Bagging Based Machine Learning Algorithms for Landslide Susceptibility Modeling

Tingyu Zhang

The ministry of natural resources: China

Quan Fu

: Shaanxi Provincial Land Engineering Construction Group

Hao Wang

: Shaanxi Provincial Land Engineering Construction Group

Fangfang Liu

: Shaanxi Provincial Land Engineering Construction Group

Huanyuan Wang (✉ whysxdj2021@163.com)

Key laboratory of Degraded and Unused Land Consolidation Engineering

Ling Han

Changan University: Chang'an University

Research Article

Keywords: Landslide susceptibility, Bagging, Best-first decision tree, Functional tree, Classification and regression tree, support vector machine

Posted Date: July 7th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-553248/v1>

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

[Read Full License](#)

Version of Record: A version of this preprint was published at Natural Hazards on August 21st, 2021. See the published version at <https://doi.org/10.1007/s11069-021-04986-1>.

Bagging based machine learning algorithms for landslide susceptibility modeling

Tingyu Zhang ^{a,b}, Quan Fu ^c, Hao Wang ^d, Fangfang Liu ^c, Huanyuan Wang ^{a,b,*}, Ling Han ^e

^a Key Laboratory of Degraded and Unused Land Consolidation Engineering, the Ministry of Natural Resources, Xi'an, Shaanxi, China.

^b Institute of Land Engineering and Technology, Shaanxi Provincial Land Engineering Construction Group Co., Ltd., Xi'an, Shaanxi, China.

^c Shaanxi Provincial Land Engineering Construction Group Land Survey Planning and Design Institute Co., Ltd., Xi'an, Shaanxi, China.

^d Hanzhong Branch of Shaanxi Land Engineering Construction Group Co., Ltd., Han Zhong, Shaanxi, China.

^e School of Land Engineering, Chang'an University, Xi'an, Shaanxi, China.

** Corresponding author at: Key Laboratory of Degraded and Unused Land Consolidation Engineering, the Ministry of Natural Resources, Institute of Land Engineering and Technology, Shaanxi Provincial Land Engineering Construction Group Co., Ltd., Xi'an 710064, Shaanxi, China.*

Email address: whysxdj2021@163.com (Huanyuan Wang).

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 (FT), support vector machine (SVM) 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 (CAE) 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 (ROC) 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

1. Introduction

As the process of urbanization have 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 (Huang and Li, 2011; Palmisano et al., 2016; Huang et al., 2019). 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 Taiebat, 2016; 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

68 landslide susceptibility assessment (Akgun and Erkan, 2016). Generally, deterministic models are more
69 suitable to evaluate stability of single slope as a result of relatively complex modelling process and higher
70 computational cost (Park et al., 2019). In recent years, under the development of data mining techniques
71 and machine learning, large quantity of novel models have been employed to classify landslide
72 susceptibility zones. According to literature, the most prevailing machine learning models include
73 support vector machine (SVM) (Chang et al., 2019; Nguyen et al., 2019a), artificial neural network (ANN)
74 (Dou et al., 2019b), multi-layer perceptron (MLP) (Hong et al., 2019; Pham et al., 2019), decision tree
75 (DT) (Kutlug Sahin and Colkesen, 2019), etc. These machine learning models can efficiently reveal the
76 intricate regulations which are hidden in huge amounts of data, which has benefits to produce more
77 reliable results. Currently, these machine learning models have been accepted by many scholars in
78 different domains due to their prominent predictive performance (Amiri et al., 2019; Choubin et al., 2019;
79 Hosseinalizadeh et al., 2019; Mohammady et al., 2019). In some cases, the performance of machine
80 learning models dramatically varies with databases, indicating that there still exists latent capacity to
81 promote in generalization performance (Pourghasemi and Rahmati, 2018). Hence, a lot of optimization
82 algorithms and ensemble strategies were integrated with conventional machine learning models to
83 improve their comprehensive performance (Hong et al., 2018b; Lee et al., 2018b; Dou et al., 2019a;
84 Moayedi et al., 2019). Especially for ensemble learning models, they can always generate more satisfying
85 results with lower variance and bias values. Compared with other ensemble techniques, the Bagging
86 ensemble exhibits superiority in well-understood theory, brief framework and promising results (Pham
87 et al., 2017c; Pham and Prakash, 2019). Although numerous Bagging based hybrid models such as J48-
88 bagging (Hong et al., 2019), BKLR (Chen et al., 2018a) and bagging-FT models (Tien Bui et al., 2016)
89 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 with 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 km², 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 m 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 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. Afterwards, the total landslide samples were divided into training data set (70%) and validation data set (30%).

Fig.1 Study area

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, $0-10^\circ$, $10-20^\circ$, $20-30^\circ$, $30-40^\circ$, $40-50^\circ$, $50-60^\circ$, $60-70^\circ$ and $70-78.59^\circ$.

