Comparison of rotation forest based intelligence models for landslide susceptibility modeling

Xia Zhao (✉ 19209071022@stu.xust.edu.cn)

Xi'an University of Science and Technology <https://orcid.org/0000-0002-9403-9330>

Wei Chen

Xi'an University of Science and Technology

Tao Li

Liupanshui Normal University

Faming Huang

Nanchang University

Chaohong Peng

Sichuan Institute of Geological Engineering Investigation Group Co.Ltd

Himan Shahabi

University of Kurdistan

Huiyuan Bian

Xi'an University of Science and Technology

Research Article

Keywords: Landslide Spatial Prediction, Rotation Forest, J48, Alternating Decision Tree, Random Forest, Ensemble Models

Posted Date: March 24th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-346721/v1>

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

[Read Full License](#)

Comparison of rotation forest based intelligence models for landslide susceptibility modeling

Xia Zhao¹, Wei Chen^{1,2}, Tao Li³, Faming Huang⁴, Chaohong Peng⁵, Himan Shahabi^{6,7},
Huiyuan Bian¹

¹College of Geology and Environment, Xi'an University of Science and Technology, Xi'an
710054, China

²Key Laboratory of Coal Resources Exploration and Comprehensive Utilization, Ministry of
Natural Resources, Xi'an 710021, China

³School of Mining & Civil Engineering, Liupanshui Normal University, Liupanshui 553000,
Guizhou, China

⁴School of Civil Engineering and Architecture, Nanchang University, Nanchang 330031, China

⁵Sichuan Institute of Geological Engineering Investigation Group Co. Ltd

⁶Department of Geomorphology, Faculty of Natural Resources, University of Kurdistan,
Sanandaj 66177-15175, Iran.

⁷Board Member of Department of Zrebar Lake Environmental Research, Kurdistan Studies
Institute, University of Kurdistan, Sanandaj 66177-15175, Iran. h.shahabi@uok.ac.ir (H. S.)

*Corresponding author. E-mail: chenwei0930@xust.edu.cn

Abstract:

The precision of landslide susceptibility assessment has always been the focus of landslide spatial prediction research. It can be considered as the possibility of landslide disaster under the action of human activities or natural factors, or both of them. For the further exploration of the mechanism of this process, Muchuan County was proposed as the study area, and four well-known machine learning models, namely rotation forest (RF), J48 decision tree (J48), alternating decision tree (ADTree) and random forest (RaF), and their ensembles (RF-J48, RF-ADTree and RF-RaF) were introduced to explore the mechanism. These models are established by twelve landslide conditioning factors, which are selected based to the local special geological environment conditions and previous related researches, including plan curvature, profile curvature, slope angle, slope aspect, elevation, topographic wetness index (TWI), land use, normalized difference vegetation index (NDVI), soil, lithology, distance to roads, and distance to rivers, as well as training (195 landslides) and validation (84 landslides) datasets were developed. The landslide prediction performance of the above conditioning factors was analyzed through the correlation attribute evaluation (CAE) model. Then, through the landslide susceptibility maps made by the above six different models, the Jenks natural breaks method is used to divide the landslide susceptibility into five grades, which are very low, low, moderate, high, and very high. In addition, the accuracy of the above six landslide susceptibility maps was verified by implementing the relative operating characteristic curve (ROC) and the area under the ROC (AUC). That is, the capabilities of the above six models

are compared and verified in the landslide spatial prediction. Finally, the obtained results show that elevation, lithology and TWI are the three most principal landslide conditioning factors in this research. The RF-RaF and RaF models in the training dataset performed best, with the AUC value of 0.75, while the RF-ADTree model (0.74), RF-J48 model (0.74), ADTree model (0.71) and J48 model (0.70) performed poorly. Meanwhile, similar results also emerge from the validation dataset, in which the RF-RaF model acquired the best performance (0.82) and the rest are the RF-ADTree model (0.80), RaF model (0.79), RF-J48 model (0.77), ADTree model (0.76) and J48 model (0.71). Last but by no means the least, the results can provide scientific references for local natural resources departments.

Keywords: Landslide Spatial Prediction, Rotation Forest, J48, Alternating Decision Tree, Random Forest, Ensemble Models.

1. Introduction

A series of complex problems such as casualties, economic losses and geological environment damage caused by landslides have brought heavy losses to China (Feng et al., 2019; Huang, 2012; Kulatilake and Ge, 2014; Peng et al., 2017). According to the latest statistics of the Ministry of emergency management of the people's Republic of China, a total of 1305 natural disasters occurred in 2020, causing 138 million people affected, 591 people died and disappeared due to the disaster, and the direct economic loss was US \$57.2 billion (<https://www.mem.gov.cn/>). Due to the special geographical location and frequent human

activities, Sichuan has become the most vulnerable province by geological disasters where occupies fourteen percent of the county's geological disasters and more than sixty percent of geological hazards are landslide (Xu et al., 2016; Yin et al., 2010). Therefore, it is imperative to evaluate the susceptibility of landslide to reduce the impact of landslide (Ge et al., 2018; Ramakrishnan et al., 2013). In recent years, various methods for landslide spatial distribution prediction have been applied to landslide susceptibility evaluation. Generally, these evaluation methods can be divided into two aspects: statistical model based on GIS and machine learning model. Statistical models that need to be parameterized by assuming appropriate structural models generally include analytic hierarchy process (AHP) (Myronidis et al., 2016; Park et al., 2013; Pourghasemi et al., 2012), logistic regression (LR) (Ayalew and Yamagishi, 2005; Li and Wang, 2019; Wu, 2015), index of entropy (IoE) (Chen et al., 2015; Mondal and Mandal, 2018), certainty factor (CF) (Devkota et al., 2013; Wang et al., 2015), evidential belief function (EBF) (Althuwaynee et al., 2012; Tehrany and Kumar, 2018; Zhang et al., 2016), frequency ratio (FR) (Mohammady et al., 2012; Sahana and Sajjad, 2017), and weights of evidence (WoE) (Dahal et al., 2008; Xie et al., 2017).

