_depth*” parameter controls the maximum depth of each tree. The parameter “*num\_iterations*” determines the number of boosting iterations, or trees to build. The “*learning\_rate*” determines the speed at which the model’s weights are updated after processing each batch of training examples.

### Categorical boosting (CatBoost)

Categorical boosting (CatBoost), firstly introduced by Prokorenkova et al. (2018), is another improved gradient boosting technique. CatBoost can process categorical data along with numerical data and needs less training data compared to other ML methods (Sahin 2022). Instead of repeatedly utilizing the same data in constructing trees, which leads to overfitting in Gradient Boosting, CatBoost uses an ordered boosting technique to combat this problem. Hence, combining the ordered boosting with the process of categorical values prevents a prediction shift stems from the special type of target leakage (Pham et al. 2022). More details about the algorithm can be found at Prokorenkova et al. (2018).

CatBoost has seven commonly used hyperparameters: *iterations*, *depth*, *learning\_rate*, *l2\_leaf\_reg*, *random\_strength*, *rsm*, and *border\_count* (Yavuz Ozalp et al. 2023). The “*iterations*” parameter specifies the maximum number of trees to be used during training. The “*depth*” parameter defines the depth of each decision tree. The “*learning\_rate*” determines the step size at each iteration while moving toward a minimum of a loss function. The “*l2\_leaf\_reg*” is the coefficient for the L2 regularization term of the cost function. The parameter “*random\_strength*”, which is used to prevent overfitting in the model, expresses the amount of randomness to be used to score splits when the tree structure is selected. The “*rsm*”, random subspace method, refers to the percentage of features to be used in each split selection when features are randomly re-selected. Lastly, the “*border\_count*” refers to the number of splits for numerical features (AWS 2024).

### Multicollinearity analysis

To enhance the accuracy of a landslide susceptibility analysis using an ML model, it is crucial to test the independence of the model’s input variables (Yu et al. 2023). Multicollinearity analysis is conducted to identify any

**Table 2** Statistical description of the conditioning factors and their sub-classes

Factor	Min	Max	Sub-classes	Reference
Altitude (m)	100	3370	<b>1:</b> 100–427, <b>2:</b> 427–754, <b>3:</b> 754–1081, <b>4:</b> 1081–1408, <b>5:</b> 1408–1735, <b>6:</b> 1735–2062, <b>7:</b> 2062–2389, <b>8:</b> 2389–2716, <b>9:</b> 2716–3043, <b>10:</b> 3043–3370	Chen et al. 2017 Kilicoglu 2021 Akinci 2022 Yavuz Ozalp et al. 2023
Aspect	-	-	<b>1:</b> Flat, <b>2:</b> North, <b>3:</b> Northeast, <b>4:</b> East, <b>5:</b> Southeast, <b>6:</b> South, <b>7:</b> Southwest, <b>8:</b> West, <b>9:</b> Northwest	Sun et al. 2022 He et al. 2023 Vega et al. 2023 Yavuz Ozalp et al. 2023
Distance to drainage (m)	0	928	<b>1:</b> 0–100, <b>2:</b> 100–200, <b>3:</b> 200–300, <b>4:</b> 300–400, <b>5:</b> 400–500, <b>6:</b> 500–600, <b>7:</b> 600–700, <b>8:</b> 700–800, <b>9:</b> 800–928	Akinci 2022 Sun et al. 2022 Arabameri et al. 2020 Yavuz Ozalp et al. 2023
Distance to Faults (m)	0	7723.34	<b>1:</b> 0–1000, <b>2:</b> 1000–2000, <b>3:</b> 2000–3000, <b>4:</b> 3000–4000, <b>5:</b> 4000–5000, <b>6:</b> 5000–6000, <b>7:</b> 6000–7000, <b>8:</b> 7000–7723.34,	Feizizadeh et al. 2014 Akinci et al. 2020 Akinci 2022 Yavuz Ozalp et al. 2023
Distance to Roads (m)	0	2197.84	<b>1:</b> 0–200, <b>2:</b> 200–400, <b>3:</b> 400–600, <b>4:</b> 600–800, <b>5:</b> 800–1000, <b>6:</b> 1000–1200, <b>7:</b> 1200–1400, <b>8:</b> 1400–1600, <b>9:</b> 1600–1800, <b>10:</b> 1800–2197.84	Akinci 2022 Zhang et al. 2022a He et al. 2023 Yavuz Ozalp et al. 2023
Lithology	-	-	Explained in Table 1	
Land cover	-	-	<b>1:</b> Water, <b>2:</b> Trees, <b>3:</b> Grass, <b>5:</b> Crops, <b>6:</b> Scrub/shrub, <b>7:</b> Built Area, <b>8:</b> Bare ground, <b>9:</b> Snow/Ice, <b>10:</b> Clouds	Lv et al. 2022 Roy et al. 2023 Yavuz Ozalp et al. 2023 Yu et al. 2023
Plan curvature	-24.50	33.33	<b>1:</b> -24.50 – -0.001, <b>2:</b> -0.001 – 0.001, <b>3:</b> 0.001 – 33.33	Kilicoglu 2021 Akinci 2022 Yavuz Ozalp et al. 2023
Profile curvature	-40.17	36.42	<b>1:</b> -40.17 – -0.001, <b>2:</b> -0.001 – 0.001, <b>3:</b> 0.001 – 36.42	Kilicoglu 2021 Akinci 2022 Yavuz Ozalp et al. 2023
Slope (°)	0	76.33	<b>1:</b> 0–5, <b>2:</b> 5–10, <b>3:</b> 10–15, <b>4:</b> 15–20, <b>5:</b> 20–25, <b>6:</b> 25–30, <b>7:</b> 30–35, <b>8:</b> 35–40, <b>9:</b> 40–45, <b>10:</b> 45–76.33	Kilicoglu 2021 Akinci 2022 Sun et al. 2022 Yavuz Ozalp et al. 2023
Slope length	0	3577.64	<b>1:</b> 0–56.12, <b>2:</b> 56.12–140.30, <b>3:</b> 140.30–252.54, <b>4:</b> 252.54–392.84, <b>5:</b> 392.84–561.20, <b>6:</b> 561.20–785.68, <b>7:</b> 785.68–1080.31, <b>8:</b> 1080.31–1487.18, <b>9:</b> 1487.18–2090.47, <b>10:</b> 2090.47–3577.64	Hong et al. 2015 Akinci 2022 Ghasemian et al. 2022 Yavuz Ozalp et al. 2023
TPI	-70.88	79.58	<b>1:</b> -70.88 – -21.32, <b>2:</b> -21.32 – -13.65, <b>3:</b> -13.65 – -7.75, <b>4:</b> -7.75 – -3.03, <b>5:</b> -3.03 – 1.10, <b>6:</b> 1.10 – 5.23, <b>7:</b> 5.23 – 9.95, <b>8:</b> 9.95 – 15.85, <b>9:</b> 15.85 – 25.29, <b>10:</b> 25.29 – 79.58	Arabameri et al. 2020 Sahin 2020 Akinci 2022 Yavuz Ozalp et al. 2023
TWI	2.26	26.18	<b>1:</b> 2.26–4.79, <b>2:</b> 4.79–5.92, <b>3:</b> 5.92–6.86, <b>4:</b> 6.86–7.80, <b>5:</b> 7.80–8.92, <b>6:</b> 8.92–10.33, <b>7:</b> 10.33–12.11, <b>8:</b> 12.11–14.46, <b>9:</b> 14.46–18.12, <b>10:</b> 18.12–26.18	Sahin 2020 Kilicoglu 2021 Akinci 2022 Yavuz Ozalp et al. 2023

linear correlation between the conditioning factors. The strong linear relationship between the variables can cause prediction results to be inaccurate and decrease the model's accuracy (Song et al. 2023). In LS mapping studies, the most commonly used indicators for multicollinearity are tolerance (TOL) and variance inflation factor (VIF), which can be calculated using Eqs. 2 and 3 (Kavzoglu

et al. 2014; Bai et al. 2015; Arabameri et al. 2020; Wang et al. 2020a, b; Yi et al. 2020; Wei et al. 2022). The VIF is a measure of the increase in the variance of a regression coefficient due to multicollinearity. The TOL is the reciprocal of the VIF value and can also be used to test for multicollinearity between variables (Yu et al. 2023). In Eq. 2,  $R^2$  represents the proportion of variance in the

target variable (Ye et al. 2022). If the VIF value is greater than 10 or the TOL value is less than 0.1, it indicates a multicollinearity problem, and the multicollinear variables should be removed from the susceptibility models.

$$TOL = 1 - R^2 \tag{2}$$

$$VIF = \frac{1}{TOL} = \frac{1}{1 - R^2} \tag{3}$$

### Performance assessment metrics

Akinci and Akinci (2023) emphasized that validation is an essential procedure to evaluate the performance of the models. An LSM that has not been validated holds no scientific value. Sahin (2022) stated that different accuracy metrics can be used to evaluate the performance of ML models. In this study, overall accuracy, precision, recall, F1-score, RMSE and area under the receiver operating characteristic curve (AUC-ROC) metrics have been used to validate the results and compare and evaluate the performance of four different ML models. Except for RMSE, the performance assessment metrics given in Table 3 are calculated using the components of the confusion matrix.

In the equations in Table 3, TP (true positive) and TN (true negative) express the number of pixels correctly classified as landslide and non-landslide, respectively; FP (false positive) and FN (false negative) denote the number of pixels misclassified as landslide and non-landslide, respectively (Ye et al. 2022).