In some cases, landslide are 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 $\ln(\frac{\alpha}{\tan \beta})$, where α indicates the flow accumulation via a point, and β is the corresponding tilt angle (Chen et al., 2021a). 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 $SPI = \alpha \tan \beta$, 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 modelling, and its value can be determined by $STI = (\frac{\alpha}{22.13})^{0.6} \times (\frac{\sin \beta}{0.0896})^{1.3}$ (Hong et al., 2018a). 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. Fig. 2 (i) and (j) demonstrates 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 $NDVI = (IR - R) / (IR + R)$, among them 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 distance to rivers and river density (Arabameri et al., 2019; 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 > 800m.

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 meters as an interval, viz., 0-200 m, 200-400 m, 400-600 m, 600-800 m and > 800m.

The connection between landslide formation and geological structure has been revealed by numerous studies (Pourghasemi et al., 2013b; Pourghasemi et al., 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 2000m as an interval segment, including 0-2000, 2000-4000, 4000-6000, 6000-8000 and > 8000.

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 - 1,000; (7) 1,000 - 1,100; (8) 1,100 - 1,200; (9) 1,200 - 1,284
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
		(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)
9	Lithology	(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
10	Soil	(1) Unused land; (2) Construction land; (3) Grass land; (4) Forest land; (5) Farm land; (6) Water
11	Land use	(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
12	NDVI	(1) 0 - 200; (2) 200 - 400; (3) 400 - 600; (4) 600 - 800; (5) > 800
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 - 2,000; (2) 2,000 - 4,000; (3) 4,000 - 6,000; (4) 6,000 - 8,000; (5) > 8,000
15	Distance to lineaments (m)	(1) 0 - 200; (2) 200 - 400; (3) 400 - 600; (4) 600 - 800; (5) > 800

203

204

205

206

207

208

209

210

211

Fig. 2 Landslide conditioning factors

212

3. Modelling 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 with 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 following formulas:

$$\text{Information Gain}(D, A) = \text{Entropy}(D) - \sum \frac{|D_i|}{|D|} \text{Entropy}(D_i) \quad (1)$$

$$\text{Gini}(p) = \sum_{k=1}^K p_k (1 - p_k) \quad (2)$$

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 p_k 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 modulus, and this approach can be seen in a number of studies such as landslide susceptibility, groundwater potential assessment, land subsidence

modelling, 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 Zhao, 2018). 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 modelling 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 (Eq. (3)), polynomial kernel function (Eq. (4)) and Gaussian kernel function (Eq. (5)).

$$K(x, z) = x \cdot z \quad (3)$$

$$K(x, z) = (\gamma x \cdot z + r)^d \quad (4)$$

$$K(x, z) = \exp(-\gamma \|x - z\|^2) \quad (5)$$

4. Results

4.1 Models results and analysis

In landslide susceptibility modelling, 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 Evaluation (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 (0.120), TWI (0.118), distance to roads (0.104), STI (0.083), SPI (0.069), distance to river (0.060), NDVI (0.042), profile curvature (0.024), lithology (0.018), soil (0.014), plan curvature (0.013), slope aspect (0.012) and distance to lineaments (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

293 4.2 Models validation

294 In this case, four benchmark models were combined with bagging strategy to generate various ensemble
295 models, viz., Bag-BFTree model, Bag-FT model, Bag-CART model and Bag-SVM model. In this section,
296 AUC value corresponding to ROC curve and validation samples were used to measure the accuracy and
297 predictive capacity of benchmark models and integrated models, respectively (Chen et al., 2021c). The
298 results are shown in Fig. 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.

Fig. 3 ROC curves for the models

Fig. 4 ROC curves for the benchmark models

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. Afterwards, the probability of landslide occurrence in each raster can be predicted using four trained bagging-based 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.

Fig. 5 Landslide susceptibility maps for hybrid models

Fig. 6 Percentages of landslide susceptibility classes

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, 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.

Declarations

Funding: This study is financially supported by Fundamental Research Funds for the Central Universities (300102351502), Opening Fund of Key Laboratory of Land Remediation of Shaanxi Province, and Inner scientific research project of Shaanxi Land Engineering Construction Group (SXDJ2021-10).

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.

Availability of data and material: Yes.

Code availability: Yes.

Authors' contributions: Tingyu Zhang: Conceptualization, Methodology, Writing-Review & Editing,

403 Funding Acquisition Quan Fu: Resources, Software, Validation Fangfang Liu: Formal Analysis, Data
404 Curation Hao Wang: Visualization Huanyuan Wang: Writing-Original Draft Preparation Ling Han:
405 Supervision.