It is very crucial to improve the prediction ability of the model for landslide susceptibility assessment. The intelligent machine learning model avoids the shortcomings of statistics and can get better results, such as fuzzy logic (FL) (Pradhan, 2010), artificial neural network (ANN) (Mandal and Mondal, 2019; Polykretis et al., 2015), decision trees (DT) (Ma et al., 2017; Park and Lee, 2014; Tien Bui et al., 2014a), support vector machine (SVM) (Pradhan, 2013; Ren et al., 2015), multivariate adaptive regression splines (MARS) (Friedman, 1991; Tsige et al., 2016), random forest (RaF) (Chu et al., 2018; Pourghasemi and Kerle, 2016),

adaptive neuro-fuzzy inference systems (ANFIS) (Bui et al., 2011; Polykretis et al., 2017) and Naive Bayes (NB) (Hong et al., 2019; Venkatesan et al., 2013). The methods mentioned above have been skillfully applied to landslide susceptibility prediction and have performed well, but some attempts to improve prediction capabilities should also be started, such as trying new ensemble machine learning algorithms. Therefore, four well-known machine learning methods, namely rotation forest (RF), J48 decision tree (J48), alternating decision tree (ADT) and random forest (RaF), are used to analyze the spatial susceptibility of regional landslides in Muchuan County, Sichuan Province, in this research.

The innovation of this research lies in the application of four machine learning methods and their integrated models to the spatial prediction of landslide susceptibility in Muchuan County. Concurrently, six landslide susceptibility maps generated by the above models are analyzed to select the best modeling method. And the six landslide susceptibility maps were verified by using the area under the receiver operating characteristic (ROC) curve, the area under the ROC (AUC) and statistical methods. Finally, the research results of landslide susceptibility can help the land use management planning and landslide disaster prevention in Muchuan County and other similar geological environment areas.

2. Study area and data preparation

Muchuan County, with an area of 1408 km², is found within the longitudes of 103°21' ~ 104°37'E, and the latitudes of 28°49' ~ 29°31'N (Fig. 1). The altitudes of the study area vary from 290 m to 1866 m according to the digital elevation model (DEM) with a 20 m regular grid. The climate of Muchuan County is warm in winter and cool in summer, abundant in rainfall, overcast and rainy, and it is a humid subtropical monsoon climate. Rainfall tends to

increase from northwest to southeast. As the elevation increases, the annual rainfall gradually increases, and the annual average precipitation is 1332mm. The coldest January is only 7.2°C, while the hottest July has an average temperature of 26.2°C. The frequency of Rainstorm in rainy season from June to September is high, accounting for 65% of the annual precipitation.

Fig. 1 Study area.

It is known that the landslide database is the basic data of landslide susceptibility research, which includes location, scale, cause, an object of threat, potential loss and so on. The 279 landslides were determined, and the landslide centroids were extracted to make landslide points and divided into a training dataset containing 195 landslides and a validation dataset containing 84 landslides at a ratio of 70/30, respectively. On the basis of previous relevant researches, combined with the local geological environment characteristics (Chen et al., 2019b; Li and Wang, 2019), this study compiled 12 landslide conditioning factor data layers, including plan curvature, profile curvature, slope angle, slope aspect, elevation, topographic wetness index (TWI), land use, normalized difference vegetation index (NDVI), soil, lithology, distance to roads and distance to rivers (Table 1 and Fig. 2). The source of conditioning factors of the above landslides include: 20m remote sensing elevation data, 1:100,000 scale land use layer provided by local government, Landsat 8 satellite imagery, 1:200,000 scale soil map provided by the Chinese Academy of Sciences, 1:200,000 scale lithologic map provided by China Geological Survey and 1:5,000 scale topographic map. Using ArcGIS 10.5 software and ENVI 5.3 software, 12 factor layers with resolution of 20m × 20m were prepared, and they were classified according to their own rules. The classification results are shown in Table 2. In addition, the description of each factor is as follows.

As a very significant factor to dominate the concentration and separation of terrain, plan curvature is directly associated with the surface water velocity and the erosion degree of rock and soil mass (Kornejady et al., 2019). The profile curvature is major reflected in the erosion of rock and soil by surface water and the accumulation of slope deposits (Pham et al., 2018; Zhuang et al., 2018). The change of slope angle affects the stress of the slope foot. When the

slope angle is larger, the stress concentration at the slope foot is more obvious. Therefore, the greater the slope angle, the more likely it is to become unstable and form landslides (Xie et al., 2019). The slope aspect has an extremely significant influence in the landslide spatial prediction, which not only affects the slope inclination direction, but also indicates and affects the slope profile, rainfall, sunshine, wind erosion and so on (Braun et al., 2019; Razavizadeh et al., 2017). Elevation, as the most important and commonly used factor in the landslide spatial prediction, will lead to its own nature to affect terrain properties, making the terrain at different altitudes present certain spatial-temporal differences; it also affects temperature changes and vegetation distribution. Temperature and vegetation distribution density are inversely related to elevation. At high elevation, the temperature is low and the vegetation is sparsely distributed. In addition, elevation can also affect the freeze-thaw cycle of rock and soil (Pradhan and Lee, 2010; Sheng and Chen, 2017). TWI considers topography and soil characteristics when considering soil moisture distribution, which has significant theoretical and practical significance for studying spatial distribution of soil moisture in watershed (Canoglu et al., 2018; Raja et al., 2017). Land use refers to the process of landscape transformation. Due to the needs of human production and life, land is divided into different categories. Different types have different characteristics. For example, they can change the characteristics of surface water flow. Surface water in cities is dominated by runoff, while surface water in cultivated land is dominated by infiltration (Diva et al., 2018; Persichillo et al., 2017). NDVI can be used to compare the characteristics of two bands in the multispectral raster dataset to express the vegetation coverage on the earth's surface (Fiorucci et al., 2019; Sun et al., 2017). The vegetation cover can consolidate the stability of the slope, so as to reduce the occurrence

