Natural Hazards (2022) 110:823–846
https://doi.org/10.1007/s11069-021-04986-1

REVIEW ARTICLE

Bagging-based machine learning algorithms for landslide susceptibility modeling

Tingyu Zhang[1,2] · Quan Fu[3] · Hao Wang[4] · Fangfang Liu[3] · Huanyuan Wang[1,2] · Ling Han[5]

Received: 22 May 2021 / Accepted: 10 August 2021 / Published online: 21 August 2021

Abstract
Landslide hazards have attracted increasing public attention over the past decades due to a series of catastrophic consequences of landslide occurrence. Thus, the mitigation and prevention of landslide hazards have been the topical issues. Thereinto, numerous research achievements on landslide susceptibility assessment have been springing up in recent years. In this paper, four benchmark models including best-first decision tree (BFTree), functional tree, support vector machine and classification regression tree (CART) and were integrated with bagging strategy. Then, these bagging-based models were applied to map regional landslide susceptibility in Jiange County, Sichuan Province, China. Fifteen conditioning factors were employed in establishing landslide susceptibility models, respectively, slope aspect, slope angle, elevation, plan curvature, profile curvature, TWI, SPI, STI, lithology, soil, land use, NDVI, distance to rivers, distance to roads and distance to lineaments. Then utilize correlation attribute evaluation method to weigh the contribution of each factor. Finally, the comprehensive performance of various bagging-based models and corresponding benchmark models was evaluated and systematically compared applying receiver operating characteristic curve and area under curve (AUC) values. Results demonstrated that bagging-based ensemble models significantly outperformed their corresponding benchmark models with validation dataset. Among them the Bag-CART model has the highest AUC value of 0.874; however, the AUC value of CART model is only 0.766, which reflected satisfying predictive capacity of integrated models in some degree. The achievements obtained in this study have some reference values for landslides prevention and land resource planning in Jiange County.

Keywords Landslide susceptibility · Bagging · Best-first decision tree · Functional tree · Classification and regression tree · Support vector machine

✉ Huanyuan Wang
whysxdj2021@163.com

Extended author information available on the last page of the article

1 Introduction

As the process of urbanization has been boosted in the past decades around mountainous areas, the negative effects of human activities on geological environment have become more significant as well. Thereinto, landslide hazards are generally considered as the most representative disasters which could induce enormous losses on residents' lives and property (Palmisano et al. 2016; Huang et al. 2019; Chen and Chen 2021). To some degree, landslide occurrence hampers the efficient utilization of land resources and economic development in mountainous regions. To prevent landslide hazards, a group of scholars and experts have spared no efforts to reveal landslide mechanism and invent a series of mitigation measures (Mohammadi and Taiebat2016; Peng et al. 2018; Huang et al. 2019; Ma et al. 2019; Wang et al. 2019a; Yang et al. 2019). Meanwhile, neoteric techniques about landslide susceptibility assessment have also attracted more extensive attention in the past years (Lee et al. 2018a; Lei et al. 2020b; Shou and Lin 2020; Li et al. 2021). Traditionally, the qualitative and semi-quantitative approaches are commonly applied in landslide susceptibility mapping (Bourenane et al. 2015; Pourghasemi and Rossi 2017; Pradhan et al. 2017). However, it should be noted that the prediction accuracy mainly depends upon subjective acknowledgment and expertise, which usually vary with each individual. To overcome the defects mentioned above, various statistical approaches, for instance frequency ratio (FR) (Aditian et al. 2018), weight of evidence (WoE) (Armaş 2012; Ding et al. 2017), index of entropy (IoE) (Jaafari et al. 2014) and certainty factor (CF) (Binaghi et al. 1998; Wu et al. 2016) have been introduced into landslide susceptibility evaluation. In thus, regional landslide susceptibility can be identified by judging the weight of conditioning factors in the landslide's occurrence. Nevertheless, there is no denying that some pre-assumptions of statistical approaches restrict their submission (Reichenbach et al. 2018). Additionally, deterministic models are also often seen in tasks of landslide susceptibility assessment (Akgun and Erkan 2016). Generally, deterministic models are more suitable to evaluate stability of single slope as a result of relatively complex modeling process and higher computational cost (Park et al. 2019). In recent years, under the development of data mining techniques and machine learning, large quantity of novel models have been employed to classify landslide susceptibility zones. According to literature, the most prevailing machine learning models include support vector machine (SVM) (Chang et al. 2019; Nguyen et al. 2019a), artificial neural network (ANN) (Dou et al. 2019b), multi-layer perceptron (MLP) (Hong et al. 2019; Pham et al. 2019), decision tree (DT) (Kutlug Sahin and Colkesen 2019), etc. These machine learning models can efficiently reveal the intricate regulations which are hidden in huge amounts of data, which has benefits to produce more reliable results. Currently, these machine learning models have been accepted by many scholars in different domains due to their prominent predictive performance (Amiri et al. 2019; Choubin et al. 2019; Hosseinalizadeh et al. 2019; Mohammady et al. 2019). In some cases, the performance of machine learning models dramatically varies with databases, indicating that there still exists latent capacity to promote in generalization performance (Pourghasemi and Rahmati 2018). Hence, a lot of optimization algorithms and ensemble strategies were integrated with conventional machine learning models to improve their comprehensive performance (Hong et al. 2018; Lee et al. 2018b; Dou et al. 2019a; Moayedi et al. 2019). Especially for ensemble learning models, they can always generate more satisfying results with lower variance and bias values. Compared to other ensemble techniques, the bagging ensemble exhibits superiority in well-understood theory, brief framework and promising results (Pham et al. 2017c; Pham and Prakash 2019). Although numerous bagging-based hybrid models such as J48-bagging (Hong et al. 2019), BKLR (Chen et al. 2018a) and bagging-FT models (Tien Bui et al. 2016) have been proposed, the researches on systematic comparison among different bagging-based machine learning algorithms are still rare now. Therefore, in this paper, four popular base classifiers, including best-first decision tree (BFTree), functional tree (FT), classification and regression tree (CART), support vector machine (SVM), were combined with bagging. Thereafter, the performance of bagging-based classifiers was evaluated and compared to that of corresponding base classifiers applying receiver operating characteristic curve (ROC) and area under curve (AUC) values. Finally, landslide susceptibility maps based on four hybrid classifiers were generated by ArcGIS tools.

2 Study area and data used

In this present case, Jiange County, covering the whole area of 3204 km^2, was selected as study area. It is located at longitude of 105°10′E-105°49′E and latitude of 31°21′N-31°31′N (Fig. 1). This subtropical humid monsoon climate dominates in the study area, which contributes to abundant rainfall and distinct seasons. It should be emphasized that local climate generally changes with elevation due to fluctuating mountainous landform within study area.

Topographically, through interpreting the 20 m regular raster digital elevation model, the altitude of the study area is between 358 and 1284 m. As a whole, altitude value shows an upward trend from southeast to northwest, which demonstrates typical characteristics of low mountainous and hilly regions. In the case of slope angle, the minimum and maximum values are 0° and 78.59° separately, and the average slope angle is 15.59° with the standard deviation (SD) of 0.03°. Concretely, areas with slope angle < 30° account for 91.41%, whereas areas with slopes between 30° and 60° cover approximately 8.55%. Correspondingly, the other areas with slope angle > 60° account for 0.04% about of the total area.

In the research, the database was prepared through historical data collection, satellite images interpretation and field survey. Ultimately, a landslide inventory map with 262

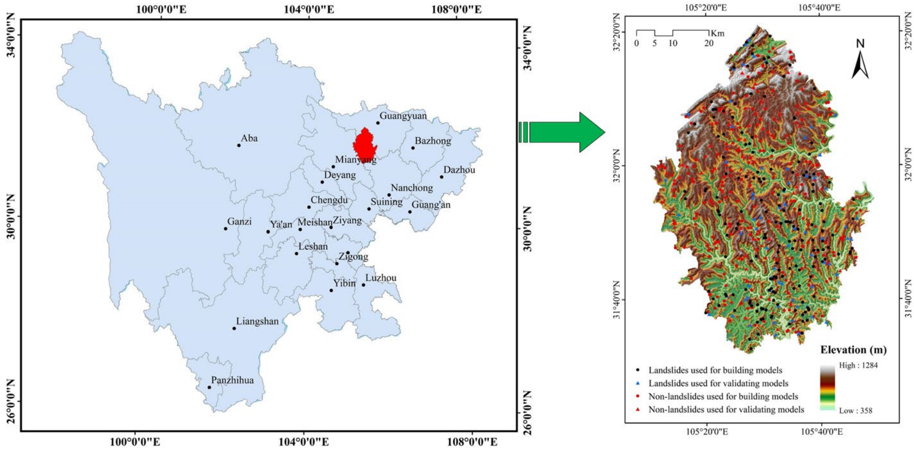


Fig.1 Study area

landslides, which can directly show space characteristic information of landslides in this study area, was plotted. Subsequently, the present study generalized each actual landslide as a single point. Besides, equivalent non-landslide points were randomly selected in study area. Afterward, the total landslide samples were divided into training data set (70%) and validation data set (30%) (Arabameri et al. 2019b; Lei et al. 2020b).

Currently, there is no recognized scheme about selection of landslide conditioning factors, besides the employed factors are always distinct in existing research achievements. In this study, fifteen landslide conditioning factors (for instance: slope aspect, elevation, slope angle, profile curvature, plan curvature, TWI, SPI, STI, lithology, soil, land use, NDVI, distance to rivers, distance to lineaments and distance to roads) were determined based on availability of data sources and relevant literature (Mandal and Mandal 2018b; Lombardo and Mai 2018; Kadavi et al. 2019). Here, NDVI in study area was derived from LANDSAT-8 satellite images (http://www.gscloud.cn/). Lithology and distance to lineaments were extracted from the geological map. Soil can be extracted from the soil maps at 1:1,000,000 scale (http://www.issas.ac.cn/). Land use map was extracted from regional land use maps with a 1:100,000 scale. The other conditioning factors can be acquired using ArcGIS tools, satellite images and DEM.