406 Ethics approval: Not applicable.

407 Consent to participate: Yes.

408 Consent for publication: Yes.

409 **References**

- 410 Abuzied, S.M., Alrefaee, H.A., 2019. Spatial prediction of landslide-susceptible zones in El-Qaá area,
411 Egypt, using an integrated approach based on GIS statistical analysis. *Bulletin of Engineering*
412 *Geology and the Environment*, 78, 2169-2195.
- 413 Aditian, A., Kubota, T., Shinohara, Y., 2018. Comparison of GIS-based landslide susceptibility models
414 using frequency ratio, logistic regression, and artificial neural network in a tertiary region of
415 Ambon, Indonesia. *Geomorphology*, 318, 101–111.
- 416 Aertsen, W., Kint, V., van Orshoven, J., Özkan, K., Muys, B., 2010. Comparison and ranking of
417 different modelling techniques for prediction of site index in Mediterranean mountain forests.
418 *Ecological Modelling*, 221, 1119-1130.
- 419 Akgun, A., Erkan, O., 2016. Landslide susceptibility mapping by geographical information system-
420 based multivariate statistical and deterministic models: in an artificial reservoir area at
421 Northern Turkey. *Arabian Journal of Geosciences*, 9, 165.
- 422 Amiri, M., Pourghasemi, H.R., Ghanbarian, G.A., Afzali, S.F., 2019. Assessment of the importance of
423 gully erosion effective factors using Boruta algorithm and its spatial modeling and mapping

424 using three machine learning algorithms. *Geoderma*, 340, 55-69.

425 Arabameri, A., Pradhan, B., Rezaei, K., Lee, S., Sohrabi, M., 2019. An ensemble model for landslide
 426 susceptibility mapping in a forested area. *Geocarto International*, 1-26.

427 Armaş, I., 2012. Weights of evidence method for landslide susceptibility mapping. Prahova
 428 Subcarpathians, Romania. *Natural Hazards*, 60, 937-950.

429 Binaghi, E., Luzi, L., Madella, P., Pergalani, F., Rampini, A., 1998. Slope Instability Zonation: a
 430 Comparison Between Certainty Factor and Fuzzy Dempster–Shafer Approaches. *Natural*
 431 *Hazards*, 17, 77-97.

432 Bourenane, H., Bouhadad, Y., Guettouche, M.S., Braham, M., 2015. GIS-based landslide susceptibility
 433 zonation using bivariate statistical and expert approaches in the city of Constantine (Northeast
 434 Algeria). *Bulletin of Engineering Geology and the Environment*, 74, 337-355.

435 Breiman, L., 1996. Bagging predictors. *Machine Learning*, 24, 123-140.

436 Chang, K.-T., Merghadi, A., Yunus, A.P., Pham, B.T., Dou, J., 2019. Evaluating scale effects of
 437 topographic variables in landslide susceptibility models using GIS-based machine learning
 438 techniques. *Scientific Reports*, 9, 12296.

439 Chen, W., Lei, X., Chakraborty, R., Chandra Pal, S., Sahana, M., Janizadeh, S., 2021a. Evaluation of
 440 different boosting ensemble machine learning models and novel deep learning and boosting
 441 framework for head-cut gully erosion susceptibility. *Journal of Environmental Management*,
 442 284, 112015.

443 Chen, W., Shahabi, H., Zhang, S., Khosravi, K., Shirzadi, A., Chapi, K., Pham, T.B., Zhang, T., Zhang,
 444 L., Chai, H., Ma, J., Chen, Y., Wang, X., Li, R., Ahmad, B.B., 2018a. Landslide Susceptibility
 445 Modeling Based on GIS and Novel Bagging-Based Kernel Logistic Regression. *Applied*

446 Sciences, 8.

447 Chen, W., Zhang, S., Li, R., Shahabi, H., 2018b. Performance evaluation of the GIS-based data mining
448 techniques of best-first decision tree, random forest, and naïve Bayes tree for landslide
449 susceptibility modeling. *Science of The Total Environment*, 644, 1006-1018.

450 Chen, Y., Chen, W., Chandra Pal, S., Saha, A., Chowdhuri, I., Adeli, B., Janizadeh, S., Dineva, A.A.,
451 Wang, X., Mosavi, A., 2021b. Evaluation Efficiency of Hybrid Deep Learning algorithms with
452 Neural Network, Decision Tree and Boosting Methods for Predicting Groundwater Potential.
453 Geocarto International, 1-20.

454 Chen, Y., Chen, W., Janizadeh, S., Bhunia, G.S., Bera, A., Pham, Q.B., Linh, N.T.T., Balogun, A.-L.,
455 Wang, X., 2021c. Deep learning and boosting framework for piping erosion susceptibility
456 modeling: spatial evaluation of agricultural areas in the semi-arid region. *Geocarto*
457 *International*, 1-27.

458 Cheng, M.-Y., Hoang, N.-D., 2015. Typhoon-induced slope collapse assessment using a novel bee
459 colony optimized support vector classifier. *Natural Hazards*, 78, 1961-1978.