of landslides. Landslides are distributed differently in different soil types, and different soil types will behave differently under the same rainfall intensity. This is mainly due to the differences in mineral composition, structure, density, porosity, shear strength and permeability coefficient among different soil types (Basher et al., 2018; Thomas et al., 2018). The strata of a slope are the material basis of the landslide. Some slopes are made up of hard rock, some of soft rock and some of soil. Due to the different lithology, their shear strength is different, leading to different possibility of landslide occurrence. Slopes of hard rock are steep and stable, with few landslides. On the contrary, slopes composed of shale, mudstone and other surface overburden, such as clay and gravel soil, are more prone to deformation and landslides (Tsangaratos and Ilia, 2017; Watakabe and Matsushi, 2019). Because of the large mountainous area in the study area, road construction is inevitable. It must be noted that the stress concentration at the bottom of the slope caused by road construction may damage the stability of the slope. In addition, the road construction will also form a new water catchment surface, thereby changing the original surface water and groundwater runoff field on the slope (Losasso et al., 2017; Sridhar et al., 2019). To some extent, the distance from the river reflects the hydraulic connections between the slope and the outside. The influence of slope on the river mainly depends on the distance to rivers, and they are positively correlated with each other. On the one hand, the river erodes the rock and soil mass at the foot of the slope through lateral erosion, which aggravates its instability. On the other hand, the water in the river will enter the slope through capillarity, increasing the mass of the slope and reducing the strength of its weak layer, aggravating its instability (Cordeira et al., 2019; Zhao et al., 2017).

Table 1 Classification of landside conditioning factors.

No.	Conditioning factors	Classes
1	Plan curvature	(1) (-26.54) – (-3.17); (2) (-3.17) – (-1.10); (3) (-1.10) - 0.56; (4) 0.56 - 2.62; (5) 2.62 - 26.21
2	Profile curvature	(1) (-34.72) – (-4.17); (2) (-4.17) – (-1.39); (3) (-1.39) - 0.88; (4) 0.88 - 3.66; (5) 3.66 - 29.66
3	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 - 77.14
4	Slope aspect	(1) Flat; (2) North; (3) Northeast; (4) East; (5) Southeast; (6) South; (7) Southwest; (8) West; (9) Northwest
5	Elevation (m)	(1) 290 - 500; (2) 500 - 700; (3) 700 - 900; (4) 900 - 1,100; (5) 1,100 - 1,300; (6) 1,300 - 1,500; (7) 1,500 - 1,700; (8) 1,700 - 1,866
6	TWI	(1) 0.35 - 1.63; (2) 1.63 - 2.41; (3) 2.41 - 3.36; (4) 3.36 - 4.81; (5) 4.81 - 14.58
7	Land use	(1) Farm land; (2) Forest land; (3) Grass land; (4) Water; (5) Residential areas; (6) Bare land
8	NDVI	(1) (-0.13) - 0.13; (2) 0.13 - 0.24; (3) 0.24 - 0.30; (4) 0.30 - 0.37; (5) 0.37 - 0.58
9	Soil	(1) Paddy soil; (2) Acid purple soil; (3) Neutral purple soil; (4) Calcareous purple soil; (5) Calcareous soil; (6) Yellow-brown soil; (7) Yellow calcareous soil; (8) Yellow soil; (9) Dark-yellow-brown soil; (10) Purple soil
10	Lithology	(1) Sandstone with mudstone and siltstone interlayers; (2) Grayish-purple arkose, siltstone with shale and limestone interlayers; Bright red mudstone with sandstone and siltstone interlayers; (3) Yellow-gray feldspar-quartz sandstone interbedded with purple-red mudstone; (4) Aubergine mudstone, quartz sandstone and siltstone with biosparite and marl interlayers; (5) Gray sandstone, siltstone and mudstone; Grayish-yellow arkose, siltstone, mudstone and coal; (6) Yellow-gray dolomite with limestone, gypsum salt and breccia interlayers; (7) Limestone, dolomite and shale; (8) Yellow-green sandstone, siltstone with mudstone and coal interlayers; (9) Limestone, dolomite, shale, siltstone; (10) Grayish-green basalt with picrite, tufaceous sand-mudstone, coal streak and siliceous rock interlayers
11	Distance to roads (m)	(1) <500; (2) 500 - 1,000; (3) 1,000 – 1,500; (4) 1,500 - 2,000; (5) >2,000
12	Distance to rivers (m)	(1) <200; (2) 200 - 400; (3) 400 - 600; (4) 600 - 800; (5) >800

Fig. 2 Landslide conditioning factors: (a) plan curvature, (b) profile curvature, (c) slope angle, (d) slope aspect, (e) elevation, (f) TWI, (g) land use, (h) NDVI, (i) soil, (j) lithology, (k) distance to roads,

and (l) distance to rivers.