Slope aspect factor is especially common in landslide sensitivity mapping (Singh and Kumar 2017). In this study district, slope aspect has firm connections with solar radiation, vegetation coverage, rainfall and so on, which could affect slope stability to some extent (Kose and Turk 2019). In the present study, slope aspects within Jiange County were reclassified as: flat, east, north, northeast, south, southeast, southwest, northwest, west.

The law is clear that stress distribution in slope varies with slope angle, and slope angle is closely connected to slope failure pattern (Wang et al. 2019b). In this essay, eight categories were rearranged with an interval of 10°, correspondingly, 0–10°, 10–20°, 20–30°, 30–40°, 40–50°, 50–60°, 60–70° and 70–78.59°.

In some cases, landslide is prone to occur in areas with a certain elevation range (Li and Wang, 2019). In addition, it has been proved that environmental conditions around slopes could be determined by elevation values (Paranunzio et al. 2019; Regmi et al. 2019). The elevation values of this study region were separated by 100 m, and arranged into nine levels, including 358–500 m, 500–600 m, 600–700 m, 700–800 m, 800–900 m, 900–1000 m, 1000–1100 m, 1100–1200 m and 1200–1284 m.

Profile curvature and plan curvature are two critical indexes to reflect slope shape from different perspectives (He et al. 2019). On the one hand, slope shape affects stress distribution around slope surface, which is correlated with slope stability. On the other hand, surface runoff characteristics mainly depend on slope shape, and surface runoff characteristics have great impacts on erosion and infiltration around slopes (Mandal and Mandal, 2018a). In this case, plan curvature and profile curvature were also taken into account when mapping landslide susceptibility. Then re-divide plan curvature value into five groups such as (− 36.05)-(− 2.40), (− 2.40)-(− 0.94), (− 0.94)-0.23, 0.23–1.69 and 1.69–38.55. Similarly, for profile curvature, five groups were also obtained by reclassification, namely, (− 2.11)-(− 3.38), (− 3.38)-(− 1.20), (− 1.20)-0.68, 0.68–2.86 and 2.86–37.53.

Topographic wetness index (TWI) was impressed on measuring local groundwater potential initially (Moore et al. 1991). The TWI value can be figured out by the following equation (Chen et al. 2021a):

$$TWI = \ln\left(\frac{\alpha}{\tan\beta}\right) \tag{1}$$

where α indicates the flow accumulation via a point, and β is the corresponding tilt angle. Finally, TWI values with research area were identified as < 1, 1–2, 2–3, 3–4 and > 4.

As a quantitative indicator of erosion capacity of surface runoff, stream power index (SPI) was also included in consideration. The calculation formula of SPI can be expressed as follows (Lei et al. 2021):

$$SPI = \alpha \tan \beta \tag{2}$$

In this formula, α and β are catchment proportion and tilt angle, respectively (Wu et al. 2017; Chen et al. 2021b). Here, all the SPI values were grouped as five classes: 0–10, 10–20, 20–30, 30–40 and > 40.

Sediment transport index (STI) is another topographic index reflecting local erosion power (Pourghasemi et al. 2013a). Thus, the STI was introduced in landslide susceptibility modeling, and its value can be determined by the following formula:

$$SPI = \left(\frac{\alpha}{22.13}\right)^{0.6} \left(\frac{\sin \beta}{0.0896}\right)^{1.3} \tag{3}$$

The parametric significance of α and β is same as that of SPI. The STI value consists of five intervals, involving 0–10, 10–20, 20–30, 30–40 and > 40.

There is no doubt that lithology and soil types decide their physical and mechanical properties, which are generally treated as the most essential factor in slope stability assessment (Abuzied and Alrefaee, 2019; Watakabe and Matsushi, 2019). In the case of lithology, eleven classes were detected ultimately while nine kinds of soil types were found out. Concretely, all the soil and lithology types were illustrated in Table 1. Figure 2i and j demonstrate the spatial distribution of multifarious soils and lithologies. Normally, strata with lower shearing strength and deformation modulus are more likely to form landslides. Among the strata in this study, the soils and lithologies owning relatively high landslide susceptibility mainly compare flood-plain alluvium, interbeds of lacustrine-shallow and lacustrine deposits, yellow soil, paddy soil, etc. It should be explained that water was taken as one special soil type in this case.

Based on literature, frequency of landslide occurrence usually fluctuates with different land use types (Hong et al. 2017). In general, the probability of landslide spreading near construction land, farm land and unused land is greater than other land use types. On the contrary, those areas which are covered by abundant vegetations could be deemed as lower landslide susceptibility zones (Wang et al. 2019c). Within Jiange County, a total of six land use types were produced, respectively, forest land, farm land, grass land, construction land, unused land and water.

Normalized difference vegetation index (NDVI) is known as the quantitative measurement of vegetation coverage degree (Choi et al. 2012). This index is described as follows:

$$NDVI = (IR - R)/(IR + R) \tag{4}$$

where R represents the red band of electromagnetic spectrum and IR means the infrared band and. The limit of NDVI values is [-1, 1], as well as higher NDVI values manifest better degree of vegetation development. In this paper, NDVI values within study area consist of five intervals: (-0.11)-0.12, 0.12–0.22, 0.22–0.27, 0.27–0.33 and 0.33–0.52.

In some regions, landslide hazards are inclined to spread along rivers, because rivers could erode slope toes and raise moisture content in slope bodies (Clapuyt et al. 2019). The tasks in regard to landslide susceptible mapping, the effects about rivers can be weighed by

Table 1 Landslide conditioning factors and their classification

No.	Factors	Classes
1	Slope aspect	(1) Flat; (2) North; (3) East; (4) South; (5) West; (6) Northeast; (7) Southeast; (8) Northwest; (9) Southwest
2	Slope angle (°)	(1) 0 – 10; (2) 10 – 20; (3)20 – 30; (4) 30 – 40; (5) 40 – 50; (6) 50 – 60; (7) 60 – 70; (8) 70 – 78.59
3	Elevation (m)	(1) 358 – 500; (2) 500 – 600; (3) 600 – 700; (4) 700 – 800; (5) 800 – 900; (6) 900 – 000; (7) 1000 – 1,100; (8) 1100 – 1200; (9) 1200 – 1284
4	Plan curvature	(1) (– 36.05) – (– 2.40); (2) (– 2.40) – (– 0.94); (3) (– 0.94) – 0.23; (4) 0.23 – 1.69; (5) 1.69 – 38.55
5	Profile curvature	(1) (– 2.11) – (– 3.38); (2) (– 3.38) – (– 1.20); (3) (– 1.20) – 0.68; (4) 0.68 – 2.86; (5) 2.86 – 37.53
6	TWI	(1) < 1; (2) 1 – 2; (3) 2- 3; (4) 3 – 4; (5) > 4
7	SPI	(1) 0 – 10; (2) 10 – 20; (3) 20 – 30; (4) 30 – 40; (5) > 40
8	STI	(1) 0 – 10; (2) 10 – 20; (3) 20 – 30; (4) 30 – 40; (5) > 40
9	Lithology	(1) Flood-plain alluvium (Quaternary); (2) Conglomerate, siltstone and mudstone (Cretaceous); (3) Piedmont and lacustrine deposits (Cretaceous); (4) Sandy conglomerate (Cretaceous); (5) Interbeds of lacustrine-shallow and lacustrine deposits, adlittoral and semi-deep lacustrine deposits (Jurassic); (6) Lacustrine deposits (Jurassic); (7) Lacustrine-swamp deposits with dark grey coal-bearing formation (Jurassic); (8) Interactive marine and terrestrial deposits (Triassic); (9) Neritic shale and marl deposits (Triassic); (10) Limestone, dolomitic limestones, shale (Permian); (11) Neritic sand-shale stone deposits (Cambrian)
10	Soil	(1) Calcaric Regosol; (2) Water; (3) Yellow soil; (4) Chromic Luvisol; (5) Yellow calcic soil; (6) Percogenic pady soil; (7) Calcareous purplish soil; (8) Pady soil; (9) Submergenic pady soil
11	Land use	(1) Unused land; (2) Construction land; (3) Grass land; (4) Forest land; (5) Farm land; (6) Water
12	NDVI	(1) (– 0.11) – 0.12; (2) 0.12 – 0.22; (3) 0.22- 0.27; (4) 0.27 – 0.33; (5) 0.33 – 0.52
13	Distance to rivers (m)	(1) 0 – 200; (2) 200 – 400; (3) 400 – 600; (4) 600 – 800; (5) > 800
14	Distance to roads (m)	(1) 0 – 200; (2) 200 – 400; (3) 400 – 600; (4) 600 – 800; (5) > 800
15	Distance to lineaments (m)	(1) 0 – 2000; (2) 2000 – 4000; (3) 4000 – 6000; (4) 6000 – 8000; (5) > 8000