460 Choi, J., Oh, H.-J., Lee, H.-J., Lee, C., Lee, S., 2012. Combining landslide susceptibility maps obtained
461 from frequency ratio, logistic regression, and artificial neural network models using ASTER
462 images and GIS. *Engineering Geology*, 124, 12-23.

463 Choubin, B., Moradi, E., Golshan, M., Adamowski, J., Sajedi-Hosseini, F., Mosavi, A., 2019. An
464 ensemble prediction of flood susceptibility using multivariate discriminant analysis,
465 classification and regression trees, and support vector machines. *Science of The Total*
466 *Environment*, 651, 2087-2096.

467 Clapuyt, F., Vanacker, V., Christl, M., Van Oost, K., Schlunegger, F., 2019. Spatio-temporal dynamics

468 of sediment transfer systems in landslide-prone Alpine catchments. *Solid Earth*, 10, 1489-
469 1503.

470 Ding, Q., Chen, W., Hong, H., 2017. Application of frequency ratio, weights of evidence and evidential
471 belief function models in landslide susceptibility mapping. *Geocarto International*, 32, 619-
472 639.

473 Dou, J., Yunus, A.P., Xu, Y., Zhu, Z., Chen, C.-W., Sahana, M., Khosravi, K., Yang, Y., Pham, B.T.,
474 2019a. Torrential rainfall-triggered shallow landslide characteristics and susceptibility
475 assessment using ensemble data-driven models in the Dongjiang Reservoir Watershed, China.
476 *Natural Hazards*, 97, 579-609.

477 Dou, J., Yunus, P.A., Tien Bui, D., Sahana, M., Chen, C.-W., Zhu, Z., Wang, W., Thai Pham, B., 2019b.
478 Evaluating GIS-Based Multiple Statistical Models and Data Mining for Earthquake and
479 Rainfall-Induced Landslide Susceptibility Using the LiDAR DEM. *Remote Sensing*, 11.

480 Du, G., Zhang, Y., Yang, Z., Guo, C., Yao, X., Sun, D., 2019. Landslide susceptibility mapping in the
481 region of eastern Himalayan syntaxis, Tibetan Plateau, China: a comparison between
482 analytical hierarchy process information value and logistic regression-information value
483 methods. *Bulletin of Engineering Geology and the Environment*, 78, 4201-4215.

484 Gama, J., 2001. Functional trees for classification, *Proceedings 2001 IEEE International Conference on*
485 *Data Mining*, pp. 147-154.

486 Gama, J., 2004. Functional Trees. *Machine Learning*, 55, 219-250.

487 Haque, U., da Silva, P.F., Devoli, G., Pilz, J., Zhao, B., Khaloua, A., Wilopo, W., Andersen, P., Lu, P.,
488 Lee, J., Yamamoto, T., Keellings, D., Wu, J.-H., Glass, G.E., 2019. The human cost of global
489 warming: Deadly landslides and their triggers (1995–2014). *Science of The Total*

490 Environment, 682, 673-684.

491 He, Q., Xu, Z., Li, S., Li, R., Zhang, S., Wang, N., Pham, T.B., Chen, W., 2019. Novel Entropy and
 492 Rotation Forest-Based Credal Decision Tree Classifier for Landslide Susceptibility Modeling.
 493 Entropy, 21.

494 Hong, H., Kornejady, A., Soltani, A., Termeh, S.V.R., Liu, J., Zhu, A.X., hesar, A.Y., Ahmad, B.B.,
 495 Wang, Y., 2018a. Landslide susceptibility assessment in the Anfu County, China: comparing
 496 different statistical and probabilistic models considering the new topo-hydrological factor
 497 (HAND). Earth Science Informatics, 11, 605-622.

498 Hong, H., Liu, J., Bui, D.T., Pradhan, B., Acharya, T.D., Pham, B.T., Zhu, A.X., Chen, W., Ahmad,
 499 B.B., 2018b. Landslide susceptibility mapping using J48 Decision Tree with AdaBoost,
 500 Bagging and Rotation Forest ensembles in the Guangchang area (China). CATENA, 163, 399-
 501 413.

502 Hong, H., Liu, J., Zhu, A.X., 2019. Landslide susceptibility evaluating using artificial intelligence
 503 method in the Youfang district (China). Environmental Earth Sciences, 78, 488.

504 Hong, H., Pradhan, B., Bui, D.T., Xu, C., Youssef, A.M., Chen, W., 2017. Comparison of four kernel
 505 functions used in support vector machines for landslide susceptibility mapping: a case study at
 506 Suichuan area (China). Geomatics, Natural Hazards and Risk, 8, 544-569.

507 Hosseinalizadeh, M., Kariminejad, N., Chen, W., Pourghasemi, H.R., Alinejad, M., Behbahani, A.M.,
 508 Tiefenbacher, J.P., 2019. Gully headcut susceptibility modeling using functional trees, na ve
 509 Bayes tree, and random forest models. Geoderma, 342, 1-11.