### SHapley additive exPlanation (SHAP)

In applications such as LS mapping, it is crucial to understand why and how a model makes a particular prediction, as well as prediction accuracy. However, the increasing arithmetic power and complexity of machine learning models makes it difficult to understand their internal mechanisms, local behaviors and decision-making processes (Zhang et al. 2023). The new generation of AI models, called explainable or interpretable AI (XAI), aims to explain the local behavior of black box models (Youssef et al. 2023). Lundberg and Lee (2017) proposed SHAP (SHapley Additive exPlanation) to provide explanations for the prediction reasons of various machine learning models, particularly opaque black box models. The SHAP method quantifies the impact of each feature on the model’s prediction. It achieves this by calculating the sum of the Shapley values of each input feature. This improves comprehension of how a model generates predictions (Zhang et al. 2023; Teke and Kavzoglu 2023).

## Results

### Multicollinearity analysis

Multicollinearity, in its simplest definition, is the presence of high correlation or linear relationship between independent variables in a regression model. Multicollinearity causes the results obtained from the model to be inaccurate. Therefore, multicollinearity analysis is applied in LS studies to test whether there is a high correlation between conditioning factors. This analysis has two natural results: i) there is no

**Table 3** Performance assessment metrics

Metric	Equation	Description
Overall accuracy (OA)	$OA = \frac{TP+TN}{TP+TN+FP+FN}$	The ratio of landslides and non-landslides that are correctly classified. This shows how well the landslide model works (Ghasemian et al. 2022)
Precision	$Precision = \frac{TP}{TP+FP}$	Precision, also called the positive predictive value, is the fraction of relevant instances (TP) amongst the retrieved instances (Azarafza et al. 2021)
Recall	$Recall = \frac{TP}{TP+FN}$	Recall (or sensitivity) is the measure at which pixels corresponding to landslides are correctly classified as a landslide having occurred. Its value is ideal as long as it is closer to 1 (Bravo-López et al. 2022)
F1-score	$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$	F1-score, one of the metrics commonly used in the evaluation of ML models, is actually a harmonic mean of precision and recall (Yu et al. 2023). The value of F1-score ranges from 0 to 1. A model is reliable if the value of F1-score is close to 1 (Ye et al. 2022)
AUC-ROC	$True\ positive\ rate\ (TPR) = \frac{TP}{TP + FN}$ $False\ positive\ rate\ (FPR) = \frac{FP}{FP + TN}$	The ROC curve plots the FPR on the X axis and the TPR on the Y axis. It shows the trade-off between the two rates (Pourghasemi et al. 2012). The area under the ROC curve (AUC) is an indicator to check the prediction performance of the model (Yilmaz 2009)
RMSE	$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_{obs} - y_{pred})^2}$	RMSE is used to measure the prediction errors of models (Nguyen et al. 2019). An RMSE value close to 0 indicates that the ML model has a good performance (Ado et al. 2022)

multicollinearity among the factors, ii) there is multicollinearity among some factors and the factors found to be correlated should be removed from the model. There are many studies with these two results in the literature. For example, in the studies conducted by He et al. (2023), Song et al. (2023), Vega et al. (2023) and Yu et al. (2023), it was found that there was no serious multicollinearity problem between conditioning factors.

Table 4 displays the outcomes of the multicollinearity analysis conducted on the conditioning factors employed in this study. The initial findings indicate a significant correlation between slope and TRI. Consequently, TRI was eliminated from the model, and a subsequent multicollinearity analysis was conducted on the remaining 13 factors. The highest VIF value 3.918890 and the lowest TOL value was 0.255174 (Table 4). Since the highest VIF value was less than 10 and the lowest TOL value is higher than 0.1, there was no multicollinearity problem between the conditioning factors used in this study.

In the LS study conducted by Yavuz Ozalp et al. (2023) in Ardeşen and Fındıklı districts of Rize (Turkey), multicollinearity analysis was performed for fifteen conditioning factors. As in this study, the researchers found that there was a high linear correlation between slope and TRI, and TRI was excluded from the susceptibility models. Wang et al. (2022) found collinearity between the factors of slope, elevation variation coefficient, surface cutting depth, relief amplitude, and surface roughness. Since slope is very effective in the occurrence of landslides, the other four factors were eliminated.

## Landslide susceptibility mapping

In this study, firstly, landslide susceptibility index (LSI) maps of the study area were produced using RF, CatBoost,

**Table 4** Multicollinearity Analysis Results

Landslide Conditioning Factors	VIF	TOL
Slope	1.694949	0.589988
Aspect	1.080109	0.925832
TWI	3.918890	0.255174
TPI	1.483154	0.674239
Altitude	1.410740	0.708848
Plan Curvature	1.301666	0.768246
Profile Curvature	1.140360	0.876916
Distance to Drainage	1.139817	0.877334
Distance to Faults	1.076383	0.929037
Distance to Roads	1.229125	0.813587
Lithology	1.294081	0.772749
Land Cover	1.031502	0.969460
Slope Length	3.059941	0.326804

XGBoost and LightGBM algorithms. The algorithms were implemented in python programming language using the scikit-learn library. The XGBoost algorithm was implemented with default hyperparameter values ( $n\_estimators=100$ ,  $max\_depth=6$ ,  $eta=0.3$ ,  $colsample\_bytree=1$ ,  $min\_child\_weight=1$ ,  $subsample=1$ ,  $gamma=0$ ). In order not to favour one algorithm over the others and to make an objective comparison, the number of trees was set to 100, the maximum tree depth to 0.6 and the learning rate to 0.3 for the other algorithms. The LSI shows the degree of susceptibility of each pixel in the study area to landslide formation. The higher the LSI value in a pixel, the higher the probability of landslide occurrence in that pixel, the lower the LSI, the lower the probability of landslide occurrence (Wubalem 2021). After the LSIs were created, they were transferred to ArcGIS software where they were reclassified and divided into five different susceptibility classes: very low, low, medium, high and very high. Thus, LSMs of the study area were obtained (Fig. 6). This classification was achieved by utilizing the natural breaks algorithm (Jenks 1967).

Table 5 presents the areal distribution ratios of susceptibility classes for each ML model. According to the areal distribution of susceptibility classes in Table 5, most of the study area is not susceptible to the landslides. 85.89% of the study area in RF, 71.86% of the study area in CatBoost, 70.16% of the study area in XGBoost and 77.31% of the study area in LightGBM is classified as very low and low susceptibility. Only 8.74% of the study area in RF, 12.85% of the study area in CatBoost, 11.37% of the study area in XGBoost and 10.43% of the study area in LightGBM is highly and very highly susceptible to the landslides.

## Performance assessment and comparison

In this study, various metrics including overall accuracy (OA), precision, recall,  $F1$ -score, RMSE and area under the receiver operating characteristic (ROC) curve (AUC) were employed to compare and evaluate the performance of the LS models. Model assessment metrics were applied for both training and validation stages. In the evaluation made in terms of OA, it was determined that the model with the lowest accuracy value for both training and validation stages was RF (Table 6). Although CatBoost, XGBoost and LightGBM have close accuracy values, LightGBM provided slightly better accuracy than other models in both training and validation stages. LightGBM also outperformed the other models in terms of precision, recall and  $F1$ -score. However, it is a known fact that ML models may perform differently depending on both the characteristics of the study area and the conditioning factors used.

One of the commonly used metrics to measure the accuracy of regression models is RMSE (Trinh et al. 2023). RMSE is used to measure the prediction error of the