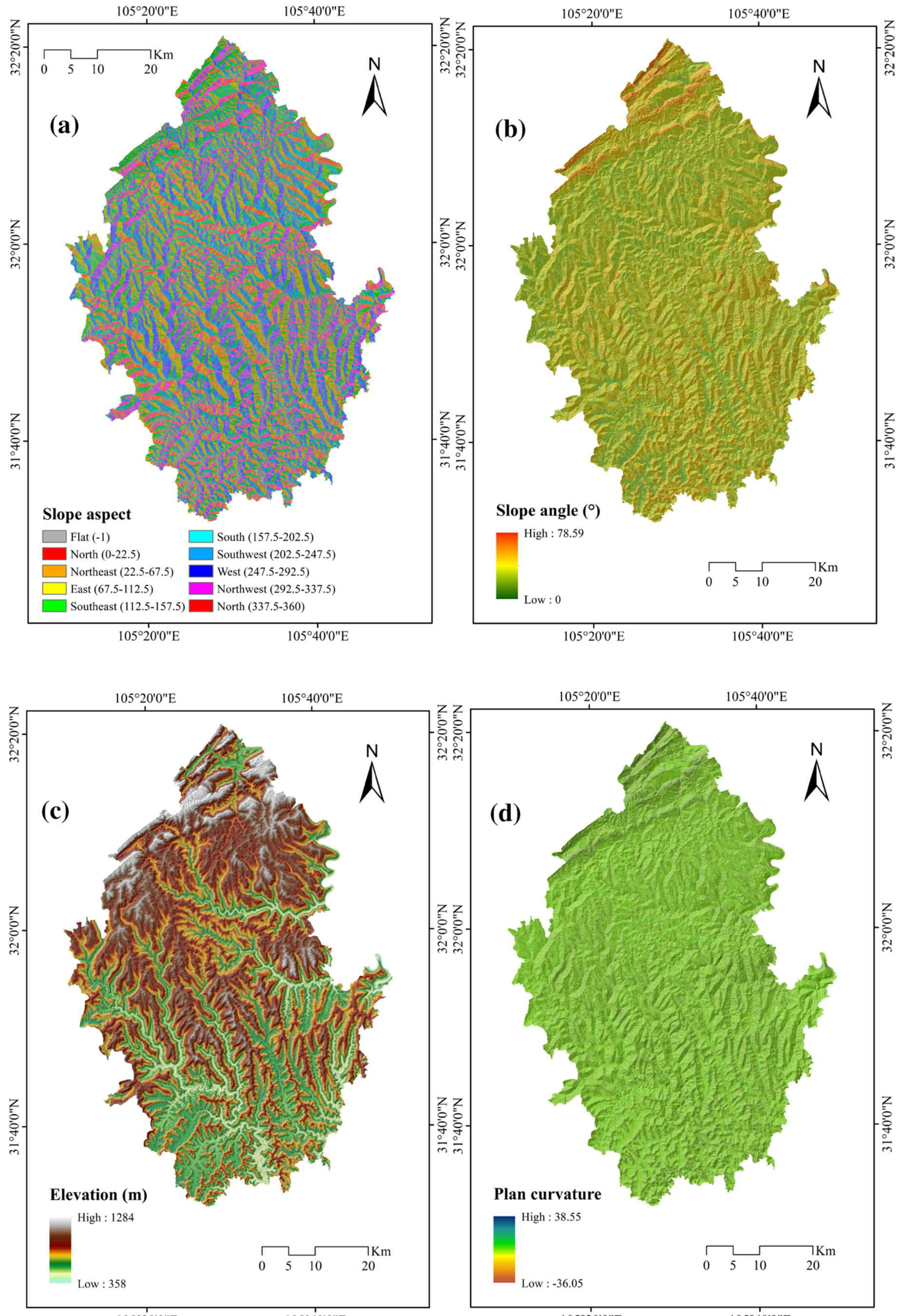


Fig. 2 Landslide conditioning factors

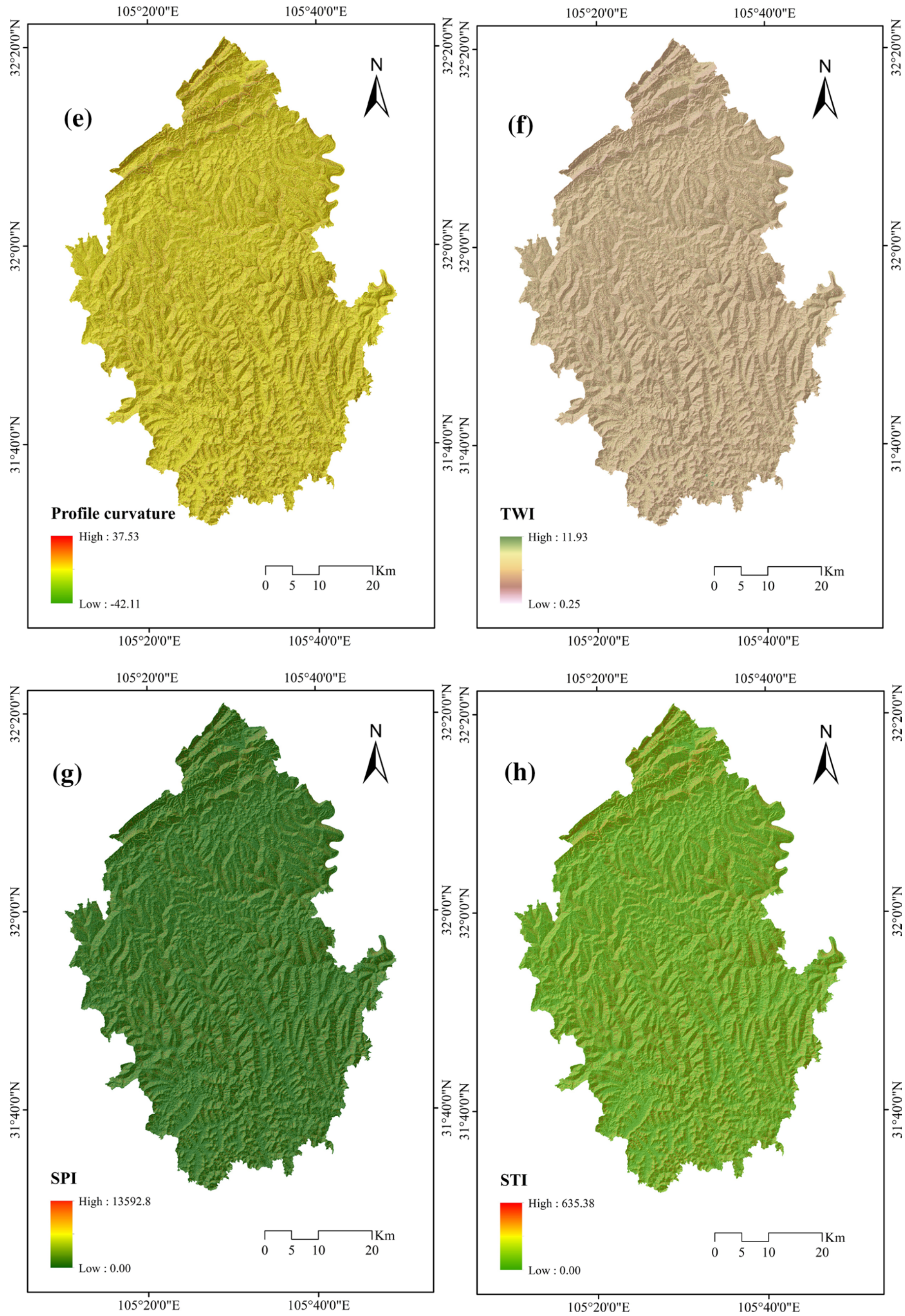


Fig. 2 (continued)

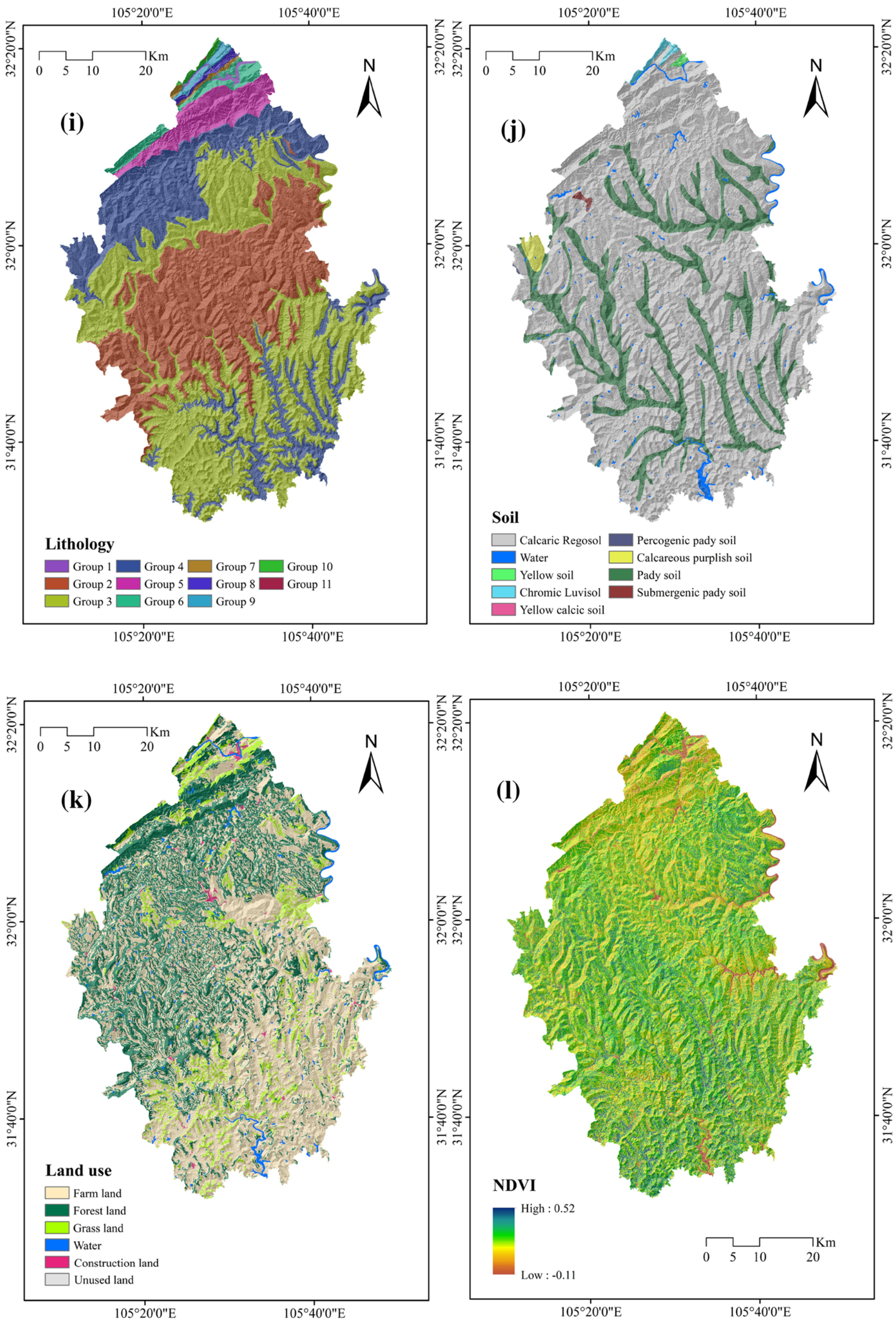


Fig. 2 (continued)