510 Huang, H., Song, K., Yi, W., Long, J., Liu, Q., Zhang, G., 2019. Use of multi-source remote sensing
 511 images to describe the sudden Shanshucao landslide in the Three Gorges Reservoir, China.

512 Bulletin of Engineering Geology and the Environment, 78, 2591-2610.

513 Huang, R., Li, W., 2011. Formation, distribution and risk control of landslides in China. Journal of

514 Rock Mechanics and Geotechnical Engineering, 3, 97-116.

515 Huang, Y., Zhao, L., 2018. Review on landslide susceptibility mapping using support vector machines.

516 Catena, 165, 520-529.

517 Jaafari, A., Najafi, A., Pourghasemi, H.R., Rezaeian, J., Sattarian, A., 2014. GIS-based frequency ratio

518 and index of entropy models for landslide susceptibility assessment in the Caspian forest,

519 northern Iran. International Journal of Environmental Science and Technology, 11, 909-926.

520 Juliev, M., Mergili, M., Mondal, I., Nurtaev, B., Pulatov, A., Hübl, J., 2019. Comparative analysis of

521 statistical methods for landslide susceptibility mapping in the Bostanlik District, Uzbekistan.

522 Science of The Total Environment, 653, 801-814.

523 Kadavi, P., Lee, C.-W., Lee, S., 2018. Application of Ensemble-Based Machine Learning Models to

524 Landslide Susceptibility Mapping. Remote Sensing, 10.

525 Kadavi, P.R., Lee, C.-W., Lee, S., 2019. Landslide-susceptibility mapping in Gangwon-do, South

526 Korea, using logistic regression and decision tree models. Environmental Earth Sciences, 78,

527 116.

528 Kasai, M., Yamada, T., 2019. Topographic effects on frequency-size distribution of landslides triggered

529 by the Hokkaido Eastern Iburi Earthquake in 2018. Earth, Planets and Space, 71, 89.

530 Kose, D.D., Turk, T., 2019. GIS-based fully automatic landslide susceptibility analysis by weight-of-

531 evidence and frequency ratio methods. Physical Geography, 40, 481-501.

532 Kumar, A., Sharma, R.K., Bansal, V.K., 2018. Landslide hazard zonation using analytical hierarchy

533 process along National Highway-3 in mid Himalayas of Himachal Pradesh, India.

534 Environmental Earth Sciences, 77, 719.

535 Kumar, N., Reddy, G.P.O., Chatterji, S., 2013. Evaluation of Best First Decision Tree on Categorical

536 Soil Survey Data for Land Capability Classification. International Journal of Computer

537 Applications, 72, 5-8.

538 Kutlug Sahin, E., Colkesen, I., 2019. Performance analysis of advanced decision tree-based ensemble

539 learning algorithms for landslide susceptibility mapping. Geocarto International, 1-23.

540 Lay, S.U., Pradhan, B., Yusoff, B.Z., Abdallah, F.A., Aryal, J., Park, H.-J., 2019. Data Mining and

541 Statistical Approaches in Debris-Flow Susceptibility Modelling Using Airborne LiDAR Data.

542 Sensors, 19.

543 Lee, C.-F., Huang, W.-K., Chang, Y.-L., Chi, S.-Y., Liao, W.-C., 2018a. Regional landslide

544 susceptibility assessment using multi-stage remote sensing data along the coastal range

545 highway in northeastern Taiwan. Geomorphology, 300, 113-127.

546 Lee, J.-H., Sameen, M.I., Pradhan, B., Park, H.-J., 2018b. Modeling landslide susceptibility in data-

547 scarce environments using optimized data mining and statistical methods. Geomorphology,

548 303, 284-298.

549 Lei, X., Chen, W., Avand, M., Janizadeh, S., Kariminejad, N., Shahabi, H., Costache, R., Shahabi, H.,

550 Shirzadi, A., Mosavi, A., 2020a. GIS-Based Machine Learning Algorithms for Gully Erosion

551 Susceptibility Mapping in a Semi-Arid Region of Iran. Remote Sensing, 12, 2478.

552 Lei, X., Chen, W., Pham, B.T., 2020b. Performance Evaluation of GIS-Based Artificial Intelligence

553 Approaches for Landslide Susceptibility Modeling and Spatial Patterns Analysis. ISPRS

554 International Journal of Geo-Information, 9, 443.

555 Li, R., Wang, N., 2019. Landslide Susceptibility Mapping for the Muchuan County (China): A

556 Comparison Between Bivariate Statistical Models (WoE, EBF, and IoE) and Their Ensembles
557 with Logistic Regression. *Symmetry*, 11.

558 Li, Y., Chen, W., Rezaie, F., Rahmati, O., Davoudi Moghaddam, D., Tiefenbacher, J., Panahi, M., Lee,
559 M.-J., Kulakowski, D., Tien Bui, D., Lee, S., 2021. Debris flows modeling using
560 anthropogenic and geo-environmental factors: developing hybridized deep-learning
561 algorithms. *Geocarto International*, 1-23.