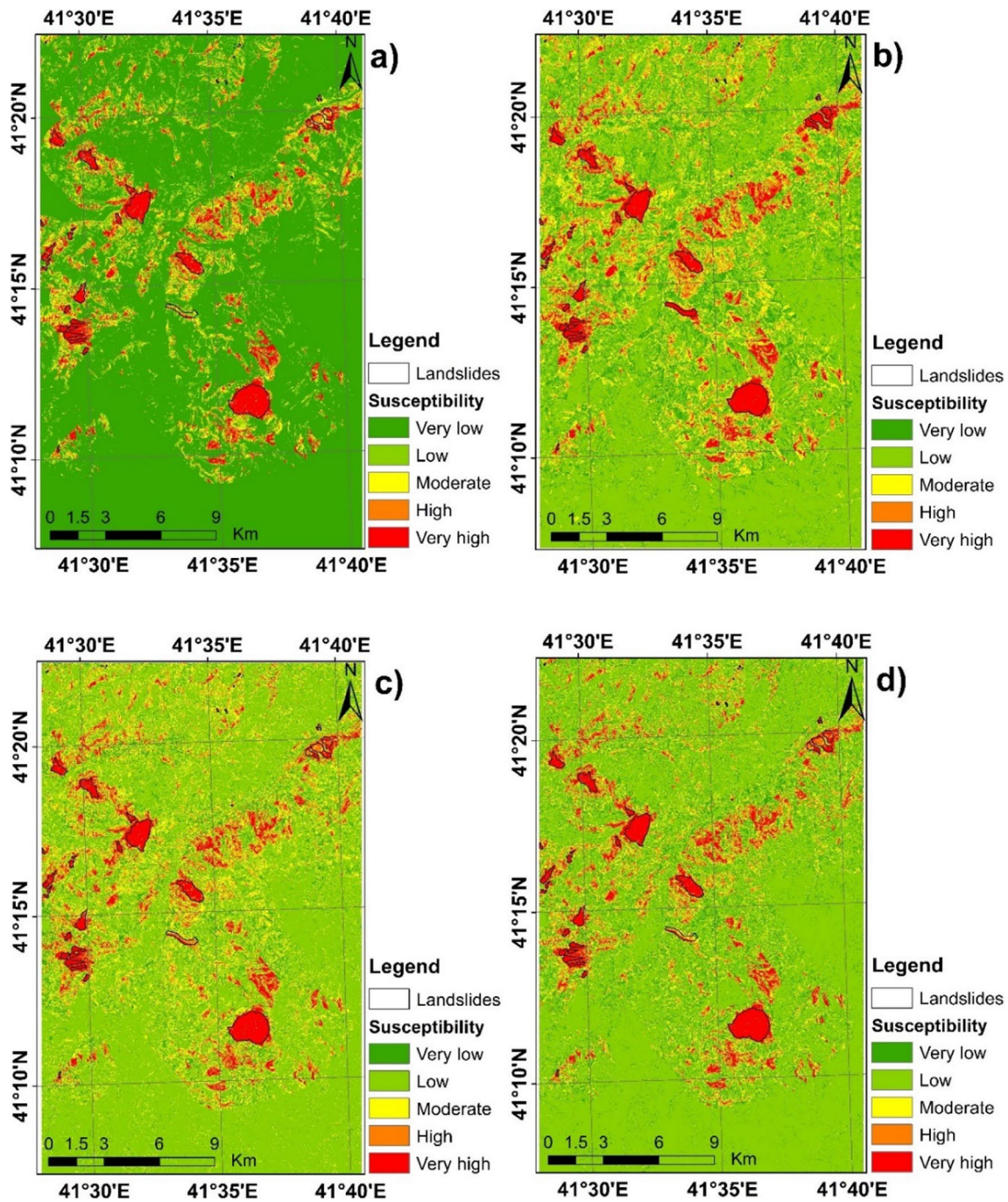


Fig. 6 LSMs: a) RF, b) CatBoost, c) XGBoost, d) LightGBM

model, and the closer it is to 0, the better the performance of the model (Nguyen et al. 2019; Ado et al. 2022; Trinh et al. 2023). Compared to precision, recall, and *F1*-score, XGBoost outperformed the other models with the lowest RMSE values.

In this study, the AUC metric, which is the most widely used performance assessment metric in LS mapping studies, was also used. Figure 7 shows the ROC curves and AUC

values of the models. When Fig. 7 is examined, it is seen that the XGBoost model has the highest AUC value (0.9773), followed by the LightGBM (0.9751), CatBoost (0.9708) and RF (0.8976) models, respectively. As with other evaluation metrics, RF lagged behind other models in terms of AUC. In the study by Sahin (2022), RF also lagged behind other ensemble learning algorithms such as CatBoost, XGBoost and LightGBM in terms of AUC value.

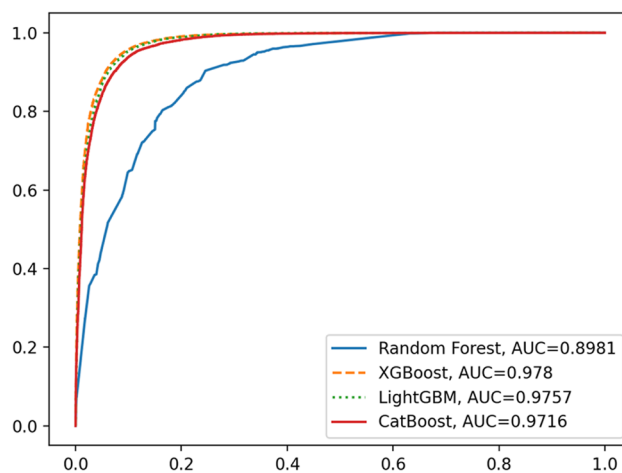
**Table 5** Comparison of the results of the ML models

Model	Susceptibility class	Landslide probability	PoA (%)	NLP	PLP (%)	FR
RF	Very low	0 – 0.08	76.50	753	0.815	0.0107
	Low	0.08 – 0.267	9.39	2303	2.491	0.2653
	Moderate	0.267 – 0.506	5.37	3566	3.857	0.7182
	High	0.506 – 0.773	4.21	10,829	11.714	2.7824
	Very high	0.773 – 1.000	4.53	74,995	81.123	17.9079
CatBoost	Very low	-0.816 – -0.057	15.95	6	0.007	0.0004
	Low	-0.057 – 0.101	55.91	29	0.031	0.0006
	Moderate	0.101 – 0.335	15.29	320	0.346	0.0226
	High	0.335 – 0.677	6.90	4309	4.661	0.6755
	Very high	0.677 – 1.311	5.95	87,782	94.955	15.9588
XGBoost	Very low	-1.214 – -0.093	8.77	127	0.138	0.0157
	Low	-0.093 – 0.059	61.39	490	0.530	0.0086
	Moderate	0.059 – 0.288	18.47	2165	2.342	0.1268
	High	0.288 – 0.658	6.44	10,517	11.376	1.7665
	Very high	0.658 – 1.561	4.93	79,147	85.614	17.3659
LightGBM	Very low	-0.811 – -0.063	13.64	158	0.171	0.0125
	Low	-0.063 – 0.099	63.67	968	1.047	0.0164
	Moderate	0.099 – 0.361	12.26	2060	2.228	0.1817
	High	0.361 – 0.721	5.18	7966	8.617	1.6635
	Very high	0.721 – 1.488	5.25	81,294	87.937	16.7499

PoA: Percentage of area, NLP: number of landslide pixels, PLP: Percentage of landslide pixels, FR: Frequency ratio

**Table 6** Performance of the ML models in training and validation stage

Stage	Metrics	RF	CatBoost	XGBoost	LightGBM
Training	Accuracy (%)	82.556	91.948	92.885	92.992
	Precision	0.82656	0.92143	0.92917	0.93124
	Recall	0.82553	0.91942	0.92884	0.92996
	F1-score	0.82541	0.91941	0.92883	0.92988
	RMSE	0.3572	0.2598	0.2411	0.2437
Validation	Accuracy (%)	82.338	91.831	92.641	92.765
	Precision	0.82433	0.92024	0.92667	0.92913
	Recall	0.82344	0.91845	0.92643	0.92755
	F1-score	0.82326	0.91828	0.92640	0.92759
	RMSE	0.3576	0.2628	0.2469	0.2481



**Fig. 7** ROC curves and AUC values of the models

**Interpretation of models with SHAP**

In LS mapping studies, it has been observed that some factors may have minimum importance in model predictions, while others may be more effective in this process. Yu et al. (2023) emphasized that determining the relative importance of conditioning factors can help to better understand the causes of landslides in a region. The importance values of conditioning factors for RF, CatBoost, XGBoost, and

LightGBM are presented in Fig. 8. The importance values of the conditioning factors vary between the different models. However, lithology, altitude, distance to faults, and aspect consistently rank in the top 4, while slope length, TWI, plan curvature, and profile curvature rank in the bottom 4.

The effects of conditioning factors on model outputs are interpreted not only by interpreting Fig. 8 but also by using SHAP values. Kavzoglu and Teke (2022) emphasized that

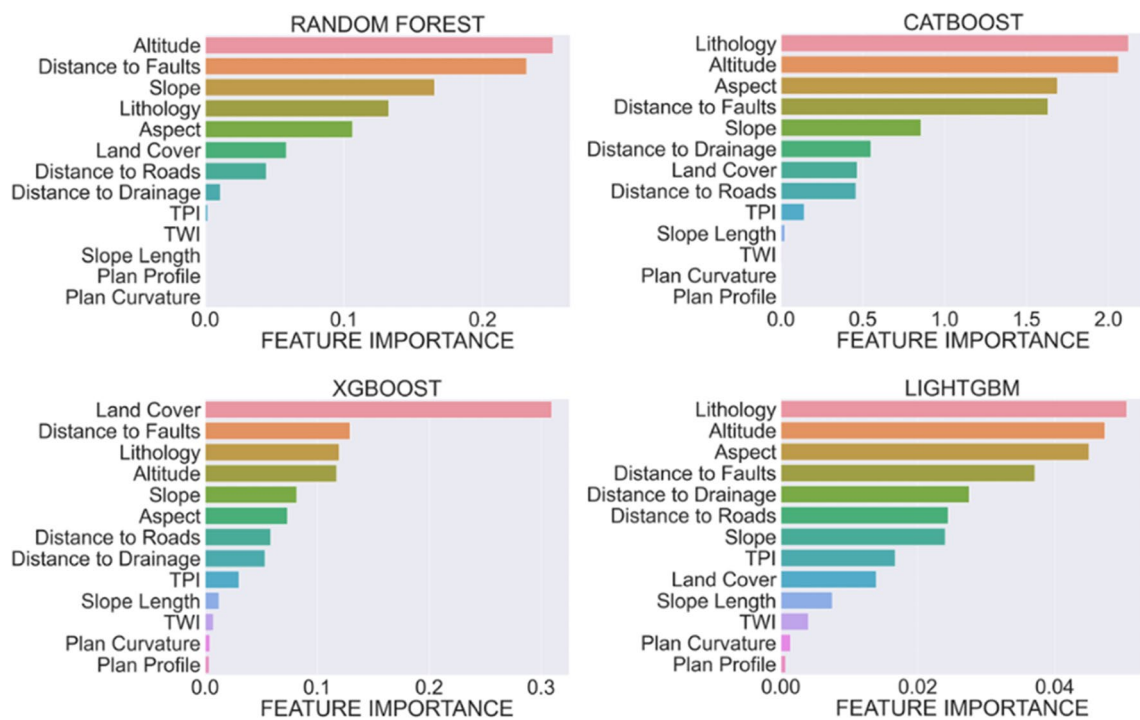


Fig. 8 Importance of the conditioning factors

unlike the feature importance function of ML algorithms, SHAP can determine whether the factors contribute positively or negatively to the model outputs. SHAP summary plot, one of the graphs offered by the SHAP library, is used to visualize the effect of each factor of the model on the prediction. Figure 9 and Fig. 10 shows the SHAP values of the thirteen conditioning factors for RF, CatBoost, XGBoost and LightGBM. As can be seen in the SHAP summary plots, although the effects of the conditioning factors on the model outputs vary in different models, altitude, lithology, aspect, distance to faults and slope were found to be the most effective factors. On the other hand, SHAP values revealed that slope length, TWI, profile curvature and plan curvature were least effective compared to other landslide conditioning factors. The results appear to be partially consistent with previous studies. For example, in the LS mapping study carried out by Akinci (2022) in Arhavi, Hopa and Kemalpaşa districts of Artvin, it was determined that TWI, slope length and curvature were the least important factors in the occurrence of landslides. In the study conducted by Akinci and Zeybek (2021), landslide susceptibility maps of Ardanuç district of Artvin province were produced using LR, SVM and RF models. In this study, the researchers determined that lithology, altitude and distance to the road parameters were the most effective factors, while TWI and curvature parameters were the least effective factors. In the study by Can et al. (2021), the predictions of the XGBoost-based landslide susceptibility model were interpreted using the SHAP summary plot. Similar to this study, the researchers

determined that lithology and altitude have a greater impact on the prediction results, while profile curvature has the least impact on the prediction results.