3. Modeling approach

3.1 Rotation forest (RF)

As a replacement technique supported the concept of feature transformation proposed by Rodriguez et al. (Rodriguez et al., 2006), the RF model can be used to improve the range and accuracy of basic classifiers (Pardo et al., 2013). Before the extraction of each sub-sample, the sample attribute sets were divided and combined randomly and using principal component analysis (PCA) to transform the data among the sub-attribute sets (Lasota et al., 2012; Zhang and Zhang, 2008). After the processes mentioned above, it not only makes each sub-sample different but also preprocess the data partially (Du et al., 2015). Finally, the diversity and accuracy of each base classifier were improved. In this research, it is assumed that $v = (v_1, v_2, \dots, v_n)$ represents the vector of conditioning factors. $l = (l_0, l_1)$ is the vector of non-landslide and landslide classes. T is the training dataset. $B_1, B_2, \dots, B_L$ are the classifiers in the ensemble framework. Q is a group of conditioning factors composed of m subsets, and each subset contains $H=m/12$ landslide conditioning factors. T_{ij} is the landslide conditioning factor of Q_{ij} in T , and the coefficient $c_{i,1}^{(1)}, c_{i,1}^{(2)}, \dots, c_{i,1}^{(H_i)}$ is obtained by transforming T_{ij} . The radio frequency model is established by the transformed rotation matrix and base classifier. The rotation matrix C_i^α is formed by rearranging the matrix of C_i as follows:

$$C_i = \begin{bmatrix} c_{i,1}^{(1)}, c_{i,1}^{(2)}, \dots, c_{i,1}^{(H_1)} & 0 & L & 0 \\ 0 & c_{i,1}^{(1)}, c_{i,1}^{(2)}, \dots, c_{i,1}^{(H_2)} & L & 0 \\ M & M & O & M \\ 0 & 0 & L & c_{i,1}^{(1)}, c_{i,1}^{(2)}, \dots, c_{i,1}^{(H_m)} \end{bmatrix} \quad (1)$$

The average grouping methodology was then used to obtain the coefficients $\beta_m(\epsilon)$ for

211 each class in a given test subset (Pham et al., 2019):

$$\beta_m(\varepsilon) = \frac{1}{L} \sum_{i=1}^L \alpha_{i,m}(\varepsilon C_i^\alpha) \quad k = 1, 2, 3, \dots, h \quad (2)$$

212 where $\alpha_{i,m}(\varepsilon C_i^\alpha)$ is the chance of the classifier B_i to the hypothesis that ε belongs to class m .

213 Finally, ε is appointed to the class with the utmost confidence.

214 3.2 J48

215 The decision tree classifier displays the sample structure based on a tree-like hierarchical
216 structure, and creates a classification tree, where non-terminal annotations indicate attribute
217 testing, and terminal annotations indicate decision results (Sahu and Mehtre, 2015). Because
218 of its stable performance, it has been used in landslide susceptibility prediction many times
219 (Pham et al., 2017; Tien Bui et al., 2014a). J48 decision tree (J48) is an optimization model
220 based on the C4.5 algorithm as a decision tree classifier and is often used to deal with
221 classification problems in machine learning and data mining (Hayashi and Nakano, 2015;
222 Kang and Michalak, 2018; Tien Bui et al., 2014b). The C4.5 algorithm was first proposed by
223 J. Ross Quinlan based on ID3. Its main purpose is to explore the mapping relationship from
224 attribute values to categories. The mapping relationship can be used to classify the results that
225 need to be classified (Moral-García et al., 2019; Sathyadevan and Nair, 2015). Also, another
226 well-known decision tree classifier method, namely the classification and regression tree
227 (CART), was also derived from the ID3 algorithm (Youssef et al., 2016).

228 In this research, the J48 model was used as one of the basic models for landslide
229 susceptibility prediction. Its principle is to sort unknown samples from root node to leaf node
230 by programming (Anzola et al., 2015). The main operation steps are: find the input data of the

conditioning factors through the training dataset with the highest gain quantitative relation and build the tree structure; the primary internal node is the root note, and then generate sub-notes by combining the sub-training data generated from the training dataset; find the gain ratio of the sub-notes to select the landslide conditioning factors with high gain ratio. Repeat till all instances within the training dataset are appointed the gain quantitative relation of the leaf nodes, or the remaining variables cannot be split (Hong et al., 2018).

3.3 Alternating decision tree (ADT)

The ADT model is a classifier, which is composed of direct representation of a decision tree combined with a boosting algorithm (Freund and Mason, 2002). However, the nodes of the ADT model are composed of prediction nodes and split note, which is distinguished from decision nodes (Kuang and Po-Leen Ooi, 2010; Sok et al., 2015). A precondition z_p , a basic condition z_b and two real numbers x and y form the basic ruler mapping from instances to real numbers (Chen et al., 2019a; Pham et al., 2016). $z_p \cap z_b$ represents the prediction as x , and conversely, the predicted value y represents $z_p \cap ^- z_b$. $F(z)$ is the total weights of the training instances. The values of x and y are calculated by the following equation:

$$x = \frac{1}{2} \ln \frac{F_+(z_p \cap z_b)}{F_-(z_p \cap z_b)}, \quad y = \frac{1}{2} \ln \frac{F_+(z_p \cap ^- z_b)}{F_-(z_p \cap ^- z_b)} \quad (3)$$

The best z_p and z_b are achieved by minimizing the $F_t(z_p, z_b)$, as shown in the following:

$$F_t(z_p, z_b) = 2\sqrt{F_+(z_p \cap z_b)F_-(z_p \cap z_b)} + \sqrt{F_+(z_p \cap ^- z_b)F_-(z_p \cap ^- z_b)} + F(^- z_b) \quad (4)$$

a represents a set of instances. $w_t(a)$ denotes the two predicted values (x and y) of each layer of the tree structure. W represents a set base rule. So, a new rule can be: $W_{t+1} = W_t + w_t$.

The classification of instances is the sign of the sum of total predicted values in W_{t+1} :

$$\text{Class}(a) = \text{sign}\left(\sum_{t=1}^T w_t(a)\right) \quad (5)$$

This method usually uses cross-validation to find the best constant prediction value of the whole dataset (Pfahring et al., 2001).