562 Loh, W.-Y., 2011. Classification and regression trees. *Wiley Interdisciplinary Reviews: Data Mining*
563 and Knowledge Discovery, 1, 14-23.

564 Lombardo, L., Mai, P.M., 2018. Presenting logistic regression-based landslide susceptibility results.
565 *Engineering geology*, 244, 14-24.

566 Ma, P., Peng, J., Wang, Q., Zhuang, J., Zhang, F., 2019. The mechanisms of a loess landslide triggered
567 by diversion-based irrigation: a case study of the South Jingyang Platform, China. *Bulletin of*
568 *Engineering Geology and the Environment*.

569 Mandal, B., Mandal, S., 2018a. Analytical hierarchy process (AHP) based landslide susceptibility
570 mapping of Lish river basin of eastern Darjeeling Himalaya, India. *Advances in Space*
571 *Research*, 62, 3114-3132.

572 Mandal, S., Mandal, K., 2018b. Bivariate statistical index for landslide susceptibility mapping in the
573 Rorachu river basin of eastern Sikkim Himalaya, India. *Spatial Information Research*, 26, 59-
574 75.

575 Maxwell, A.E., Warner, T.A., Fang, F., 2018. Implementation of machine-learning classification in
576 remote sensing: an applied review. *International Journal of Remote Sensing*, 39, 2784-2817.

577 McKenney, D.W., Pedlar, J.H., 2003. Spatial models of site index based on climate and soil properties

578 for two boreal tree species in Ontario, Canada. *Forest Ecology and Management*, 175, 497-
579 507.

580 Moayedi, H., Mehrabi, M., Mosallanezhad, M., Rashid, A.S.A., Pradhan, B., 2019. Modification of
581 landslide susceptibility mapping using optimized PSO-ANN technique. *Engineering with*
582 *Computers*, 35, 967-984.

583 Mohammadi, S., Taiebat, H., 2016. Finite element simulation of an excavation-triggered landslide
584 using large deformation theory. *Engineering Geology*, 205, 62-72.

585 Mohammady, M., Pourghasemi, H.R., Amiri, M., 2019. Land subsidence susceptibility assessment
586 using random forest machine learning algorithm. *Environmental Earth Sciences*, 78, 1-12.

587 Mokhtari, M., Abedian, S., 2019. Spatial prediction of landslide susceptibility in Taleghan basin, Iran.
588 *Stochastic Environmental Research and Risk Assessment*, 33, 1297-1325.

589 Moore, I.D., Grayson, R.B., Ladson, A.R., 1991. Digital terrain modelling: A review of hydrological,
590 geomorphological, and biological applications. *Hydrological Processes*, 5, 3-30.

591 Naghibi, A.S., Vafakhah, M., Hashemi, H., Pradhan, B., Alavi, J.S., 2018. Groundwater Augmentation
592 through the Site Selection of Floodwater Spreading Using a Data Mining Approach (Case
593 study: Mashhad Plain, Iran). *Water*, 10.

594 Nguyen, T.P., Tuyen, T.T., Shirzadi, A., Pham, T.B., Shahabi, H., Omidvar, E., Amini, A., Entezami, H.,
595 Prakash, I., Phong, V.T., Vu, B.T., Thanh, T., Saro, L., Bui, T.D., 2019a. Development of a
596 Novel Hybrid Intelligence Approach for Landslide Spatial Prediction. *Applied Sciences*, 9.

597 Nguyen, V.V., Pham, T.B., Vu, T.B., Prakash, I., Jha, S., Shahabi, H., Shirzadi, A., Ba, N.D., Kumar, R.,
598 Chatterjee, M.J., Tien Bui, D., 2019b. Hybrid Machine Learning Approaches for Landslide
599 Susceptibility Modeling. *Forests*, 10.

600 Palmisano, F., Vitone, C., Cotecchia, F., 2016. Methodology for Landslide Damage Assessment.
601 Procedia Engineering, 161, 511-515.

602 Paranunzio, R., Chiarle, M., Laio, F., Nigrelli, G., Turconi, L., Luino, F., 2019. New insights in the
603 relation between climate and slope failures at high-elevation sites. Theoretical and Applied
604 Climatology, 137, 1765-1784.

605 Park, H.J., Jang, J.Y., Lee, J.H., 2019. Assessment of rainfall-induced landslide susceptibility at the
606 regional scale using a physically based model and fuzzy-based Monte Carlo simulation.
607 Landslides, 16, 695-713.

608 Peng, J., Tong, X., Wang, S., Ma, P., 2018. Three-dimensional geological structures and sliding factors
609 and modes of loess landslides. Environmental Earth Sciences, 77, 675.