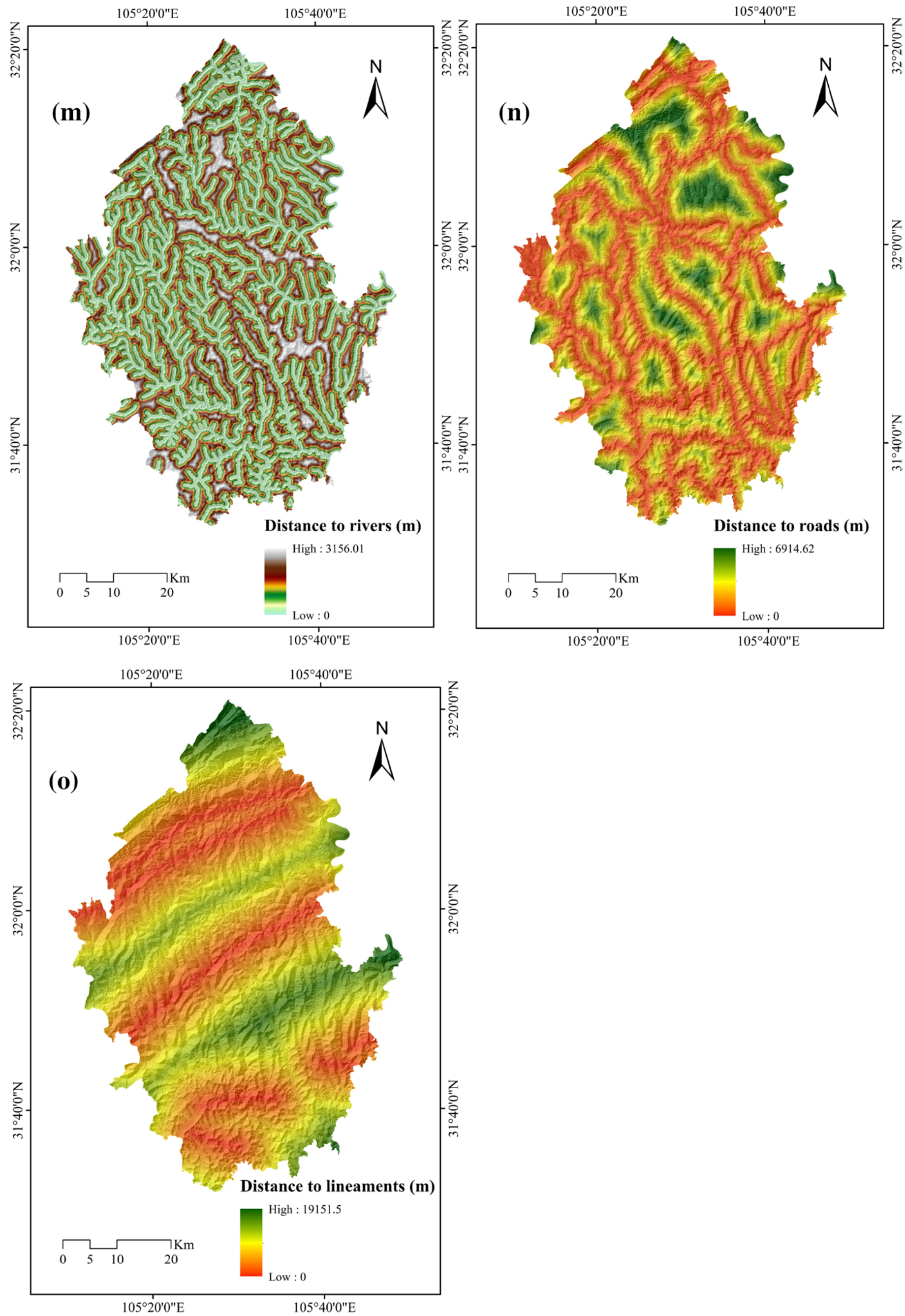


Fig. 2 (continued)

distance to rivers and river density (Arabameri et al. 2019a; Kose and Turk, 2019). During the present research, distance to rivers was selected to construct landslide susceptibility model. With an interval of 200 m, five sections were acquired, comprising 0–200 m, 200–400 m, 400–600 m, 600–800 m and > 800 m.

Road construction is another non-negligible conditioning factor in landslide susceptibility assessment. Actually, some road engineering in mountainous areas could make radical changes on slope shapes, and engineering vibration may increase probability of slope deformation or failure (Du et al. 2019). In Jiange County, the distance to roads was analyzed and reclassifies it into five classes with 200 m as an interval, viz., 0–200 m, 200–400 m, 400–600 m, 600–800 m and > 800 m.

The connection between landslide formation and geological structure has been revealed by numerous studies (Pourghasemi et al. 2013b, 2018; Juliev et al. 2019). Commonly, landslide susceptibility has negative association with distance to lineal geological structures such as faults and surface cracks. Based on ArcGIS software, the distance between each raster and lineaments was calculated, and the results were rearranged into five categories with 2000 m as an interval segment, including 0–2000, 2000–4000, 4000–6000, 6000–8000 and > 8000.

3 approach

3.1 Bagging

The full designation of bagging is bootstrap aggregation approach, belonging to the group of machine learning ensemble meta algorithms (Kadavi et al. 2018). Facts have proved that bagging retains an outstanding function on improving stability and generalization capacity of multiple base classifiers (Pham et al. 2017a; Truong et al. 2018). The core concept of bagging is bootstrap sampling technique which can be simply described as sampling with replacement. Specifically, new subsamples are generated by randomly sampling with replacement from initial training dataset. Then, all the subsamples were used to train base classifiers, and the final result could be found out through aggregating various base classifiers based on majority voting strategy (Breiman 1996). Compared to boosting technique, bagging improves predictive performance by decreasing variance corresponding to the landslide susceptibility model. In the research of this paper, four base classifiers such as best-first decision tree, functional tree, support vector machine, classification and regression tree were prepared to generate ensemble models using bagging strategy.

3.2 Best-first decision tree

Best-first decision tree (BFTree) is proven to be a remarkable member inter tree-based algorithms (Chen et al. 2018b). The most important feature of BFTree is that the best node should expand in depth-first order (Nguyen et al. 2019b; Lei et al. 2020a). The best node could be determined by measuring information gain and Gini index, which are used to assess node impurity (Lay et al. 2019; Shirzadi et al. 2019). For this algorithm, there exist two termination conditions, including that labels of all the instances are defined or the best value of splitting criteria is a negative value (Kumar et al. 2013). The calculation of information gain and Gini index can be implemented using the following formulas:

$$\text{Information Gain}\pi(D, A) = \text{Entropy}(D) - \sum \frac{|D_i|}{|D|}\text{Entropy}(D) \tag{5}$$

$$\text{Gini}(P) = \sum_{k=1}^{k} p_k\left(1 - p_k\right) \tag{6}$$

Among them, A represents any selected landslide conditioning factor, and D represents the training sample. K denotes the count of labels (non-landslide and landslide here), and pk means the probability of an instance belonging to the k-th class.

3.3 Functional trees

Functional tree (FT) algorithm is another tree classifier, which has excellent performance on reducing bias and variance (Gama 2001). In FT model, a logistic regression function is adopted to realize splitting and prediction on inner nodes and leaves namely (Tien Bui et al. 2019). Moreover, the comprehensive capability of the FT model is controlled by some parameters, like the cycle of bootstrap iterations, the lowest value of each leaf node instance and functional trees (Gama 2004). Before establishing functional trees, the linear bayes test function is applied in producing the probability distribution of non-landslide class and landslide class. Then, the initial conditioning factors could be extended to form new dataset, and the original dataset and split-new dataset are both involved in constructing classification trees (Pham et al. 2017b).

3.4 Classification and regression tree

Classification and regression tree (CART) as a nonparametric modus, and this approach can be seen in a number of studies such as landslide susceptibility, groundwater potential assessment, land subsidence modeling, etc. (Schifman et al. 2018; Naghibi et al. 2018; Sarkar et al. 2019; Rahmati et al. 2019; Choubin et al. 2019). The most notable benefit of CART is that this approach can deal with problems of outliers and missing values (Loh 2011). Additionally, the CART approach is more applicable for processing data with various types likes numeric, binary and categorical (Aertsen et al. 2010). In this process of CART construction, Gini values are computed and employed as the criteria of node splitting. Generally, the CART which is constructed with training data needs pruning with some validation data, furthermore, in order to simplify model framework and optimize model comprehensive performance (McKenney and Pedlar 2003).

3.5 Support vector machine

Support vector machine (SVM) is a particularly prevailing classifiers used in landslide susceptibility assessment in recent years (Xiao et al. 2018; Chang et al. 2019; Mokhtari and Abedian 2019). In contrast with other machine learning algorithm, the SVM model is more suitable for datasets consisting of a small number of samples (Huang and Zhao2018). Conventional SVM is a typical binary classification model which can be implemented with linearly separable data and linearly inseparable data (Maxwell et al. 2018). The basic principle of SVM is to explore out the optimal hyperplane which can divide training specimen into two categories with the highest accuracy. Actually, most data used in landslide susceptibility modeling are linearly inseparable. Therefore, nonlinear transformation should be carried out to map initial data to higher dimensional space in which samples are linearly separable. The operation of nonlinear transformation can be achieved using different kernel functions, mainly including linear kernel function, polynomial kernel function and Gaussian kernel function.

4 Results

4.1 Models results and analysis

In landslide susceptibility modeling, there may exist some conditioning factors which have no contribution to landslide occurrence. Hence, removing those minor affecting factors is a basic step to ameliorate the framework and rationality of the landslide susceptibility model. This article engrains the correlation attribute evaluatrion (CAE) means to calculate average merit (AM) value of each affecting factor (He et al. 2019). Concretely, a positive AM value means that the corresponding conditioning factor indeed contributes to landslide susceptibility model, and the conditioning factor with higher AM value is of greater importance. Table 2 illustrates the AM values toward fifteen influencing factors in the present examination. It can be concluded from the table that the contribution of slope angle is the topmost (AM = 0.311), and its standard deviation is ± 0.013. And second highest AM value of 0.151 belongs to land use, the other ones are elevation (AM = 0.120), TWI (AM = 0.118), distance to roads (AM = 0.104), STI (AM = 0.083), SPI (AM = 0.069), distance to river (AM = 0.060), NDVI (AM = 0.042), profile curvature (AM = 0.024), lithology (AM = 0.018), soil (AM = 0.014), plan curvature (AM = 0.013), slope aspect (AM = 0.012) and distance to lineaments (AM = 0.012). It is obvious that every affecting factor has correlations with landslide happening. Consequently, fifteen conditioning factors were included in building landslide susceptibility model ultimately.