3.4 Random forest (RaF)

The RaF model is an integrated machine learning technique proposed by Breiman, used for regression, classification and different works (Breiman, 2001). It is composed of multiple decision trees (Breiman, 2001; Deng and Guo, 2011). The basic unit of the RAF model is the decision tree. Its essence is a part of machine ensemble learning, and it is an algorithm that integrates holistic learning with multiple trees (Bernard et al., 2008; Mantas et al., 2018; Vens, 2013). Because the model has the advantages of generating high accuracy classifiers, not easy to over fitting, excellent anti-noise ability, and fast training speed, the RaF algorithm is widely utilized in difficult prediction work, especially in nonlinear high-dimensional landslide spatial evaluation (Dang et al., 2018).

In this research, the process of model processing data can be briefly summarized as the following: (i) selecting samples from the training dataset to generate a new training sample set, (ii) using the landslide conditioning factors to establish a decision tree; (iii) repeating the above steps to generate decision trees and form the RaF; (iv) using the remaining data of training dataset and validation dataset to verify the model; (v) The model was applied to the whole study space and classified into landslide susceptibility classes.

4. Results

4.1 Predictive ability of landslide conditioning factors

In recent years, some researches have shown that the predictive ability of landslide conditioning factors in different research areas is different, which plays a positive or negative affect within the prevalence of landslides (Lee et al., 2018; Wu et al., 2019). In other words, they may not have any positive effect on the occurrence of landslides, that is, to inhibit the occurrence of landslides. Due to the differences in hydrogeology, engineering geology, and human activities between different regions, a lack of unified, reasonable and effective criteria for landslide conditioning factors has not been formed. The reasonability of the selected conditioning factors mainly dependent on whether the researchers have an accurate understanding of the geological environment conditions in the landslide area. In this study, the average merit (AM) and standard deviation (St.d) of correlation attribute evaluation (CAE) model are used as the evaluation criteria to evaluate the prediction ability of conditioning factors on landslide occurrence and the values are shown in Table 2. If the conditioning factors promote landslide susceptibility, the AM value is greater than zero. The CAE model verifies that the selected conditioning factors in this study all play a positive role in promoting landslides. Among them, lithology, elevation and TWI have the most obvious promoting effect; the promoting effect of slope aspect exists but is not obvious. In conclusion, the conditioning factors selected in this study have relatively high rationality.

Table 2 Predictive ability of conditioning factors based on the CAE model.

Factors	average merit (AM)	standard deviation (St.d)
Elevation	0.231	±0.012
Lithology	0.224	±0.016
TWI	0.214	±0.013
Distance to roads	0.205	±0.016
Slope angle	0.174	±0.015
Landuse	0.153	±0.018
Distance to rivers	0.092	±0.009
NDVI	0.059	±0.015

Plan curvature	0.057	±0.014
Profile curvature	0.048	±0.017
Soil	0.020	±0.011
Slope aspect	0.009	±0.006

4.2 Generation of landslide susceptibility maps

The last and most critical step of the research is to generate the landslide susceptibility maps. It is carried out under the combined action of selected and analyzed 12 landslide conditioning factors and six trained models and directly graphically shows the spatial distribution characteristics of landslide susceptibility differences in the research area. Generally speaking, the generation of the landslide susceptibility map is divided into three steps. Firstly, the landslide susceptibility index (LSI) of each evaluation unit was calculated by the six established models. Then, all the LSIs were introduced into the drawn landslide inventory map to become property on the map. Thirdly, using the appropriate method to classify the landslide susceptibility grades. The Jenks natural fracture method was applied to classify the landslide susceptibility grades. This is mainly because the classification results of the Jenks natural breaks method are characterized by small differences within groups and large differences between groups. Furthermore, it can effectively analyze the characteristics of the data to make the results objective and reliable. Finally, the landslide susceptibility grades of this research were partition into five grades, namely very low, low, moderate, high and very high. It can be seen that the six maps in Fig. 3 are the final maps created by the J48, ADTree, RaF, RF-J48, RF-ADTree and RF-RaF models, respectively. The area proportion of the six models in the five landslide susceptibility grades and the percentage of the landslides number in each grade are shown in Fig. 4. Each detail data is also shown on the offside of the bar graph in Fig. 4. Different models reflect the dominant distribution of very low susceptibility

308 areas, and most landslides occur in very high susceptibility grade.

Fig. 3 Landslide susceptibility maps: (a) J48 model, (b) ADTree model, (c) RaF model, (d) RF-J48 model, (e) RF-ADTree model, and (f) RF-RaF model.

Fig. 4 Percentages of landslide susceptibility grades.

4.3 Models performances and analysis

The scientific nature, accuracy and later applicability of the research results largely depend on

whether the research model is selected appropriately. Hence, the model performance and analysis are an indispensable section of landslide susceptibility evaluation research. Without this section, the rigor of the research process will be weakened, and the results will not have any practical scientific significances (Amato et al., 2019; Merghadi et al., 2020). Based on the above reasons, the relative operating characteristic (ROC) curve and its parameters, such as AUC value, standard error, 95% confidence interval (CI) and P value, were adopted to analyze the performance of the six models used in this research. The AUC value of the ROC curve generated from the training dataset is called the success rate, while the AUC value of the ROC curve generated from the validation datasets is called the prediction rate. As can be seen from Fig. 5 and Fig. 6, there are six ROC curves in each figure, which respectively represent the six susceptibility research models used in this study. RF-RaF and RaF models performed best in the training dataset with the highest success rate of 0.75, followed by the RF-ADTree model (0.74), RF-J48 model (0.74), ADTree model (0.71) and J48 model (0.70). At the same time, model performance in the validation dataset expressed similar results, namely RF-RaF model achieves the best performance with the prediction rate of 0.82, RaF model (0.79) achieved the best performance. And the RF-ADTree, RF-J48, ADTree and J48 models indicate that the AUC values of the prediction rates are 0.80, 0.77, 0.76 and 0.71, respectively. Additionally, the results of other parameters of ROC curves in Table 3 and Table 4 are the same as AUC values, and each model has extremely significant statistical differences. For example, in the training and validation datasets, the RF-RaF model has the smallest standard error, 95% CI and P value, which indicate the RF-RaF model performed outstandingly in this study.