610 Pham, B.T., Bui, D.T., Dholakia, M.B., Prakash, I., Pham, H.V., Mehmood, K., Le, H.Q., 2017a. A
611 novel ensemble classifier of rotation forest and Naïve Bayer for landslide susceptibility
612 assessment at the Luc Yen district, Yen Bai Province (Viet Nam) using GIS. Geomatics,
613 Natural Hazards and Risk, 8, 649-671.

614 Pham, B.T., Nguyen, M.D., Bui, K.-T.T., Prakash, I., Chapi, K., Bui, D.T., 2019. A novel artificial
615 intelligence approach based on Multi-layer Perceptron Neural Network and Biogeography-
616 based Optimization for predicting coefficient of consolidation of soil. CATENA, 173, 302-
617 311.

618 Pham, B.T., Prakash, I., 2019. A novel hybrid model of Bagging-based Naïve Bayes Trees for landslide
619 susceptibility assessment. Bulletin of Engineering Geology and the Environment, 78, 1911-
620 1925.

621 Pham, B.T., Tien Bui, D., Pourghasemi, H.R., Indra, P., Dholakia, M., 2017b. Landslide susceptibility

622 assessment in the Uttarakhand area (India) using GIS: a comparison study of prediction
 623 capability of naïve bayes, multilayer perceptron neural networks, and functional trees
 624 methods. *Theoretical and Applied Climatology*, 128, 255-273.

625 Pham, B.T., Tien Bui, D., Prakash, I., 2017c. Landslide Susceptibility Assessment Using Bagging
 626 Ensemble Based Alternating Decision Trees, Logistic Regression and J48 Decision Trees
 627 Methods: A Comparative Study. *Geotechnical and Geological Engineering*, 35, 2597-2611.

628 Pourghasemi, H.R., Moradi, H.R., Fatemi Aghda, S.M., 2013a. Landslide susceptibility mapping by
 629 binary logistic regression, analytical hierarchy process, and statistical index models and
 630 assessment of their performances. *Natural Hazards*, 69, 749-779.

631 Pourghasemi, H.R., Pradhan, B., Gokceoglu, C., Mohammadi, M., Moradi, H.R., 2013b. Application of
 632 weights-of-evidence and certainty factor models and their comparison in landslide
 633 susceptibility mapping at Haraz watershed, Iran. *Arabian Journal of Geosciences*, 6, 2351-
 634 2365.

635 Pourghasemi, H.R., Rahmati, O., 2018. Prediction of the landslide susceptibility: Which algorithm,
 636 which precision? *CATENA*, 162, 177-192.

637 Pourghasemi, H.R., Rossi, M., 2017. Landslide susceptibility modeling in a landslide prone area in
 638 Mazandarn Province, north of Iran: a comparison between GLM, GAM, MARS, and M-AHP
 639 methods. *Theoretical and Applied Climatology*, 130, 609-633.

640 Pourghasemi, H.R., Teimoori Yansari, Z., Panagos, P., Pradhan, B., 2018. Analysis and evaluation of
 641 landslide susceptibility: a review on articles published during 2005–2016 (periods of 2005–
 642 2012 and 2013–2016). *Arabian Journal of Geosciences*, 11, 193.

643 Pradhan, B., Seeni, M.I., Kalantar, B., 2017. Performance evaluation and sensitivity analysis of expert-

644 based, statistical, machine learning, and hybrid models for producing landslide susceptibility
645 maps, Laser scanning applications in landslide assessment. Springer, pp. 193-232.

646 Rahmati, O., Falah, F., Naghibi, S.A., Biggs, T., Soltani, M., Deo, R.C., Cerdà, A., Mohammadi, F.,
647 Tien Bui, D., 2019. Land subsidence modelling using tree-based machine learning algorithms.
648 Science of The Total Environment, 672, 239-252.

649 Regmi, N.R., McDonald, E.V., Rasmussen, C., 2019. Hillslope response under variable microclimate.
650 Earth Surface Processes and Landforms, 0.

651 Reichenbach, P., Rossi, M., Malamud, B.D., Mihir, M., Guzzetti, F., 2018. A review of statistically-
652 based landslide susceptibility models. Earth-Science Reviews, 180, 60-91.

653 Sarkar, S., Raj, R., Vinay, S., Maiti, J., Pratihari, D.K., 2019. An optimization-based decision tree
654 approach for predicting slip-trip-fall accidents at work. Safety Science, 118, 57-69.

655 Schifman, L.A., Prues, A., Gilkey, K., Shuster, W.D., 2018. Realizing the opportunities of black carbon
656 in urban soils: Implications for water quality management with green infrastructure. Science
657 of The Total Environment, 644, 1027-1035.

658 Schmidt, A.H., Denn, A.R., Hidy, A.J., Bierman, P.R., Tang, Y., 2019. Human and natural controls on
659 erosion in the Lower Jinsha River, China. Journal of Asian Earth Sciences, 170, 351-359.