Table 2 Importance of affecting factors based on CAE method

Factors	AM	Sd
Slope angle	0.311	±0.013
Land use	0.151	±0.013
Elevation	0.120	±0.012
TWI	0.118	±0.017
Distance to roads	0.104	±0.012
STI	0.083	±0.017
SPI	0.069	±0.020
Distance to rivers	0.060	±0.010
NDVI	0.042	±0.014
Profile curvature	0.024	±0.013
Lithology	0.018	±0.012
Soil	0.014	±0.008
Plan curvature	0.013	±0.005
Slope aspect	0.012	±0.006
Distance to lineaments	0.012	±0.009

4.2 Models validation

In this case, four benchmark models were combined with bagging strategy to generate various ensemble models, viz., Bag-BFTree model, Bag-FT model, Bag-CART model and Bag-SVM model. In this section, AUC value corresponding to ROC curve and validation samples were used to measure the accuracy and predictive capacity of benchmark models and integrated models, respectively (Chen et al. 2021c). The results are shown in Figs. 3 and 4. The AUC values of Bag-BFTree, Bag-FT, Bag-CART and Bag-SVM models are 0.869, 0.763, 0.874 and 0.729 namely. For those benchmark models, AUC value of CART model is 0.766, exhibiting a relatively better predictive capacity, the following are BFTree model (0.748), FT model (0.694) and SVM model (0.650). The AUC values obtained by various models are larger than 0.5, indicating that all the models have predictive abilities in landslide susceptibility within study area. However, it should be clarified that there are significant differences in the overall performance of the four selected models, even the results of some models (such as SVM and FT models) may not be satisfactory.

4.3 Generation of landslide susceptibility maps

According to the results obtained from model validation, four ensemble models based on bagging were selected to establish landslide susceptibility mapping in Jiange County. Before generating landslide susceptibility maps, fifteen conditioning factors of each raster within study area were extracted using ArcGIS tools. Afterward, the probability of landslide occurrence in each raster can be predicted using four trained bagging-based models.

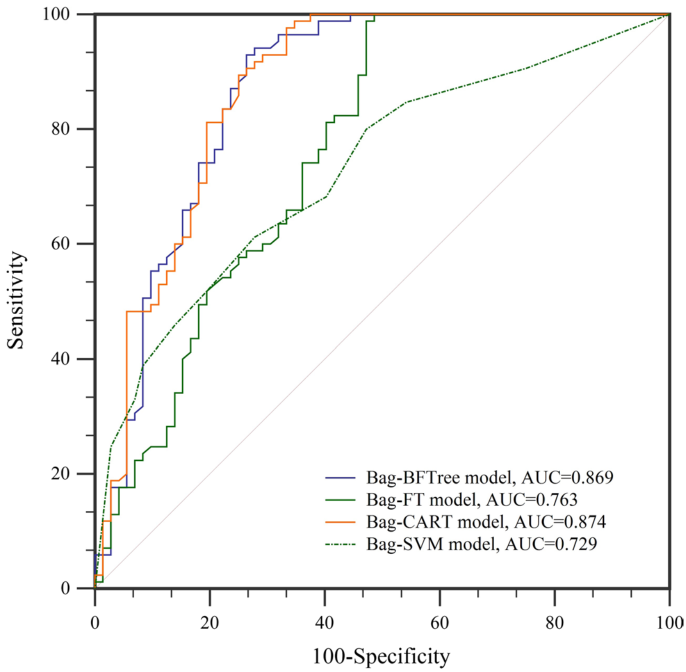


Fig. 3 ROC curves for the models

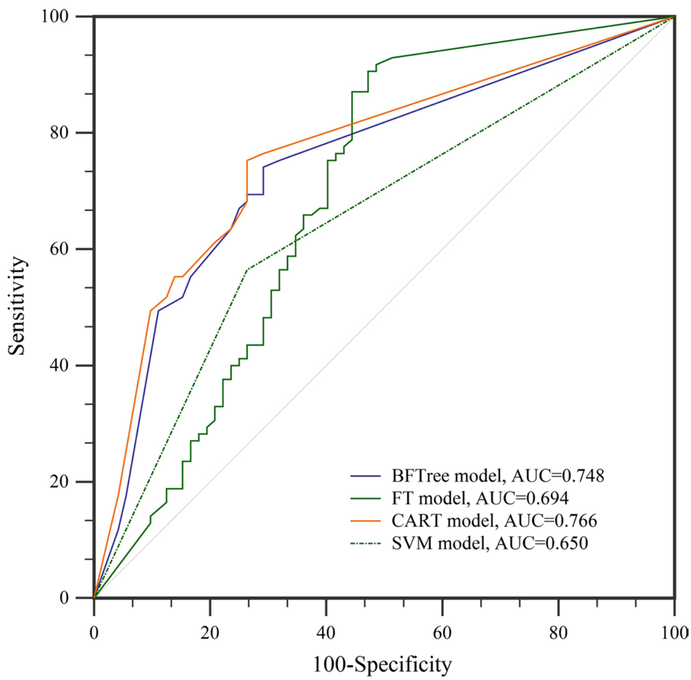


Fig. 4 ROC curves for the benchmark models

Furthermore, to identify different classes of landslide susceptibility, natural break method (Aditian et al. 2018; Kumar et al. 2018) was adopted here, after that the susceptibility of regional landslide was segmented into very high, high, moderate, low, and very low. Finally, all the results received are visualized in ArcGIS software, as shown in Fig. 5. Meanwhile, the area proportion of each class of landslide susceptibility level were computed and listed in Fig. 6.

5 Discussion

Landslides befall frequently in mountain region and could be considered as a horrible natural disaster (Haque et al. 2019). Currently, landslide susceptibility modeling process has introduced a large number of mathematical models in order to generate relatively more precise results (Pourghasemi et al. 2018). In view of excellent performance of bagging strategy, in the present study, four benchmark classifiers such as BFTree, FT, CART and SVM were combined with bagging. To systematically evaluate and contrast the quality of prediction capability between various bagging-based models and benchmark models is one of the staple intentions of this research. Additionally, four bagging-based models (Bag-BFTree, Bag-FT, Bag-CART and Bag-SVM) were utilized to the production of the landslide susceptibility mapping in Jiange County, which may have significance on local landslide mitigation and control.

In the course of this research, the influencing factors which were employed to construct landslide susceptibility models were selected based on local geological environment,

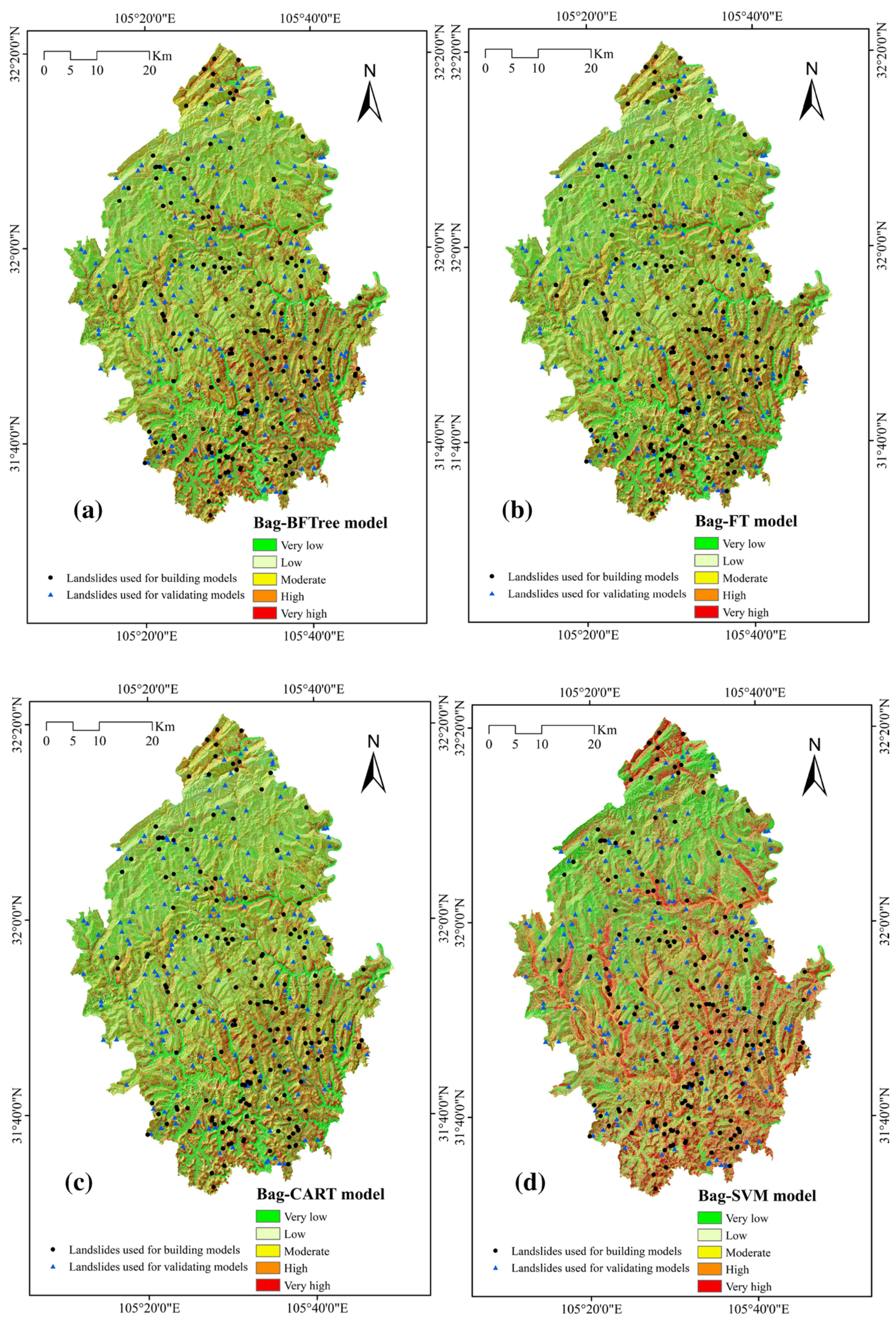
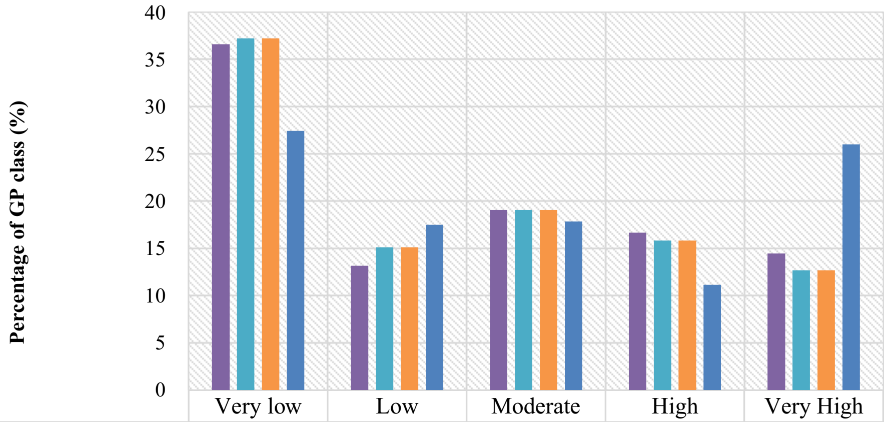