Fig. 5 ROC curves of the six models using training dataset.

Fig. 6 ROC curves of the six models using validation dataset.

Table 3 Parameters of ROC curves with training dataset.

Test Variables	RF-J48	RF-ADTree	RF-RaF	J48	ADTree	RaF
AUC	0.730	0.739	0.753	0.699	0.707	0.752
Standard Error	0.026	0.025	0.024	0.027	0.026	0.025
95% CI	0.680-0.780	0.689-0.788	0.705-0.800	0.647-0.751	0.656-0.759	0.704-0.800
P Value	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001

Table 4 Parameters of ROC curves with validation dataset.

Test Variables	RF-J48	RF-ADTree	RF-RaF	J48	ADTree	RaF
----------------	--------	-----------	--------	-----	--------	-----

AUC	0.770	0.799	0.819	0.707	0.755	0.792
Standard Error	0.036	0.034	0.032	0.041	0.037	0.034
95% CI	0.699-0.841	0.732-0.865	0.756-0.882	0.627-0.786	0.683-0.828	0.725-0.859
P Value	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001

5. Discussion

In the present study, J48, ADTree, RaF, RF-J48, RF-ADTree and RF-RaF were implemented to landslide spatial prognosis research in Muchuan County, China. Firstly, 279 landslides were collected through field investigation and satellite images interpretation, and a landslide inventory map was compiled. Then, twelve conditioning factors, for example elevation, lithology, TWI, distance to roads, slope angle and so on, were selected based on local geo-environmental characteristics and related landslide research literature. Also, the selected conditioning factors distinguish the predictive ability of landslide occurrence through the CAE model, and elevation, lithology and TWI were identified to be the three most significant factors. Finally, the above-mentioned six algorithms establish their landslide susceptibility models based on twelve conditioning factors, and the landslide spatial prediction in Muchuan County is studied. As everyone knows, elevation, lithology and TWI play a vital role in the occurrence of landslides (Zhou et al., 2018). Multitudinous researchers have been searching for the relationship between elevation and landslides. Sheng and Chen proposed a landslide detection method based on elevation: to enhance the distinguishability of the geological environment characteristics of the research area, adopt the slow feature analysis (SFA) method, and then apply intelligent algorithms to model and compare with the actual landslide distribution (Sheng and Chen, 2017). Lithology is critical to the occurrence of landslides, and it is the basis of the material composition of slopes. The shear strength varies drastically in different lithologies, which makes the probability of landslide occurrence differently. Many

experts and scholars emphasize the key role of lithology change in landslide (Jaafari et al., 2014). Wu and Qiao found that in addition to the relative height difference, slope and curvature, lithology was also an essential conditioning factor. According to the field investigation in the Three Gorges Reservoir area, they found lithologies of J_{3s}, J_{3p} and T_{2b} contribute the most (Wu and Qiao, 2009). TWI is commonly accustomed to describe the spatial distribution characteristics of soil moisture. Generally speaking, with the increase of soil water content, the internal friction angle and cohesion of soil will decrease significantly, which will affect the occurrence of the landslide. By simulating the response process of the slope to precipitation and introducing the soil moisture provided by field sensors, Marino et al. conducted an experimental study on the impact of soil moisture on shallow landslides (Marino et al., 2020).

The landslide susceptibility maps generated by six different models and related data statistics show that the landslide susceptibility maps classified by Jenks natural breaks method are reasonable, which effectively and intuitively show the objective landslide susceptibility zoning results (Fig. 3). From the area proportion and the relative proportion of landslide number in each grade of relevant data statistics (Fig. 4), it can be seen that the area and landslide proportion of all model zones follow the Jenks natural fracture rules and are reasonable distribution. The RF-RaF model has the best partition result, and the landslide number distribution in each grade is the most reasonable. The reason is that the RF algorithm is improved base on the RaF algorithm, which has the advantages of good code readability and high precision. The RF model has been used in scientific research many times in recent years. For example, Naghibi et al. Considered that the groundwater susceptibility map

generated by the RF model in meshgin Shahr of Iran has higher accuracy than traditional models, such as enhanced regression tree, classified regression tree and evidence belief function (Naghibi et al., 2019). Artetxe et al. conducted a susceptibility study on patient readmission based on the RF model. The prediction accuracy of this model is more than 70%, which indicates that the RF model can be used for on-site analysis of medical staff (Artetxe et al., 2017). Although the RF model is a relatively advanced algorithm, in order to further explore the advantages of the algorithm and meet the needs of scientific research, it is also an extensional application of the model to build integrated models by coupling with other algorithms. Therefore, in this paper, four other algorithms widely used in landslide susceptibility research were introduced based on the RF model. This study was conducted by comparing the performance of three benchmark models, and three RF integrated hybrid models (J48, ADTree, RaF, RF-J48, RF-ADTree and RF-RaF models).

The performance of the above six models during this research was demonstrated and compared by ROC curves, especially by the parameter AUC value of each model. As shown in Table 3 and Table 4, the RaF or RF-based model performs eminently in the training and the validation datasets. Although the success rate of the RaF model in the training dataset is 0.75, which is equal to RF-RaF model that consists of RaF and RF models, it cannot be denied that the RF-RaF model has an excellent performance. In case of the validation dataset, the advantage of RF-RaF model becomes more obvious, and its AUC value (0.82) is significantly higher than other models. Other parameters in the training and validation datasets, such as standard error and 95% CI, show the same results too. And the integrated model is more predictive than the single model. Therefore, it can be considered that the landslide

susceptibility maps of all the above models can be used as the data of natural resource management and land use planning in Muchuan County, and the map drawn by RF-RaF model is the most convincing. At the same time, it also provides a reference for different regions under similar geological environment.