660 Shirzadi, A., Solaimani, K., Roshan, M.H., Kaviani, A., Chapi, K., Shahabi, H., Keesstra, S., Ahmad,
661 B.B., Bui, D.T., 2019. Uncertainties of prediction accuracy in shallow landslide modeling:
662 Sample size and raster resolution. CATENA, 178, 172-188.

663 Shou, K.-J., Lin, J.-F., 2020. Evaluation of the extreme rainfall predictions and their impact on
664 landslide susceptibility in a sub-catchment scale. Engineering Geology, 265, 105434.

665 Singh, K., Kumar, V., 2017. Landslide hazard mapping along national highway-154A in Himachal

666 Pradesh, India using information value and frequency ratio. *Arabian Journal of Geosciences*,
667 10, 539.

668 Tien Bui, D., Ho, T.-C., Pradhan, B., Pham, B.-T., Nhu, V.-H., Revhaug, I., 2016. GIS-based modeling
669 of rainfall-induced landslides using data mining-based functional trees classifier with
670 AdaBoost, Bagging, and MultiBoost ensemble frameworks. *Environmental Earth Sciences*,
671 75, 1101.

672 Tien Bui, D., Shahabi, H., Omidvar, E., Shirzadi, A., Geertsema, M., Clague, J.J., Khosravi, K.,
673 Pradhan, B., Pham, T.B., Chapi, K., Barati, Z., Bin Ahmad, B., Rahmani, H., Gróf, G., Lee, S.,
674 2019. Shallow Landslide Prediction Using a Novel Hybrid Functional Machine Learning
675 Algorithm. *Remote Sensing*, 11.

676 Truong, X., Mitamura, M., Kono, Y., Raghavan, V., Yonezawa, G., Truong, X., Do, T., Tien Bui, D.,
677 Lee, S., 2018. Enhancing Prediction Performance of Landslide Susceptibility Model Using
678 Hybrid Machine Learning Approach of Bagging Ensemble and Logistic Model Tree. *Applied*
679 *Sciences*, 8.

680 Wang, J., Zhang, D., Wang, N., Gu, T., 2019a. Mechanisms of wetting-induced loess slope failures.
681 *Landslides*, 16, 937-953.

682 Wang, K., Zhang, S., DelgadoTéllez, R., Wei, F., 2019b. A new slope unit extraction method for
683 regional landslide analysis based on morphological image analysis. *Bulletin of Engineering*
684 *Geology and the Environment*, 78, 4139-4151.

685 Wang, X., Huang, Z., Hong, M.M., Zhao, Y.F., Ou, Y.S., Zhang, J., 2019c. A comparison of the effects
686 of natural vegetation regrowth with a plantation scheme on soil structure in a geological
687 hazard-prone region. *European Journal of Soil Science*, 70, 674-685.

688 Wang, Y., Wu, X., Chen, Z., Ren, F., Feng, L., Du, Q., 2019d. Optimizing the Predictive Ability of
689 Machine Learning Methods for Landslide Susceptibility Mapping Using SMOTE for Lishui
690 City in Zhejiang Province, China. *International Journal of Environmental Research and Public*
691 *Health*, 16.

692 Watakabe, T., Matsushi, Y., 2019. Lithological controls on hydrological processes that trigger shallow
693 landslides: Observations from granite and hornfels hillslopes in Hiroshima, Japan. *CATENA*,
694 180, 55-68.

695 Wu, Y., Li, W., Wang, Q., Liu, Q., Yang, D., Xing, M., Pei, Y., Yan, S., 2016. Landslide susceptibility
696 assessment using frequency ratio, statistical index and certainty factor models for the Gangu
697 County, China. *Arabian Journal of Geosciences*, 9, 84.

698 Wu, Z., Wu, Y., Yang, Y., Chen, F., Zhang, N., Ke, Y., Li, W., 2017. A comparative study on the
699 landslide susceptibility mapping using logistic regression and statistical index models. *Arabian*
700 *Journal of Geosciences*, 10, 187.

701 Xiao, L., Zhang, Y., Peng, G., 2018. Landslide Susceptibility Assessment Using Integrated Deep
702 Learning Algorithm along the China-Nepal Highway. *Sensors*, 18.

703 Yang, Z., Cai, H., Shao, W., Huang, D., Uchimura, T., Lei, X., Tian, H., Qiao, J., 2019. Clarifying the
704 hydrological mechanisms and thresholds for rainfall-induced landslide: in situ monitoring of
705 big data to unsaturated slope stability analysis. *Bulletin of Engineering Geology and the*
706 *Environment*, 78, 2139-2150.

707 Yue, X.-l., Wu, S.-h., Huang, M., Gao, J.-b., Yin, Y.-h., Feng, A.-q., Gu, X.-p., 2018. Spatial association
708 between landslides and environmental factors over Guizhou Karst Plateau, China. *Journal of*
709 *Mountain Science*, 15, 1987-2000.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Graphicalabstract.docx](#)