Fig. 5 Landslide susceptibility maps for hybrid models



	Very low	Low	Moderate	High	Very High
Bag-BFTree model	36.64	13.15	19.09	16.66	14.46
Bag-FT model	37.27	15.13	19.07	15.84	12.69
Bag-CART model	37.27	15.13	19.07	15.84	12.69
Bag-SVM model	27.46	17.50	17.87	11.15	26.02

Fig. 6 Percentages of landslide susceptibility classes

existing literature and accessibility of data resources. Furthermore, the CAE method was adopted to assess relative importance of each conditioning factor. As a result, through calculating acquired AM values of various influencing factors are all positive, demonstrating that all affecting factors contributed to landslide occurrence within study area. Concretely, the contribution of slope angle influence factor is identified as the superior because it has the topmost AM value, 0.311. Generally, areas within a certainty scope of slope angles are especially prone to place a premium on landslides (Kasai and Yamada 2019). In addition to slope angle, land use was judged as another critical evaluation criteria in landslide susceptibility assessment. According to the previous report, human-induced landslides events have been rising in the past few decades (Yue et al. 2018; Schmidt et al. 2019). Hence, construction land and farm land where human activities may be intense usually have higher susceptibility to landslide occurrence. There is another conditioning factor which should be explained furtherly. It could be seen that contribution of distance to lineament to landslide occurrence was relatively limited. The reason may be that majority of landslides in this study area were triggered by rainfall, river, engineering construction and so on rather than linear geological structure.

According to results of model validation, the optimal classifiers with validation dataset are Bag-BFTree model and Bag-CART model, respectively, among them the Bag-CART possesses the best predictive performance. It should be noticed that AUC values of Bag-SVM and SVM models are relatively lower, which means that applicability of SVM is inferior in this study and its performance could be boosted using optimization algorithms (Cheng and Hoang 2015; Wang et al. 2019d). Moreover, it is obvious that bagging-based models outperform those benchmark models, proving that bagging strategy is efficient on advancing model accuracy and predictive capacity. In the present study, for validation data, bagging strategy makes the BFTree model and CART model have higher AUC increments, which are 0.121 and 0.108 separately. It should be noted that effects of model performance improvement definitely have connections with benchmark model types. Consequently, bagging may be not the most suitable approach for some benchmark classifiers, and it is necessary to evaluate and compare comprehensive performance of hybrid models which are constructed by various ensemble frameworks and multiple base classifiers. In other words, the best landslide susceptibility model is not fixed and unique for different datasets, and the employed model should be determined by data characteristics and computing efficiency (Pourghasemi and Rahmati 2018).

In the process of generating landslide susceptibility maps, natural break method embedded in ArcGIS was applied to identify five landslide susceptibility classes. According to literature, the most common classification methods involve natural break method, standard deviation method, equal interval method, quantile method and so on (He et al. 2019). Thereinto, natural break is usually regarded as the most prevailing method, which can maximize differences among various categories by automatically searching breakpoints. In this case, the final results produced by natural break method are reasonable and functional for landslide prevention in study area.

6 Conclusions

According to the content of this article, the following key conclusions can be drawn:

(1) The total of fifteen landslide conditioning factors were put into use in creating landslide susceptibility model here. Based on results of CAE analysis, slope angle is of the most importance while distance to lineament had the lowest importance. Moreover, all the landslide conditioning factors indeed contributed to landslide occurrence in Jiange County.

(2) Bagging can significantly improve performance of benchmark models such as BFTree, FT, CART and SVM. Among those bagging-based models in this paper, Bag-CART model exhibited the best predictive performance with the AUC value of 0.874, the following is Bag-BFTree model, which two models possessed the more optimal accuracy with testing dataset. Additionally, the performance of FT and SVM models might be not satisfying due to lower AUC values.

(3) Based on natural break method, four classifiers which bagging-based models are used to create landslide sensitivity mappings holding five levels viz., very high, high, moderate, low and very low, respectively. Concretely, the low-sensitivity category has the highest percentage of area.

In a nutshell, the thoughts and framework of this study have reference significance on later relevant researches. Besides, the conclusions and achievements in this paper can be developed as effective tools to reduce probability and threat of landslide occurrence in Jiange County.

Authors' contributions Tingyu Zhang: Conceptualization, Methodology, Writing-Review & Editing, Funding Acquisition Quan Fu: Resources, Software, Validation Fangfang Liu: Formal Analysis, Data Curation Hao Wang: Visualization Huanyuan Wang: Writing-Original Draft Preparation Ling Han: Supervision.

Funding This study is financially supported by Fundamental Research Funds for the Central Universities (300102351502), Shaanxi Province Youth Talent Support Program Project (2021-1-2), Shaanxi Land Construction-Xi'an Jiaotong University Land Engineering and Human Settlement Environment Tecnology

Innovation Center Open Fund Project (2021WHZ0089) and Inner scientific research project of Shaanxi Land Engineering Construction Group (SXDJ2021-10, SXDJ2021-30, SXDJ2020-22). The author wish to express their sincere thanks to Chaohong Peng (Sichuang Institute of Geological Engineering Investigation Group Co.Ltd) for useful information provided.

Declarations

Conflicts of interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Abuzied SM, Alrefaee HA (2019) Spatial prediction of landslide-susceptible zones in El-Qaá area, Egypt, using an integrated approach based on GIS statistical analysis. Bull Eng Geol Env 78:2169–2195

Aditian 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

Aertsen W, Kint V, van Orshoven J, Özkan K, Muys B (2010) Comparison and ranking of different modelling techniques for prediction of site index in Mediterranean mountain forests. Ecol Model 221:1119–1130

Akgun A, Erkan O (2016) Landslide susceptibility mapping by geographical information system-based multivariate statistical and deterministic models: in an artificial reservoir area at Northern Turkey. Arab J Geosci 9:165

Amiri M, Pourghasemi HR, Ghanbarian GA, Afzali SF (2019) Assessment of the importance of gully erosion effective factors using Boruta algorithm and its spatial modeling and mapping using three machine learning algorithms. Geoderma 340:55–69

Arabameri A, Pradhan B, Rezaei K, Lee S, Sohrabi M (2019a) An ensemble model for landslide susceptibility mapping in a forested area. Geocarto Int 35:1–26

Arabameri A, Pradhan B, Rezaei K, Sohrabi M, Kalantari Z (2019b) GIS-based landslide susceptibility mapping using numerical risk factor bivariate model and its ensemble with linear multivariate regression and boosted regression tree algorithms. J Mount Sci 16:595–618

Armaş I (2012) Weights of evidence method for landslide susceptibility mapping. Prahova Subcarpathians. Romania Nat Haz 60:937–950

Binaghi E, Luzi L, Madella P, Pergalani F, Rampini A (1998) slope instability zonation: a comparison between certainty factor and fuzzy dempster-shafer approaches. Nat Hazards 17:77–97

Bourenane H, Bouhadad Y, Guettouche MS, Braham M (2015) GIS-based landslide susceptibility zonation using bivariate statistical and expert approaches in the city of Constantine (Northeast Algeria). Bull Eng Geol Env 74:337–355

Breiman L (1996) Bagging predictors. Mach Learn 24:123–140

Chang K-T, Merghadi A, Yunus AP, Pham BT, Dou J (2019) Evaluating scale effects of topographic variables in landslide susceptibility models using GIS-based machine learning techniques. Sci Rep 9:12296

Chen X, Chen W (2021) GIS-based landslide susceptibility assessment using optimized hybrid machine learning methods. CATENA 196:104833

Chen W, Shahabi H, Zhang S, Khosravi K, Shirzadi A, Chapi K, Pham TB, Zhang T, Zhang L, Chai H, Ma J, Chen Y, Wang X, Li R, Ahmad BB (2018a) Landslide susceptibility modeling based on GIS and novel bagging-based kernel logistic regression. Appl Sci 8:2540–2562

Chen W, Zhang S, Li R, Shahabi H (2018b) 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 W, Lei X, Chakrabortty R, Chandra Pal S, Sahana M, Janizadeh S (2021a) Evaluation of different boosting ensemble machine learning models and novel deep learning and boosting framework for head-cut gully erosion susceptibility. J Environ Manag 284:112015

Chen Y, Chen W, Chandra Pal S, Saha A, Chowdhuri I, Adeli B, Janizadeh S, Dineva AA, Wang X, Mosavi A (2021b) Evaluation efficiency of hybrid deep learning algorithms with neural network, decision tree and boosting methods for predicting groundwater potential. Geocarto Int 36:1–20