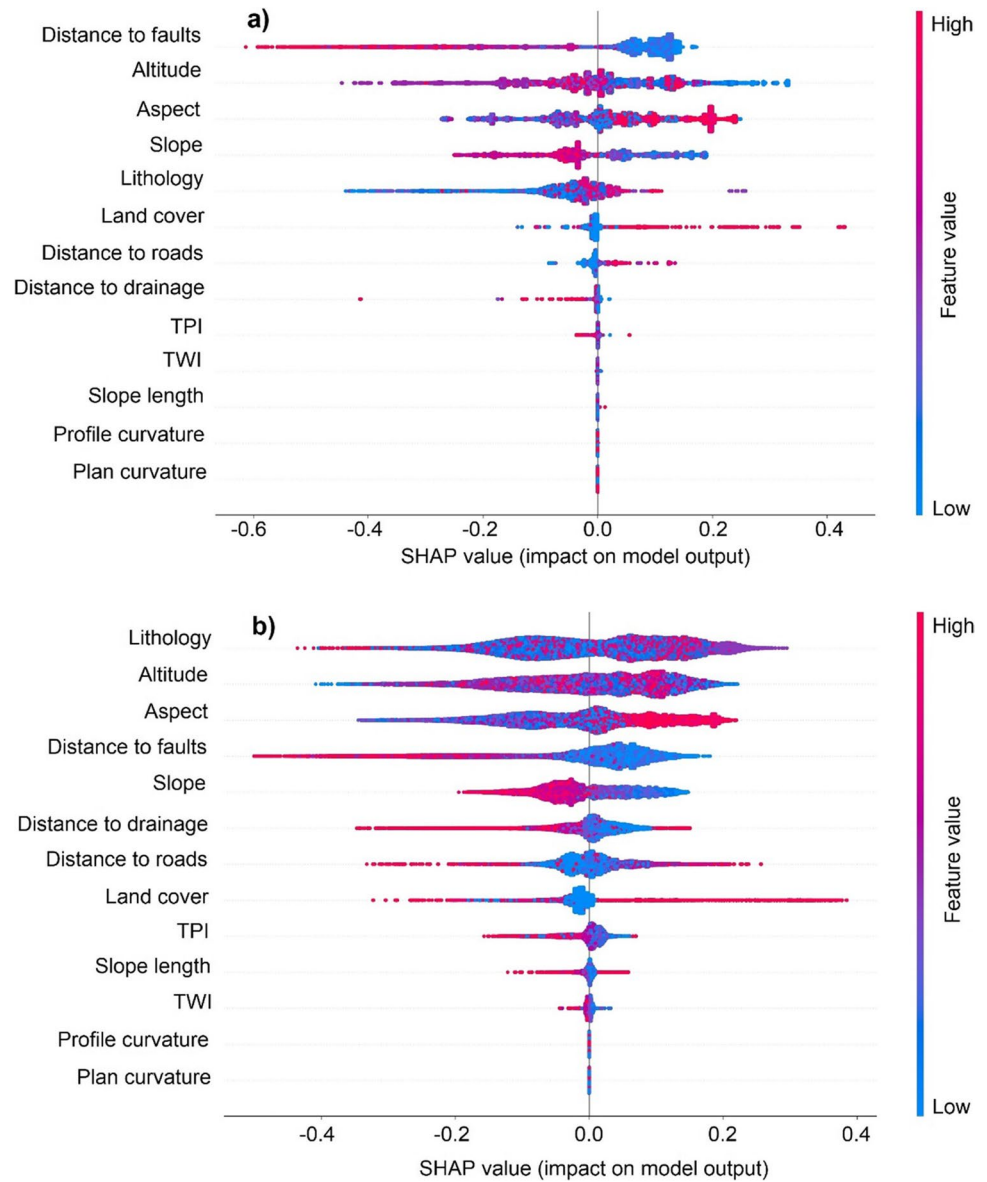
On the other hand, we interpreted the local contribution of the conditioning factors to the model outputs using SHAP waterfall plots (refer to Fig. 11 and Fig. 12). In the waterfall plot, the red arrow indicates that the SHAP value of the factor is greater than zero, meaning that the factor provides a positive gain to the landslides, while the blue arrow indicates a negative gain (Chang et al. 2023). Figure 11 shows that the distance to faults in the RF model and aspect in the CatBoost model had the largest negative impact on the occurrence of landslides in the study area. Lithology was also found to be a significant factor in both models. According to Fig. 12, altitude, aspect, distance to faults, distance to drainage and lithology in the XGBoost model (Fig. 12a) and distance to faults, altitude, land cover, aspect and slope in the LightGBM model (Fig. 12b) were determined as the factors with the largest positive contribution to the occurrence of landslides in the study area.

## Discussion

### Comparison of factors

ML algorithms have different decision-making mechanisms. Therefore, in ML models using the same factors,

**Fig. 9** SHAP summary plots: a) RF, b) CatBoost



the contributions of the factors to the model predictions are also different. However, ML models can only indicate the relative importance of factors on prediction results through feature importance functions. This function is insufficient to interpret how the models work when making predictions, which is why ML models are often referred to as “black box” models. To overcome this problem and explain the decision-making mechanism of ML-based landslide susceptibility models, methods such as SHAP should be used (Lu et al. 2023; Pradhan et al. 2023). The interpretability of a model is directly proportional to the ease of understanding the reasoning behind a specific decision. For this reason, we see that the SHAP method is used in studies such as LS mapping (Pradhan et al. 2023; Vega et al. 2023; Zhang et al. 2023) and wildfire susceptibility mapping (Iban and Sekertekin 2022).