6. Conclusions

Landslide spatial prediction research is the most basic and critical part in landslide disaster prevention, which determines the probability difference of landslide spatial distribution, that is, landslide susceptibility grade. The purpose of this research is to establish landslide spatial prediction models in Muchuan County by J48, ADTree, RaF, RF-J48, RF-ADTree and RF-RaF models. Then, according to the local geo-environmental conditions and previous research results, 12 landslide conditioning factors are selected, and the prediction ability of each conditioning factor is analyzed and evaluated by using CAE model. Next, six landslide susceptibility maps were obtained by the six models, and the grades of susceptibility were separated into five grades by the Jenks natural breaks method, i.e., very low, low, moderate, high and very high. Last, the ROC curves were introduced to predict the spatial evaluation capabilities of those models. It was indicated that RF-RaF, RaF and RF-ADTree model are the three well performed models in both training and validation datasets, and the AUC values of RF-RaF model in the training and validation datasets are 0.75 and 0.82 respectively. In addition, this study also proves that ensemble model has more powerful spatial prediction capability than a single model. Finally, the RF-RaF model can be considered as the most suitable model for landslide spatial assessment within the study area, and the results can provide scientific references of landslide prevention and management for native natural

resources departments.

Acknowledgements This study has been supported by the National Natural Science Foundation of China (Grant No. 41807285).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

- Althuwaynee OF, Pradhan B, Lee S (2012) Application of an evidential belief function model in landslide susceptibility mapping. *Computers & Geosciences* 44:120-135. doi:<https://doi.org/10.1016/j.cageo.2012.03.003>
- Amato G, Eisank C, Castro-Camilo D, Lombardo L (2019) Accounting for covariate distributions in slope-unit-based landslide susceptibility models. A case study in the alpine environment. *Engineering Geology* 260:105237. doi:<https://doi.org/10.1016/j.enggeo.2019.105237>
- Anzola J, Cuartas K, Tarazona G (2015) Classification Methodology Of Research Topics Based In Decision Trees: J48 And Randomtree. *International Journal of Applied Engineering Research* 10:19413.
- Artetxe A, Ayerdi B, Graña M, Rios S (2017) Using Anticipative Hybrid Extreme Rotation Forest to predict emergency service readmission risk. *Journal of Computational Science* 20:154-161. doi:<https://doi.org/10.1016/j.jocs.2016.12.008>
- Ayalew L, Yamagishi H (2005) The application of GIS-based logistic regression for landslide susceptibility mapping in the Kakuda-Yahiko Mountains, Central Japan. *Geomorphology* 65:15-31.
- Basher L, Betts H, Lynn I, Marden M, McNeill S, Page M, Rosser B (2018) A preliminary assessment of the impact of landslide, earthflow, and gully erosion on soil carbon stocks in New Zealand. *Geomorphology* 307:93-106. doi:<https://doi.org/10.1016/j.geomorph.2017.10.006>
- Bernard S, Heutte L, Adam S (2008) Forest-RK: A New Random Forest Induction Method. In: Huang D-S, Wunsch DC, Levine DS, Jo K-H (eds), *Advanced Intelligent Computing Theories and Applications With Aspects of Artificial Intelligence*. Springer Berlin Heidelberg, Berlin, Heidelberg. pp 430-437
- Braun A, Garcia Urquia EL, Moncada Lopez R, Yamagishi H (2019) Landslide Susceptibility Mapping in Tegucigalpa, Honduras, Using Data Mining Methods. In: Shakoor A, Cato K (eds), *IAEG/AEG Annual Meeting Proceedings, San Francisco, California, 2018 - Volume 1*. Springer International Publishing, Cham. pp 207-215

467 Breiman L (2001) Random Forests. Machine Learning 45:5-32.
 468 doi:10.1023/A:1010933404324

469 Bui D, Pradhan B, Löfman O, Revhaug I, Dick Ø (2011) Landslide susceptibility mapping at
 470 Hoa Binh province (Vietnam) using an adaptive neuro-fuzzy inference system and GIS.
 471 Computers & Geosciences 45. doi:10.1016/j.cageo.2011.10.031

472 Canoglu MC, Aksoy H, Ercanoglu M (2018) Integrated approach for determining
 473 spatio-temporal variations in the hydrodynamic factors as a contributing parameter in
 474 landslide susceptibility assessments. Bulletin of Engineering Geology and the
 475 Environment. doi:10.1007/s10064-018-1337-z

476 Chen W, Li W, Hou E, Bai H, Chai H, Wang D, Cui X, Wang Q (2015) Application of
 477 frequency ratio, statistical index, and index of entropy models and their comparison in
 478 landslide susceptibility mapping for the Baozhong Region of Baoji, China. Arabian
 479 Journal of Geosciences 8:1829-1841. doi:10.1007/s12517-014-1554-0

480 Chen W, Pourghasemi HR, Kornejady A, Xie X (2019a) GIS-Based Landslide Susceptibility
 481 Evaluation Using Certainty Factor and Index of Entropy Ensembled with Alternating
 482 Decision Tree Models. In: Pourghasemi HR, Rossi M (eds). Natural Hazards GIS-Based
 483 Spatial Modeling Using Data Mining Techniques. Springer International Publishing,
 484 Cham. pp 225-251

485 Chen W, Tsangaratos P, Ilia I, Duan Z, Chen X (2019b) Groundwater spring potential
 486 mapping using population-based evolutionary algorithms and data mining methods.
 487 Science of The Total Environment. doi:<https://doi.org/10.1016/j.scitotenv.2019.05.312>