Chen Y, Che W, Janizadeh S, Bhunia GS, Bera A, Pham QB, Linh NTT, Balogun A-L, Wang X (2021c) Deep learning and boosting framework for piping erosion susceptibility modeling: spatial evaluation of agricultural areas in the semi-arid region. Geocarto Int 36:1–27

Cheng M-Y, Hoang N-D (2015) Typhoon-induced slope collapse assessment using a novel bee colony optimized support vector classifier. Nat Hazards 78:1961–1978

Choi J, Oh H-J, Lee H-J, Lee C, Lee S (2012) Combining landslide susceptibility maps obtained from frequency ratio, logistic regression, and artificial neural network models using ASTER images and GIS. Eng Geol 124:12–23

Choubin B, Moradi E, Golshan M, Adamowski J, Sajedi-Hosseini F, Mosavi A (2019) An ensemble prediction of flood susceptibility using multivariate discriminant analysis, classification and regression trees, and support vector machines. Sci Total Environ 651:2087–2096

Clapuyt F, Vanacker V, Christl M, Van Oost K, Schlunegger F (2019) Spatio-temporal dynamics of sediment transfer systems in landslide-prone Alpine catchments. Solid Earth 10:1489–1503

Ding Q, Chen W, Hong H (2017) Application of frequency ratio, weights of evidence and evidential belief function models in landslide susceptibility mapping. Geocarto Int 32:619–639

Dou J, Yunus AP, Xu Y, Zhu Z, Chen C-W, Sahana M, Khosravi K, Yang Y, Pham BT (2019a) Torrential rainfall-triggered shallow landslide characteristics and susceptibility assessment using ensemble data-driven models in the Dongjiang Reservoir Watershed, China. Nat Hazards 97:579–609

Dou J, Yunus PA, Tien Bui D, Sahana M, Chen C-W, Zhu Z, Wang W, Thai Pham B (2019b) Evaluating GIS-based multiple statistical models and data mining for earthquake and rainfall-induced landslide susceptibility using the LiDAR DEM. Remote Sens 11:638–659

Du G, Zhang Y, Yang Z, Guo C, Yao X, Sun D (2019) Landslide susceptibility mapping in the region of eastern Himalayan syntaxis, Tibetan Plateau, China: a comparison between analytical hierarchy process information value and logistic regression-information value methods. Bull Eng Geol Env 78:4201–4215

Gama J (2001) Functional trees for classification, Proceedings 2001 IEEE International Conference on Data Mining, pp 147–154

Gama J (2004) Functional trees. Mach Learn 55:219–250

Haque U, da Silva PF, Devoli G, Pilz J, Zhao B, Khaloua A, Wilopo W, Andersen P, Lu P, Lee J, Yamamoto T, Keellings D, Wu J-H, Glass GE (2019) The human cost of global warming: deadly landslides and their triggers (1995–2014). Sci Total Environ 682:673–684

He Q, Xu Z, Li S, Li R, Zhang S, Wang N, Pham TB, Chen W (2019) Novel entropy and rotation forest-based credal decision tree classifier for landslide susceptibility modeling. Entropy 21:106–130

Hong H, Pradhan B, Bui DT, Xu C, Youssef AM, Chen W (2017) Comparison of four kernel functions used in support vector machines for landslide susceptibility mapping: a case study at Suichuan area (China). Geomat Nat Haz Risk 8:544–569

Hong H, Liu J, Bui DT, Pradhan B, Acharya TD, Pham BT, Zhu AX, Chen W, Ahmad BB (2018) Landslide susceptibility mapping using J48 Decision Tree with AdaBoost, Bagging and Rotation Forest ensembles in the Guangchang area (China). CATENA 163:399–413

Hong H, Liu J, Zhu AX (2019) Landslide susceptibility evaluating using artificial intelligence method in the Youfang district (China). Environ Earth Sci 78:488

Hosseinalizadeh M, Kariminejad N, Chen W, Pourghasemi HR, Alinejad M, Behbahani AM, Tiefenbacher JP (2019) Gully headcut susceptibility modeling using functional trees, na ve Bayes tree, and random forest models. Geoderma 342:1–11

Huang Y, Zhao L (2018) Review on landslide susceptibility mapping using support vector machines. CATENA 165:520–529

Huang H, Song K, Yi W, Long J, Liu Q, Zhang G (2019) Use of multi-source remote sensing images to describe the sudden Shanshucao landslide in the Three Gorges Reservoir, China. Bull Eng Geol Env 78:2591–2610

Jaafari A, Najafi A, Pourghasemi HR, Rezaeian J, Sattarian A (2014) GIS-based frequency ratio and index of entropy models for landslide susceptibility assessment in the Caspian forest, northern Iran. Int J Environ Sci Technol 11:909–926

Juliev M, Mergili M, Mondal I, Nurtaev B, Pulatov A, Hübl J (2019) Comparative analysis of statistical methods for landslide susceptibility mapping in the Bostanlik District, Uzbekistan. Sci Total Environ 653:801–814

Kadavi P, Lee C-W, Lee S (2018) Application of ensemble-based machine learning models to landslide susceptibility mapping. Remote Sens 10:1252–1274

Kadavi PR, Lee C-W, Lee S (2019) Landslide-susceptibility mapping in Gangwon-do, South Korea, using logistic regression and decision tree models. Environ Earth Sci 78:116

Kasai M, Yamada T (2019) Topographic effects on frequency-size distribution of landslides triggered by the Hokkaido Eastern Iburi Earthquake in 2018. Earth, Planets Space 71:89–101
Kose DD, Turk T (2019) GIS-based fully automatic landslide susceptibility analysis by weight-of-evidence and frequency ratio methods. Phys Geogr 40:481–501
Kumar N, Reddy GPO, Chatterji S (2013) Evaluation of best first decision tree on categorical soil survey data for land capability classification. Int J Comput Appl 72:5–8
Kumar A, Sharma RK, Bansal VK (2018) Landslide hazard zonation using analytical hierarchy process along National Highway-3 in mid Himalayas of Himachal Pradesh, India. Environ Earth Sci 77:719
Kutlug Sahin E, Colkesen I (2019) Performance analysis of advanced decision tree-based ensemble learning algorithms for landslide susceptibility mapping. Geocarto Int 34:1–23
Lay SU, Pradhan B, Yusoff BZ, Abdallah FA, Aryal J, Park H-J (2019) Data mining and statistical approaches in debris-flow susceptibility modelling using airborne LiDAR Data. Sensors 19:3451–3472
Lee C-F, Huang W-K, Chang Y-L, Chi S-Y, Liao W-C (2018a) Regional landslide susceptibility assessment using multi-stage remote sensing data along the coastal range highway in northeastern Taiwan. Geomorphology 300:113–127
Lee J-H, Sameen MI, Pradhan B, Park H-J (2018b) Modeling landslide susceptibility in data-scarce environments using optimized data mining and statistical methods. Geomorphology 303:284–298
Lei X, Chen W, Avand M, Janizadeh S, Kariminejad N, Shahabi H, Costache R, Shahabi H, Shirzadi A, Mosavi A (2020a) GIS-based machine learning algorithms for gully erosion susceptibility mapping in a semi-arid region of Iran. Remote Sens 12:2478
Lei X, Chen W, Pham BT (2020b) Performance evaluation of GIS-based Artificial Intelligence approaches for landslide susceptibility modeling and spatial patterns analysis. ISPRS Int J Geo Inf 9:443
Lei X, Chen W, Panahi M, Falah F, Rahmati O, Uuemaa E, Kalantari Z, Sofia Santos Ferreira C, Rezaie F, Tiefenbacher JP, Lee S, Bian H (2021) Urban flood modeling using deep-learning approaches in Seoul, South Korea. J Hydrol 601:126684–126701
Li R, Wang N (2019) Landslide susceptibility mapping for the Muchuan County (China): a comparison between bivariate statistical models (WoE, EBF, and IoE) and their ensembles with logistic regression. Symmetry 11:762–784
Li Y, Chen W, Rezaie F, Rahmati O, Davoudi Moghaddam D, Tiefenbacher J, Panahi M, Lee M-J, Kulakowski D, Tien Bui D, Lee S (2021) Debris flows modeling using anthropogenic and geo-environmental factors: developing hybridized deep-learning algorithms. Geocarto Int 36:1–23
Loh W-Y (2011) Classification and regression trees. Wiley Interdiscip Rev Data MinKnowl Disc 1:14–23
Lombardo L, Mai PM (2018) Presenting logistic regression-based landslide susceptibility results. Eng Geol 244:14–24
Ma P, Peng J, Wang Q, Zhuang J, Zhang F (2019) The mechanisms of a loess landslide triggered by diversion-based irrigation: a case study of the South Jingyang Platform, China. Bull Eng Geol Environ 78:4954–4693
Mandal B, Mandal S (2018a) Analytical hierarchy process (AHP) based landslide susceptibility mapping of Lish river basin of eastern Darjeeling Himalaya, India. Adv Space Res 62:3114–3132
Mandal S, Mandal K (2018b) Bivariate statistical index for landslide susceptibility mapping in the Rorachu river basin of eastern Sikkim Himalaya, India. Spat Inf Res 26:59–75
Maxwell AE, Warner TA, Fang F (2018) Implementation of machine-learning classification in remote sensing: an applied review. Int J Remote Sens 39:2784–2817
McKenney DW, Pedlar JH (2003) Spatial models of site index based on climate and soil properties for two boreal tree species in Ontario, Canada. For Ecol Manage 175:497–507
Moayedi H, Mehrabi M, Mosallanezhad M, Rashid ASA, Pradhan B (2019) Modification of landslide susceptibility mapping using optimized PSO-ANN technique. Eng Comput 35:967–984
Mohammadi S, Taiebat H (2016) Finite element simulation of an excavation-triggered landslide using large deformation theory. Eng Geol 205:62–72
Mohammady M, Pourghasemi HR, Amiri M (2019) Land subsidence susceptibility assessment using random forest machine learning algorithm. Environ Earth Sci 78:1–12
Mokhtari M, Abedian S (2019) Spatial prediction of landslide susceptibility in Taleghan basin, Iran. Stoch Environ Res Risk Assess 33:1297–1325
Moore ID, Grayson RB, Ladson AR (1991) Digital terrain modelling: a review of hydrological, geomorphological, and biological applications. Hydrol Process 5:3–30
Naghibi AS, Vafakhah M, Hashemi H, Pradhan B, Alavi JS (2018) Groundwater augmentation through the site selection of floodwater spreading using a data mining approach (Case study: Mashhad Plain, Iran). Water 10:1405–1605