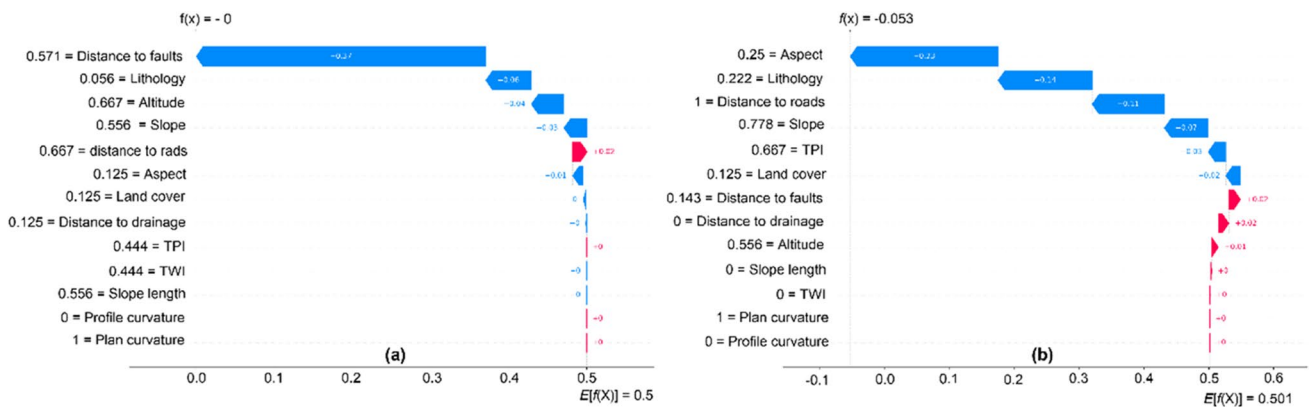
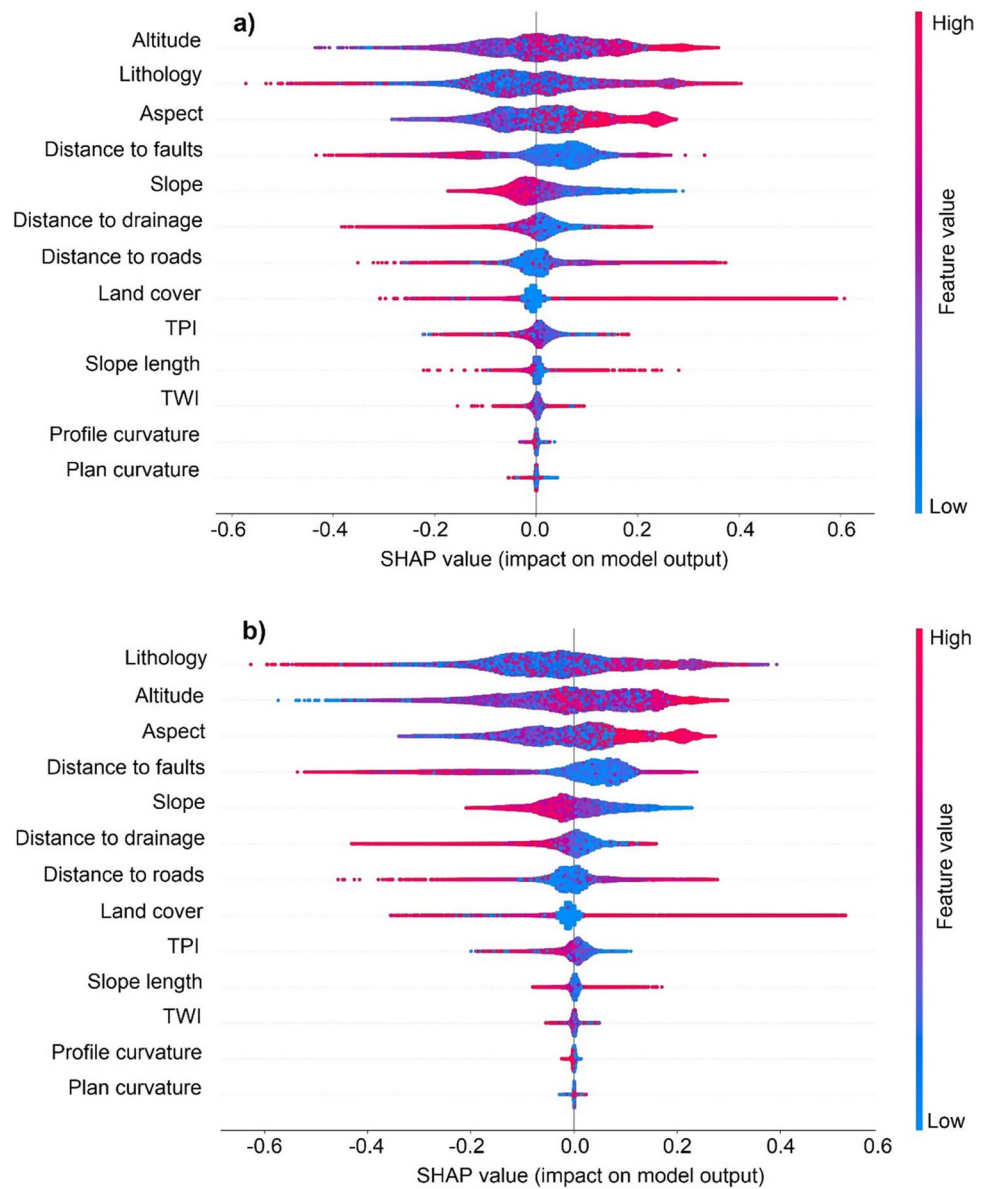
When SHAP summary plots for XGBoost and LightGBM models are analyzed, it is seen that elevation and lithology are the most effective factors in model outputs (Fig. 10). This means that areas with more anthropogenic activities and weak lithological units are prone to landslides. Agricultural areas are often concentrated at specific altitudes and aspects within settlements, leading to increased human activity in these regions. Uncontrolled irrigation and excavation works can trigger landslides in these areas.

### Comparison of models

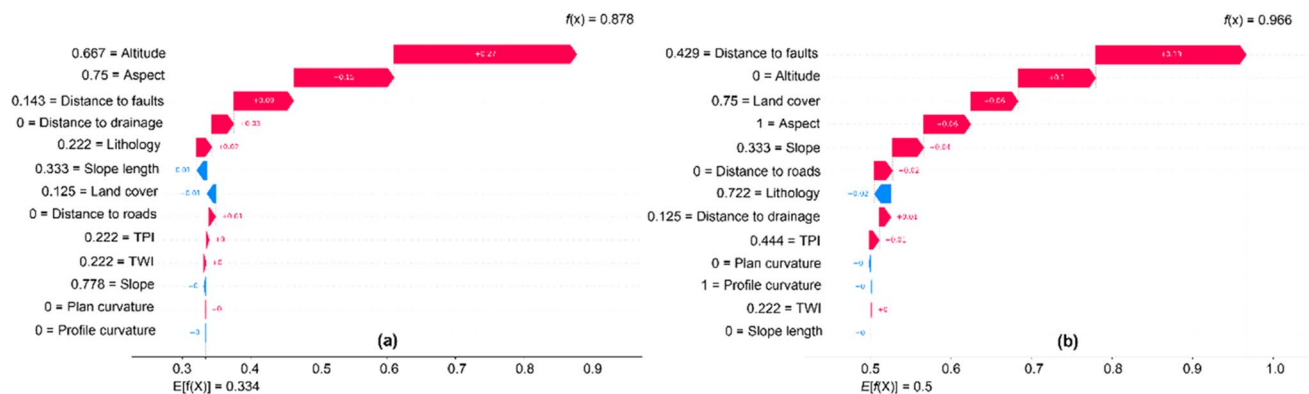
In this study, the prediction performance of ensemble learning algorithms were evaluated using accuracy metrics (OA, F1-score, precision, recall, RMSE and AUC) commonly used in LS mapping studies. There are many recent studies in



**Fig. 10** SHAP summary plots:  
**a)** XGBoost, **b)** LightGBM



**Fig. 11** SHAP waterfall plots: **a)** RF, **b)** CatBoost



**Fig. 12** SHAP waterfall plots: **a)** XGBoost, **b)** LightGBM

the LS mapping literature that use these metrics to assess the accuracy of susceptibility models (Kavzoglu and Teke 2022; He et al. 2023; Sun et al. 2023a, b; Vega et al. 2023; Yu et al. 2023; Zhang et al. 2023). Kavzoglu and Teke (2022) suggest that OA is a reliable metric for measuring model robustness in LS mapping studies. Ye et al. (2022) explain that the  $F1$ -score ranges from 0 to 1, and a value close to 1 indicates model reliability. As with OA, the model with the lowest  $F1$ -score was again the RF (Table 6). It was seen that the highest  $F1$ -score belonged to the LightGBM model and this model had a high classification capacity for both the training and validation dataset. In summary, LightGBM was found to be slightly superior to other models in all metrics except AUC. In the LS mapping study conducted by Sun et al. (2023b), OA, precision, recall and  $F1$ -score metrics were used to evaluate the accuracy of XGBoost and LightGBM models, and as in this study, LightGBM gave better results than XGBoost in all metrics. Tree-based ensemble learning algorithms are widely used in flood and forest fire susceptibility mapping apart from landslide susceptibility. For example, Saber et al. (2022) was used RF, LightGBM and CatBoost models for flash flood susceptibility prediction and determined that LightGBM outperformed other models in terms of evaluation metrics and processing time.

Another statistical criterion used in the study was AUC, which represents the area under the ROC curve. The AUC value of an ROC curve ranges from 0.5 to 1, and an AUC value close to 1 indicates that the model is excellent (Sahin 2022). According to Chen et al. (2017), Jiao et al. (2019), and Wu et al. (2020), the AUC value is generally classified in 5 ways: poor (0.5–0.6), average (0.6–0.7), good (0.7–0.8), very good (0.8–0.9), and excellent (0.9–1). According to this classification, RF performed very good while CatBoost, LightGBM and XGBoost performed excellent. When evaluated in terms of AUC, it is seen that this study is consistent with other studies in the literature. Because in many studies in the literature, it has been stated that ensemble learning

algorithms are more successful than single ML algorithms (Kavzoglu and Teke 2022).

On the other hand, in order to assess the practical usability of an LSM, it is important to consider the accuracy of the ML model as well as the rationality of the generated LSM. To ensure rationality, an LSM needs to be able to accurately classify existing landslides in the study area. In this study, the LSMs generated were compared with the LIM, and the distribution of landslide pixels across the susceptibility classes in the LSMs was determined (Table 5). Such an LSM is considered rational if the existing landslides in the LIM remain in areas of high susceptibility as much as possible, and the areas of very high susceptibility in the LSM are as small as possible (Guo et al. 2021). In the RF, CatBoost, XGBoost and LightGBM models, the percentages of very high susceptible areas were determined as 4.53%, 5.95%, 4.93% and 5.25%, and percentages of landslides were determined as 81.123%, 94.955%, 85.614% and 87.937%, respectively. The frequency ratios were 17.9079, 15.9588, 17.3659, and 16.7499 (Table 5). LSMs showed a similar trend in general. As the susceptibility levels increased, the frequency ratios also tended to increase. In all models, the highest number of landslides was observed in the very high susceptible class. This comparison showed that the landslide occurrence rate gradually increased from very low prone areas to very high prone areas and the LSMs produced were reasonable. However, the LSM produced by the CatBoost model seems to be more rational and reasonable than other models. Because in this model, approximately 99.6% of the existing landslides remain in high and very highly landslide susceptible areas. In the LS mapping study by Yavuz Ozalp et al. (2023), RF, GBM, XGBoost and CatBoost algorithms were used. The researchers also evaluated the rationality of the LSMs produced by these algorithms. The evaluation results showed that the LSMs produced by CatBoost and XGBoost models are reasonable and rational for the study area.

## Limitations and future suggestions

The ML models used in the study were able to predict the probability of landslide occurrence in the study area well. However, it is considered that the study has an important limitation. As it is known, ML models need accurate, up-to-date and complete data during the training phase. If the data is not complete or up-to-date, the model may perform poorly because it is not trained correctly. In this context, landslide inventory data are of vital importance in ML-based landslide susceptibility mapping studies. The most important constraint in this study is related to the up-to-dateness of landslide inventory data. Because Artvin Provincial Disaster and Emergency Directorate last updated the landslide inventory map of the study area in November 2016. Since landslides that occurred in recent years were not included in the inventory map, the models may have been trained with missing data. This situation was also mentioned in the Artvin Provincial Disaster Risk Reduction Plan and updating the landslide inventory maps was added to the action plan. Therefore, future studies in the study area should primarily focus on updating landslide inventory maps using field surveys and high-resolution satellite images. In addition, deep learning algorithms such as CNN and RNN can be used in future studies to produce highly accurate landslide susceptibility maps and the performance of deep learning algorithms can be compared with tree-based ensemble learning algorithms.

## Conclusion

Artvin is a province prone to landslides due to its highly rugged topography. Landslides constitute a large part of the natural disasters occurring throughout the province. In the provincial disaster risk reduction plan, it is stated that landslides are the priority disaster type in the region and various strategic actions are proposed for risk reduction. Landslides are observed in all districts of Artvin. Landslides in the region cause damage to transport networks, energy transmission and drinking water lines, degradation in agricultural and forest areas, damage to buildings and loss of life. Therefore, landslide susceptibility mapping is of crucial importance to reduce the damages caused by landslides in all districts of Artvin province. In this study, the landslide susceptibility of Murgul district of Artvin province was analyzed. Four different ML models, random forest, XGBoost, LightGBM and CatBoost were employed to generate LSMs. The models considered multiple factors, including land cover, aspect, slope, lithology, elevation, plan curvature, profile curvature, distance to drainage networks, distance to faults, distance to roads, slope length, topographic position index (TPI), and topographic wetness index (TWI). The LIM provided by the

GDMRE and Artvin Provincial Directorate of Disaster and Emergency, which documented 54 landslide polygons, was utilized in the analysis. The resulting LSMs classified the study area into five classes: very low, low, medium, high, and very high, based on the natural breaks (jenks) classification. The performance of these models was evaluated using metrics like overall accuracy, precision, recall, *F1*-score, RMSE and AUC-ROC. The models that performed the best were XGBoost in terms of RMSE and AUC values, and LightGBM in terms of accuracy and *F1*-score. The LSMs produced through XGBoost and LightGBM are highly valuable for landslide risk assessment and risk mitigation studies in the study area. The maps produced can guide engineers for disaster and emergency planning, decision makers and urban planners for sustainable land use planning and insurance companies for natural disaster insurance.

**Author contribution** Ziya Usta and Alper Tunga Akın performed algorithmic development stages. Ziya Usta and Halil Akıncı prepared the data and the figures. The content of the article was determined by Halil Akıncı. Ziya Usta, Halil Akıncı and Alper Tunga Akın wrote the main article text.

**Funding** Open access funding provided by the Scientific and Technological Research Council of Türkiye (TÜBİTAK). This study was supported by the Scientific Research Projects Office of Artvin Çoruh University (AÇÜBAP) (Scientific Research Project No. 2022. F40.02.01).

**Data availability** The datasets used and/or analyzed during the current study are available from the corresponding author upon reasonable request.

## Declarations

**Conflict of interest** The authors declare no conflict of interest.

**Competing interests** The authors declare no competing interests.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Achour Y, Pourghasemi HR (2020) How do machine learning techniques help in increasing accuracy of landslide susceptibility maps? *Geosci Front* 11:871–883
- Adition A, Kubota T, Shinohara Y (2018) Comparison of GIS-based landslide susceptibility models using frequency ratio, logistic

- regression, and artificial neural network in a tertiary region of Ambon, Indonesia. *Geomorphology* 318:101–111
- Ado M, Amitab K, Maji AK, Jasińska E, Gono R, Leonowicz Z, Jasiński M (2022) Landslide Susceptibility Mapping Using Machine Learning: A Literature Survey. *Remote Sens* 14:3029. <https://doi.org/10.3390/rs14133029>
- AFAD (2020) Overview of 2019 within the scope of disaster management and statistics on natural events, Ministry of Interior of the Republic of Türkiye, Disaster and Emergency Management Presidency, [https://www.afad.gov.tr/kurumlar/afad.gov.tr/e\\_kutuphane/kurumsal-raporlar/afet\\_istatistikleri\\_2020\\_web.pdf](https://www.afad.gov.tr/kurumlar/afad.gov.tr/e_kutuphane/kurumsal-raporlar/afet_istatistikleri_2020_web.pdf). Accessed 19 Jan 2023
- Akgun A, Dag S, Bulut F (2008) Landslide susceptibility mapping for a landslide-prone area (Findikli, NE of Turkey) by likelihood-frequency ratio and weighted linear combination models. *Environ Geol* 54:1127–1143
- Akinci H (2022) Assessment of rainfall-induced landslide susceptibility in Artvin, Turkey using machine learning techniques. *J Afr Earth Sc* 191:104535. <https://doi.org/10.1016/j.jafrearsci.2022.104535>
- Akinci HA, Akinci H (2023) Machine learning based forest fire susceptibility assessment of Manavgat district (Antalya). *Turkey Earth Science Informatics* 16(1):397–414
- Akinci H, Yavuz Ozalp A (2021) Landslide susceptibility mapping and hazard assessment in Artvin (Turkey) using frequency ratio and modified information value model. *Acta Geophys* 69:725–745
- Akinci H, Zeybek M (2021) Comparing classical statistic and machine learning models in landslide susceptibility mapping in Ardanc (Artvin). *Turkey Nat Hazards* 108:1515–1543. <https://doi.org/10.1007/s11069-021-04743-4>
- Akinci H, Kilicoglu C, Dogan S (2020) Random Forest-Based Landslide Susceptibility Mapping in Coastal Regions of Artvin. *Turkey ISPRS Int J Geo-Inf* 9:553. <https://doi.org/10.3390/ijgi9090553>
- Akinci H, Zeybek M, Dogan S (2021) Evaluation of landslide susceptibility of Şavşat District of Artvin Province (Turkey) using machine learning techniques. In: *Landslides*. IntechOpen
- Arabameri A, Saha S, Roy J, Chen W, Blaschke T, Tien Bui D (2020) Landslide susceptibility evaluation and management using different machine learning methods in the Gallicash River Watershed. *Iran Remote Sensing* 12(3):475. <https://doi.org/10.3390/rs12030475>
- AWS (2024) CatBoost hyperparameters. Amazon SageMaker: Developer Guide. <https://docs.aws.amazon.com/sagemaker/latest/dg/catboost-hyperparameters.html>. Accessed 29 Jan 2024
- Azarafza M, Azarafza M, Akgün H, Atkinson PM, Derakhshani R (2021) Deep learning-based landslide susceptibility mapping. *Sci Rep* 11:24112. <https://doi.org/10.1038/s41598-021-03585-1>
- Bai SB, Lu P, Wang J (2015) Landslide susceptibility assessment of the Youfang catchment using logistic regression. *J Mt Sci* 12:816–827
- Bravo-López E, Fernández Del Castillo T, Sellers C, Delgado-García J (2022) Landslide susceptibility mapping of landslides with artificial neural networks: Multi-approach analysis of backpropagation algorithm applying the neuralnet package in Cuenca. *Ecuador Remote Sensing* 14(14):3495. <https://doi.org/10.3390/rs14143495>
- Breiman L (2001) Random forests. *Mach Learn* 45:5–32
- Can R, Kocaman S, Gokceoglu C (2021) A Comprehensive Assessment of XGBoost Algorithm for Landslide Susceptibility Mapping in the Upper Basin of Ataturk Dam. *Turkey Appl Sci* 11:4993. <https://doi.org/10.3390/app11114993>
- Cao J, Zhang Z, Du J, Zhang L, Song Y, Sun G (2020) Multi-geohazards susceptibility mapping based on machine learning—A case study in Jiuzhaigou. *China Nat Hazards* 102:851–871
- Catani F, Lagomarsino D, Segoni S, Tofani V (2013) Landslide susceptibility estimation by random forests technique: sensitivity and scaling issues. *Nat Hazards Earth Syst Sci* 13(11):2815–2831
- Chang L, Xing G, Yin H, Fan L, Zhang R, Zhao N, Huang F, Ma J (2023) Landslide susceptibility evaluation and interpretability analysis of typical loess areas based on deep learning. *Natural Hazards Research* 3:155–169. <https://doi.org/10.1016/j.nhres.2023.02.005>
- Chen W, Xie X, Wang J, Pradhan B, Hong H, Tien Bui D, Duan Z, Ma JA (2017) comparative study of logistic model tree, random forest, and classification and regression tree models for spatial prediction of landslide susceptibility. *CATENA* 151:147–160
- 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* 644:1006–1018
- Chen T, Zhu L, Niu RQ, Trinder CJ, Peng L, Lei T (2020) Mapping landslide susceptibility at the Three Gorges Reservoir, China, using gradient boosting decision tree, random forest and information value models. *J Mt Sci* 17:670–685
- Chen T, Guestrin C (2016) Xgboost: A scalable tree boosting system. In: *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pp 785–794
- Colkesen I, Sahin EK, Kavzoglu T (2016) Susceptibility mapping of shallow landslides using kernel-based Gaussian process, support vector machines and logistic regression. *J Afr Earth Sc* 118:53–64. <https://doi.org/10.1016/j.jafrearsci.2016.02.01>
- Dağ S, Akgün A, Kaya A, Alemdağ S, Bostancı HT (2020) Medium scale earthflow susceptibility modelling by remote sensing and geographical information systems based multivariate statistics approach: an example from Northeastern Turkey. *Environ Earth Sci* 79:468. <https://doi.org/10.1007/s12665-020-09217-7>
- Dai L, Zhu M, He Z, He Y, Zheng Z, Zhou G (2021) Landslide risk classification based on ensemble machine learning. In: *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, pp 3924–3927
- Dalkes M, Korkmaz MS (2023) Comparison of Analytic Hierarchy Process and Frequency Ratio Methods in Landslide Susceptibility Analysis: Example of Akçaabat and Düzköy districts of Trabzon province. *Journal of Natural Hazards and Environment* 9(1):16–38. <https://doi.org/10.21324/dacd.1105000>
- Das S, Sarkar S, Kanungo DP (2023) A critical review on landslide susceptibility zonation: recent trends, techniques, and practices in Indian Himalaya. *Nat Hazards* 115:23–72
- Demir G (2019) GIS-based landslide susceptibility mapping for a part of the North Anatolian Fault Zone between Reşadiye and Koyulhisar (Turkey). *CATENA* 183:104211. <https://doi.org/10.1016/j.catena.2019.104211>
- Duman TY, Çan T (2023) Characteristics of landslides and assessment of deep-seated landslide susceptibility in Northern Turkey. *Characteristics of landslides and assessment of deep-seated landslide susceptibility in Northern Turkey*. *Mediterranean Geoscience Reviews* 5:131–157. <https://doi.org/10.1007/s42990-023-00105-3>
- Erener A, Mutlu A, Düzgün HS (2016) A comparative study for landslide susceptibility mapping using GIS-based multi-criteria decision analysis (MCDA), logistic regression (LR) and association rule mining (ARM). *Eng Geol* 203:45–55
- Fang M, Chen Y, Xue R, Wang H, Chakraborty N, Su T, Dai Y (2021) A hybrid machine learning approach for hypertension risk prediction. *Neural Comput Appl* 35:14487–14497. <https://doi.org/10.1007/s00521-021-06060-0>
- Feizizadeh B, Blaschke T, Nazmfar H (2014) GIS based ordered weighted averaging and Dempster-Shafer methods for landslide susceptibility mapping in the Urmia Lake Basin. *Iran*