Nguyen TP, Tuyen TT, Shirzadi A, Pham TB, Shahabi H, Omidvar E, Amini A, Entezami H, Prakash I, Phong VT, Vu BT, Thanh T, Saro L, Bui TD (2019a) Development of a novel hybrid intelligence approach for landslide spatial prediction. Appl Sci 9:2824–2850

Nguyen VV, Pham TB, Vu TB, Prakash I, Jha S, Shahabi H, Shirzadi A, Ba ND, Kumar R, Chatterjee MJ, Tien Bui D (2019b) Hybrid machine learning approaches for landslide susceptibility modeling. Forests 10:1–27

Palmisano F, Vitone C, Cotecchia F (2016) Methodology for landslide damage assessment. Procedia Engineering 161:511–515

Paranunzio R, Chiarle M, Laio F, Nigrelli G, Turconi L, Luino F (2019) New insights in the relation between climate and slope failures at high-elevation sites. Theoret Appl Climatol 137:1765–1784

Park HJ, Jang JY, Lee JH (2019) Assessment of rainfall-induced landslide susceptibility at the regional scale using a physically based model and fuzzy-based Monte Carlo simulation. Landslides 16:695–713

Peng J, Tong X, Wang S, Ma P (2018) Three-dimensional geological structures and sliding factors and modes of loess landslides. Environ Earth Sci 77:675

Pham BT, Prakash I (2019) A novel hybrid model of Bagging-based Naïve Bayes Trees for landslide susceptibility assessment. Bull Eng Geol Env 78:1911–1925

Pham BT, Bui DT, Dholakia MB, Prakash I, Pham HV, Mehmood K, Le HQ (2017a) A novel ensemble classifier of rotation forest and Naïve Bayer for landslide susceptibility assessment at the Luc Yen district, Yen Bai Province (Viet Nam) using GIS. Geomat Nat Haz Risk 8:649–671

Pham BT, Tien Bui D, Pourghasemi HR, Indra P, Dholakia M (2017b) Landslide susceptibility assesssment in the Uttarakhand area (India) using GIS: a comparison study of prediction capability of naïve bayes, multilayer perceptron neural networks, and functional trees methods. Theoret Appl Climatol 128:255–273

Pham BT, Tien Bui D, Prakash I (2017c) Landslide susceptibility assessment using bagging ensemble based alternating decision trees, logistic regression and j48 decision trees methods: a comparative study. Geotech Geol Eng 35:2597–2611

Pham BT, Nguyen MD, Bui K-TT, Prakash I, Chapi K, Bui DT (2019) A novel artificial intelligence approach based on Multi-layer Perceptron Neural Network and Biogeography-based Optimization for predicting coefficient of consolidation of soil. CATENA 173:302–311

Pourghasemi HR, Rahmati O (2018) Prediction of the landslide susceptibility: Which algorithm, which precision? CATENA 162:177–192

Pourghasemi HR, Rossi M (2017) Landslide susceptibility modeling in a landslide prone area in Mazandarn Province, north of Iran: a comparison between GLM, GAM, MARS, and M-AHP methods. Theoret Appl Climatol 130:609–633

Pourghasemi HR, Moradi HR, Fatemi Aghda SM (2013a) Landslide susceptibility mapping by binary logistic regression, analytical hierarchy process, and statistical index models and assessment of their performances. Nat Hazards 69:749–779

Pourghasemi HR, Pradhan B, Gokceoglu C, Mohammadi M, Moradi HR (2013b) Application of weights-of-evidence and certainty factor models and their comparison in landslide susceptibility mapping at Haraz watershed. Iran Arab J Geosci 6:2351–2365

Pourghasemi HR, Teimoori Yansari Z, Panagos P, Pradhan B (2018) Analysis and evaluation of landslide susceptibility: a review on articles published during 2005–2016 (periods of 2005–2012 and 2013–2016). Arab J Geosci 11:193

Pradhan B, Seeni MI, Kalantar B (2017) Performance evaluation and sensitivity analysis of expert-based, statistical, machine learning, and hybrid models for producing landslide susceptibility maps, Laser scanning applications in landslide assessment. Springer, pp 193–232

Rahmati O, Falah F, Naghibi SA, Biggs T, Soltani M, Deo RC, Cerdà A, Mohammadi F, Tien Bui D (2019) Land subsidence modelling using tree-based machine learning algorithms. Sci Total Environ 672:239–252

Regmi NR, McDonald EV, Rasmussen C (2019) Hillslope response under variable microclimate. Earth Surf Proc Land 44:2615–2627

Reichenbach P, Rossi M, Malamud BD, Mihir M, Guzzetti F (2018) A review of statistically-based landslide susceptibility models. Earth Sci Rev 180:60–91

Sarkar S, Raj R, Vinay S, Maiti J, Pratihar DK (2019) An optimization-based decision tree approach for predicting slip-trip-fall accidents at work. Saf Sci 118:57–69

Schifman LA, Prues A, Gilkey K, Shuster WD (2018) Realizing the opportunities of black carbon in urban soils: Implications for water quality management with green infrastructure. Sci Total Environ 644:1027–1035

Schmidt AH, Denn AR, Hidy AJ, Bierman PR, Tang Y (2019) Human and natural controls on erosion in the Lower Jinsha River, China. J Asian Earth Sci 170:351–359
Shirzadi A, Solaimani K, Roshan MH, Kavian A, Chapi K, Shahabi H, Keesstra S, Ahmad BB, Bui DT (2019) Uncertainties of prediction accuracy in shallow landslide modeling: Sample size and raster resolution. CATENA 178:172–188
Shou K-J, Lin JF (2020) Evaluation of the extreme rainfall predictions and their impact on landslide susceptibility in a sub-catchment scale. Eng Geol 265:105434
Singh K, Kumar V (2017) Landslide hazard mapping along national highway-154A in Himachal Pradesh, India using information value and frequency ratio. Arab J Geosci 10:539
Tien Bui D, Ho T-C, Pradhan B, Pham B-T, Nhu V-H, Revhaug I (2016) GIS-based modeling of rainfall-induced landslides using data mining-based functional trees classifier with AdaBoost, Bagging, and MultiBoost ensemble frameworks. Environ Earth Sci 75:1101
Tien Bui D, Shahabi H, Omidvar E, Shirzadi A, Geertsema M, Clague JJ, Khosravi K, Pradhan B, Pham TB, Chapi K, Barati Z, Bin Ahmad B, Gróf Rahmani H, Lee G (2019) Shallow landslide prediction using a novel hybrid functional machine learning algorithm. Remote Sens 11:931–953
Truong X, Mitamura M, Kono Y, Raghavan V, Yonezawa G, Truong X, Do T, Tien Bui D, Lee S (2018) Enhancing prediction performance of landslide susceptibility model using hybrid machine learning approach of bagging ensemble and logistic model tree. Appl Sci 8:1046–1067
Wang J, Zhang D, Wang N, Gu T (2019a) Mechanisms of wetting-induced loess slope failures. Landslides 16:937–953
Wang K, Zhang S, DelgadoTéllez R, Wei F (2019b) A new slope unit extraction method for regional landslide analysis based on morphological image analysis. Bull Eng Geol Env 78:4139–4151
Wang X, Huang Z, Hong MM, Zhao YF, Ou YS, Zhang J (2019c) A comparison of the effects of natural vegetation regrowth with a plantation scheme on soil structure in a geological hazard-prone region. Eur J Soil Sci 70:674–685
Wang Y, Wu X, Chen Z, Ren F, Feng L, Du Q (2019d) Optimizing the predictive ability of machine learning methods for landslide susceptibility mapping using SMOTE for Lishui City in Zhejiang Province, China. Int J Environ Res Pub Health 16:368–400
Watakabe T, Matsushi Y (2019) Lithological controls on hydrological processes that trigger shallow landslides: observations from granite and hornfels hillslopes in Hiroshima, Japan. CATENA 180:55–68
Wu Y, Li W, Wang Q, Liu Q, Yang D, Xing M, Pei Y, Yan S (2016) Landslide susceptibility assessment using frequency ratio, statistical index and certainty factor models for the Gangu County, China. Arab J Geosci 9:84
Wu Z, Wu Y, Yang Y, Chen F, Zhang N, Ke Y, Li W (2017) A comparative study on the landslide susceptibility mapping using logistic regression and statistical index models. Arab J Geosci 10:187
Xiao L, Zhang Y, Peng G (2018) Landslide susceptibility assessment using integrated deep learning algorithm along the China-Nepal Highway. Sensors 18:4436–4449
Yang Z, Cai H, Shao W, Huang D, Uchimura T, Lei X, Tian H, Qiao J (2019) Clarifying the hydrological mechanisms and thresholds for rainfall-induced landslide: in situ monitoring of big data to unsaturated slope stability analysis. Bull Eng Geol Env 78:2139–2150
Yue X-L, Wu S-H, Huang M, Gao J-B, Yin Y-H, Feng A-Q, Gu X-P (2018) Spatial association between landslides and environmental factors over Guizhou Karst Plateau, China. J Mt Sci 15:1987–2000

Authors and Affiliations

Tingyu Zhang[1,2] · Quan Fu[3] · Hao Wang[4] · Fangfang Liu[3] · Huanyuan Wang[1,2] · Ling Han[5]

[1] Key Laboratory of Degraded and Unused Land Consolidation Engineering, The Ministry of Natural Resources, Xi'an, Shaanxi, China

[2] Institute of Land Engineering and Technology, Shaanxi Provincial Land Engineering Construction Group Co., Ltd., Xi'an, Shaanxi, China

[3] Shaanxi Provincial Land Engineering Construction Group Land Survey Planning and Design Institute Co., Ltd., Xi'an, Shaanxi, China

[4] Hanzhong Branch of Shaanxi Land Engineering Construction Group Co., Ltd., Han Zhong, Shaanxi, China

[5] School of Land Engineering, Chang'an University, Xi'an, Shaanxi, China