- Int. J. Digit. Earth 7:688–708. <https://doi.org/10.1080/17538947.2012.749950>
- Ghasemian B, Shahabi H, Shirzadi A, Al-Ansari N, Jaafari A, Geertsema M et al (2022) Application of a novel hybrid machine learning algorithm in shallow landslide susceptibility mapping in a mountainous area. *Front Environ Sci* 10:897254. <https://doi.org/10.3389/fenvs.2022.897254>
- 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>
- Gomez H, Kavzoglu T (2005) Assessment of shallow landslide susceptibility using artificial neural networks in Jabonosa River Basin. *Venezuela Eng Geol* 78(1–2):11–27. <https://doi.org/10.1016/j.enggeo.2004.10.004>
- Guo Z, Shi Y, Huang F, Fan X, Huang J (2021) Landslide susceptibility zonation method based on C5.0 decision tree and K-means cluster algorithms to improve the efficiency of risk management. *Geoscience Frontiers* 12(6):101249. <https://doi.org/10.1016/j.gsf.2021.101249>
- He W, Chen G, Zhao J, Lin Y, Qin B, Yao W, Cao Q (2023) Landslide Susceptibility Evaluation of Machine Learning Based on Information Volume and Frequency Ratio: A Case Study of Weixin County. *China Sensors* 23:2549. <https://doi.org/10.3390/s23052549>
- Hong H, Xu C, Tien Bui D (2015) Landslide susceptibility assessment at the Xiushui area (China) using frequency ratio model. *Procedia Environ Sci* 15:513–517. <https://doi.org/10.1016/j.proeps.2015.08.065>
- Iban MC, Sekertekin A (2022) Machine learning based wildfire susceptibility mapping using remotely sensed fire data and GIS: A case study of Adana and Mersin provinces. *Turkey Ecological Informatics* 69:101647. <https://doi.org/10.1016/j.ecoinf.2022.101647>
- Jenks GF (1967) The data model concept in statistical mapping. *Int Yearb Cartogr* 7:186–190
- Jiao Y, Zhao D, Ding Y, Liu Y, Xu Q, Qiu Y et al (2019) Performance evaluation for four GIS-based models purposed to predict and map landslide susceptibility: A case study at a World Heritage site in Southwest China. *CATENA* 183:104221. <https://doi.org/10.1016/j.catena.2019.104221>
- Kavzoglu T, Teke A (2022) Predictive Performances of Ensemble Machine Learning Algorithms in Landslide Susceptibility Mapping Using Random Forest, Extreme Gradient Boosting (XGBoost) and Natural Gradient Boosting (NGBoost). *Arab J Sci Eng* 47:7367–7385. <https://doi.org/10.1007/s13369-022-06560-8>
- Kavzoglu T, Sahin EK, Colkesen I (2014) Landslide susceptibility mapping using GIS-based multi-criteria decision analysis, support vector machines, and logistic regression. *Landslides* 11:425–439. <https://doi.org/10.1007/s10346-013-0391-7>
- Kaya Topaçlı Z, Özcan AK, Gökçeoğlu C (2024) Performance Comparison of Landslide Susceptibility Maps Derived from Logistic Regression and Random Forest Models in the Bolaman Basin. *Türkiye Nat Hazards Rev* 25(1):04023054. <https://doi.org/10.1061/NHREFO.NHENG-1771>
- Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, Ye Q, Liu T-Y (2017) LightGBM: A highly efficient gradient boosting decision tree. In: *Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS 2017)*, December 4–9, Long Beach, CA, USA, pp 3149–3157
- Keskin I (2013a) 1:100,000 Scale Geological Map of Turkey, No:178 Artvin-F46 Map Sheet. General Directorate of Mineral Research and Exploration, Geological Research Department, Ankara, Turkey. (in Turkish)
- Keskin I (2013b) 1:100,000 Scale Geological Map of Turkey, No:179 Artvin-E47 and F47 Map Sheet. General Directorate of Mineral Research and Exploration, Geological Research Department, Ankara, Turkey. (in Turkish)
- Kilicoglu C (2021) Investigation of the effects of approaches used in the production of training and validation data sets on the accuracy of landslide susceptibility mapping models: Samsun (Turkey) example. *Arab J Geosci* 14:2106. <https://doi.org/10.1007/s12517-021-08312-8>
- Lee S, Choi J (2004) Landslide susceptibility mapping using GIS and the weight-of-evidence model. *Int J Geogr Inf Sci* 18(8):789–814. <https://doi.org/10.1080/13658810410001702003>
- Liu Q, Tang A, Huang Z, Sun L, Han X (2022) Discussion on the tree-based machine learning model in the study of landslide susceptibility. *Nat Hazards* 113:887–911. <https://doi.org/10.1007/s11069-022-05329-4>
- Lu J, Ren C, Yue W, Zhou Y, Xue X, Liu Y, Ding C (2023) Investigation of Landslide Susceptibility Decision Mechanisms in Different Ensemble-Based Machine Learning Models with Various Types of Factor Data. *Sustainability* 15:13563. <https://doi.org/10.3390/su151813563>
- Lundberg SM, Lee S-I (2017) A unified approach to interpreting model predictions. In: *Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS 2017)*, December 4–9, Long Beach, CA, USA, pp 4768–4777
- Lv L, Chen T, Dou J, Plaza A (2022) A hybrid ensemble-based deep-learning framework for landslide susceptibility mapping. *Int J Appl Earth Obs Geoinf* 108:102713. <https://doi.org/10.1016/j.jag.2022.102713>
- Nguyen VV, Pham BT, Vu BT, Prakash I, Jha S, Shahabi H, Shirzadi A, Ba DN, Kumar R, Chatterjee JM, Tien Bui D (2019) Hybrid Machine Learning Approaches for Landslide Susceptibility Modeling. *Forests* 10:157. <https://doi.org/10.3390/f10020157>
- Omotehinwa TO, Oyewola DO, Dada EG (2023) A Light Gradient-Boosting Machine algorithm with Tree-Structured Parzen Estimator for breast cancer diagnosis. *Healthcare Analytics* 4:100218. <https://doi.org/10.1016/j.health.2023.100218>
- Park S, Kim J (2019) Landslide susceptibility mapping based on random forest and boosted regression tree models, and a comparison of their performance. *Appl Sci* 9(5):942. <https://doi.org/10.3390/app9050942>
- Pham K, Kim D, Le CV, Choi H (2022) Dual tree-boosting framework for estimating warning levels of rainfall-induced landslides. *Landslides* 19(9):2249–2262
- Pourghasemi HR, Rahmati O (2018) Prediction of the landslide susceptibility: which algorithm, which precision? *CATENA* 162:177–192
- Pourghasemi HR, Gayen A, Park S, Lee CW, Lee S (2018) Assessment of Landslide-Prone Areas and Their Zonation Using Logistic Regression, LogitBoost, and Naïve Bayes Machine-Learning Algorithms. *Sustainability* 10:3697. <https://doi.org/10.3390/su10103697>
- Pradhan B (2013) A comparative study on the predictive ability of the decision tree, support vector machine and neuro-fuzzy models in landslide susceptibility mapping using GIS. *Comput Geosci* 51:350–365. <https://doi.org/10.1016/j.cageo.2012.08.023>
- Pradhan B, Lee S (2010) Landslide susceptibility assessment and factor effect analysis: backpropagation artificial neural networks and their comparison with frequency ratio and bivariate logistic regression modelling. *Environ Model Softw* 25:747–759
- Pradhan B, Dikshit A, Lee S, Kim H (2023) An explainable AI (XAI) model for landslide susceptibility modeling. *Appl Soft Comput* 142:110324. <https://doi.org/10.1016/j.asoc.2023.110324>

- Prokhorenkova L, Gusev G, Vorobev A, Dorogush AV, Gulina A (2018) CatBoost: Unbiased boosting with categorical features. In: Proceedings of the 32nd Conference on Neural Information Processing Systems (NeurIPS 2018), December 2–8, Montréal, Canada
- Roy D, Sarkar A, Kundu P, Paul S, Sarkar BC (2023) An ensemble of evidence belief function (EBF) with frequency ratio (FR) using geospatial data for landslide prediction in Darjeeling Himalayan region of India. *Quaternary Science Advances* 11:100092. <https://doi.org/10.1016/j.qsa.2023.100092>
- Saber M, Boulmaiz T, Guermoui M, Abdrabo KI, Kantoush SA, Sumi T et al (2022) Examining LightGBM and CatBoost models for wadi flash flood susceptibility prediction. *Geocarto Int* 37(25):7462–7487
- Sahin EK (2020) Assessing the predictive capability of ensemble tree methods for landslide susceptibility mapping using XGBoost, gradient boosting machine, and random forest. *SN Appl Sci* 2:1308. <https://doi.org/10.1007/s42452-020-3060-1>
- Sahin EK (2022) Comparative analysis of gradient boosting algorithms for landslide susceptibility mapping. *Geocarto Int* 37(9):2441–2465
- Sifa SF, Mahmud T, Abdullah Tarin M, Enamul Haque DM (2020) Event-based landslide susceptibility mapping using weights of evidence (WoE) and modified frequency ratio (MFR) model: a case study of Rangamati district in Bangladesh. *Geology, Ecology, and Landscapes* 4(3):222–235
- Song Y, Yang D, Wu W, Zhang X, Zhou J, Tian Z, Wang C, Song Y (2023) Evaluating Landslide Susceptibility Using Sampling Methodology and Multiple Machine Learning Models. *ISPRS Int J Geo-Inf* 12:197. <https://doi.org/10.3390/ijgi12050197>
- Sun D, Ding Y, Zhang J, Wen H, Wang Y, Xu J, Zhou X (2022) Liu R (2022) Essential insights into decision mechanism of landslide susceptibility mapping based on different machine learning models. *Geocarto Int* 10(1080/10106049):2146763
- Sun D, Chen D, Zhang J, Mi C, Gu Q, Wen H (2023a) Landslide Susceptibility Mapping Based on Interpretable Machine Learning from the Perspective of Geomorphological Differentiation. *Land* 12:1018. <https://doi.org/10.3390/land12051018>
- Sun D, Wu X, Wen H, Gu Q (2023b) A LightGBM-based landslide susceptibility model considering the uncertainty of non-landslide samples. *Geomat Nat Haz Risk* 14(1):2213807. <https://doi.org/10.1080/19475705.2023.2213807>
- Taalab K, Cheng T, Zhang Y (2018) Mapping landslide susceptibility and types using Random Forest. *Big Earth Data* 2:159–178
- Teke A, Kavzoglu T (2023) Explainable artificial intelligence empowered landslide susceptibility mapping using Extreme Gradient Boosting (XGBoost). *Advanced Engineering Days* 6:74–76
- Tien Bui D, Pradhan B, Lofman O, Revhaug I (2012) Landslide Susceptibility Assessment in Vietnam Using Support Vector Machines, Decision Tree, and Naive Bayes Models. *Math Probl Eng* 2012:974638. <https://doi.org/10.1155/2012/974638>
- Trinh T, Luu BT, Le Thi TH, Nguyen DH, Tran TV, Nguyen THV, Nguyen KQ, Nguyen LT (2023) A comparative analysis of weight-based machine learning methods for landslide susceptibility mapping in Ha Giang area. *Big Earth Data* 7(4):1005–1034. <https://doi.org/10.1080/20964471.2022.2043520>
- TURKSTAT (2023) Address based population registration system results 2021. <https://biruni.tuik.gov.tr/medas/?kn=95&locale=tr>. Accessed 1 Feb 2023
- Varnes DJ (1978) Slope movement types and processes. In: Schuster RL, Krizek RJ (eds) Special Report 176: Landslides: Analysis and Control. TRB, National Research Council, Washington, DC, pp 11–33
- Vega J, Sepúlveda-Murillo FH, Parra M (2023) Landslide Modeling in a Tropical Mountain Basin Using Machine Learning Algorithms and Shapley Additive Explanations. *Air, Soil and Water Research* 16:1–20. <https://doi.org/10.1177/11786221231195824>
- Wang G, Chen X, Chen W (2020a) Spatial Prediction of Landslide Susceptibility Based on GIS and Discriminant Functions. *ISPRS Int J Geo-Inf* 9:144. <https://doi.org/10.3390/ijgi9030144>
- Wang Z, Hong T, Piette MA (2020b) Building thermal load prediction through shallow machine learning and deep learning. *Appl Energy* 263:114683. <https://doi.org/10.1016/j.apenergy.2020.114683>
- Wang X, Zhang X, Bi J, Zhang X, Deng S, Liu Z, Wang L, Guo H (2022) Landslide Susceptibility Evaluation Based on Potential Disaster Identification and Ensemble Learning. *Int J Environ Res Public Health* 19:14241. <https://doi.org/10.3390/ijerph191214241>
- Wei A, Yu K, Dai F, Gu F, Zhang W, Liu Y (2022) Application of Tree-Based Ensemble Models to Landslide Susceptibility Mapping: A Comparative Study. *Sustainability* 14:6330. <https://doi.org/10.3390/su14106330>
- 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. <https://doi.org/10.1016/j.catena.2019.104396>
- Wubalem A (2021) Landslide susceptibility mapping using statistical methods in Uatzau catchment area, northwestern Ethiopia. *Geoenviron Disasters* 8:1. <https://doi.org/10.1186/s40677-020-00170-y>
- Yavuz Ozalp A, Akinci H, Zeybek M (2023) Comparative Analysis of Tree-Based Ensemble Learning Algorithms for Landslide Susceptibility Mapping: A Case Study in Rize. *Turkey Water* 15:2661. <https://doi.org/10.3390/w15142661>
- 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 Nat Hazards* 113:965–995
- Yi Y, Zhang Z, Zhang W, Jia H, Zhang J (2020) Landslide susceptibility mapping using multiscale sampling strategy and convolutional neural network: A case study in Jiuzhaigou region. *CATENA* 195:104851. <https://doi.org/10.1016/j.catena.2020.104851>
- Yilmaz I (2009) Landslide susceptibility mapping using frequency ratio, logistic regression, artificial neural networks and their comparison: A case study from Kat landslides (Tokat—Turkey). *Comput Geosci* 35(6):1125–1138
- Youssef AM, Pourghasemi HR (2021) Landslide susceptibility mapping using machine learning algorithms and comparison of their performance at Abha Basin, Asir Region. *Saudi Arabia Geosci Front* 12(2):639–655
- Youssef AM, Pourghasemi HR, Pourtaghi ZS, Al-Katheeri MM (2016) Landslide susceptibility mapping using random forest, boosted regression tree, classification and regression tree, and general linear models and comparison of their performance at Wadi Tayyah Basin, Asir Region, Saudi Arabia. *Landslides* 13:839–856
- Youssef K, Shao K, Moon S, Bouchard L-S (2023) Landslide susceptibility modeling by interpretable neural network. *Communications Earth & Environment* 4:162. <https://doi.org/10.1038/s43247-023-00806-5>
- Yu H, Pei W, Zhang J, Chen G (2023) Landslide Susceptibility Mapping and Driving Mechanisms in a Vulnerable Region Based on Multiple Machine Learning Models. *Remote Sensing* 15(7):1886. <https://doi.org/10.3390/rs15071886>
- Zhang RH, Wu CZ, Goh ATC, Böhlke T, Zhang WG (2020a) Estimation of diaphragm wall deflections for deep braced excavation in anisotropic clays using ensemble learning. *Geosci Front* 12(1):365–373

- Zhang Y, Zhao Z, Zheng J (2020b) CatBoost: A new approach for estimating daily reference crop evapotranspiration in arid and semi-arid regions of Northern China. *J Hydrol* 588:125087. <https://doi.org/10.1016/j.jhydrol.2020.125087>
- Zhang T, Fu Q, Wang H, Liu F, Wang H, Han L (2022a) Bagging-based machine learning algorithms for landslide susceptibility modeling. *Nat Hazards* 110:823–846
- Zhang W, Wu C, Tang L, Gu X, Wang L (2022b) Efficient time-variant reliability analysis of Bazimen landslide in the Three Gorges Reservoir Area using XGBoost and LightGBM algorithms. *Gondwana Res* 123:41–53
- 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. *J Environ Manage* 332:117357. <https://doi.org/10.1016/j.jenvman.2023.117357>
- Zhao Y, Wang R, Jiang Y, Liu H, Wei Z (2019) GIS-based logistic regression for rainfall-induced landslide susceptibility mapping under different grid sizes in Yueqing, Southeastern China. *Eng Geol* 259:105147. <https://doi.org/10.1016/j.enggeo.2019.105147>
- Zhou Y, Wang W, Wang K, Song J (2022) Application of LightGBM Algorithm in the Initial Design of a Library in the Cold Area of China Based on Comprehensive Performance. *Buildings* 12:1309. <https://doi.org/10.3390/buildings12091309>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.