488 Chu L, Wang L-J, Jiang J, Liu X, Sawada K, Zhang J (2018) Comparison of landslide
 489 susceptibility maps using random forest and multivariate adaptive regression spline
 490 models in combination with catchment map units. Geosciences Journal.
 491 doi:10.1007/s12303-018-0038-8

492 Cordeira JM, Stock J, Dettinger MD, Young AM, Kalansky JF, Ralph FM (2019) A 142-year
 493 Climatology of Northern California Landslides and Atmospheric Rivers. Bulletin of the
 494 American Meteorological Society. doi:10.1175/BAMS-D-18-0158.1

495 Dahal RK, Hasegawa S, Nonomura A, Yamanaka M, Masuda T, Nishino K (2008) GIS-based
 496 weights-of-evidence modelling of rainfall-induced landslides in small catchments for
 497 landslide susceptibility mapping. Environmental Geology 54:311-324.
 498 doi:10.1007/s00254-007-0818-3

499 Dang V-H, Dieu TB, Tran X-L, Hoang N-D (2018) Enhancing the accuracy of
 500 rainfall-induced landslide prediction along mountain roads with a GIS-based random
 501 forest classifier. Bulletin of Engineering Geology and the Environment.
 502 doi:10.1007/s10064-018-1273-y

503 Deng C, Guo MZ (2011) A new co-training-style random forest for computer aided diagnosis.
 504 Journal of Intelligent Information Systems 36:253-281. doi:10.1007/s10844-009-0105-8

505 Devkota KC, Regmi AD, Pourghasemi HR, Yoshida K, Pradhan B, Ryu IC, Dhital MR,
 506 Althuwaynee OF (2013) Landslide susceptibility mapping using certainty factor, index
 507 of entropy and logistic regression models in GIS and their comparison at Mugling–
 508 Narayanghat road section in Nepal Himalaya. Natural Hazards 65:135-165.
 509 doi:10.1007/s11069-012-0347-6

510 Diva IH, Irwanto U, Nizam K, Annur L, Sekarjati D, Putra BG, Safitri Y, Giovandi EA,

- Nofrizal AY, Hanif M, Abe A (2018) Investigation Volcanic Land Form and Mapping Landslide Potential at Mount Talang. Sumatra Journal of Disaster, Geography and Geography Education; Vol 2 No 1 (2018): Sumatra Journal of Disaster, Geography and Geography Education Volume 2 Number 1 - June 2018 : DisasterDO - 1024036/sjdggev2i1130.
- Du P, Samat A, Waske B, Liu S, Li Z (2015) Random Forest and Rotation Forest for fully polarized SAR image classification using polarimetric and spatial features. ISPRS Journal of Photogrammetry and Remote Sensing 105:38-53. doi:<https://doi.org/10.1016/j.isprsjprs.2015.03.002>
- Feng Z, Zhang N, Yan H, Dai Z (2019) Failure Mechanisms of the Lotus Pond Landslide: A Reactivated Landslide from Large-Scale Cataclinal Slope Failure in the Three Gorges Reservoir Area of China. In: Shakoor A, Cato K (eds), IAEG/AEG Annual Meeting Proceedings, San Francisco, California, 2018 - Volume 1. Springer International Publishing, Cham. pp 19-23
- Fiorucci F, Ardizzone F, Mondini AC, Viero A, Guzzetti F (2019) Visual interpretation of stereoscopic NDVI satellite images to map rainfall-induced landslides. Landslides 16:165-174. doi:10.1007/s10346-018-1069-y
- Freund Y, Mason L (2002) *The Alternating Decision Tree Learning Algorithm*
- Friedman JH (1991) Multivariate Adaptive Regression Splines. Annals of Statistics 19:1-67.
- Ge Y, Chen H, Zhao B, Tang H, Lin Z, Xie Z, Lv L, Zhong P (2018) Correction to: A comparison of five methods in landslide susceptibility assessment: a case study from the 330-kV transmission line in Gansu Region, China. Environmental Earth Sciences 77:716. doi:10.1007/s12665-018-7900-x
- Hayashi Y, Nakano S (2015) Use of a Recursive-Rule eXtraction algorithm with J48graft to achieve highly accurate and concise rule extraction from a large breast cancer dataset. Informatics in Medicine Unlocked 1:9-16. doi:<https://doi.org/10.1016/j.imu.2015.12.002>
- 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. doi:<https://doi.org/10.1016/j.catena.2018.01.005>
- Hong H, Liu J, Zhu AX (2019) Landslide susceptibility evaluating using artificial intelligence method in the Youfang district (China). Environmental Earth Sciences 78:488. doi:10.1007/s12665-019-8415-9
- Huang R (2012) Mechanisms of large-scale landslides in China. Bulletin of Engineering Geology and the Environment 71:161-170. doi:10.1007/s10064-011-0403-6
- 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. International Journal of Environmental Science and Technology 11:909-926. doi:10.1007/s13762-013-0464-0
- Kang K, Michalak J (2018) *Enhanced version of AdaBoostM1 with J48 Tree learning method*
- Kornejady A, Pourghasemi HR, Afzali SF (2019) Presentation of RFFR New Ensemble Model for Landslide Susceptibility Assessment in Iran. In: Pradhan SP, Vishal V, Singh TN (eds). Landslides: Theory, Practice and Modelling. Springer International Publishing, Cham. pp 123-143

555 Kuang Y, Po-Leen Ooi M (2010) *Complex feature alternating decision tree*
 556 Kulatilake PHSW, Ge Y (2014) Investigation of Stability of the Critical Rock Blocks that
 557 Initiated the Jiweishan Landslide in China. *Geotechnical and Geological Engineering*
 558 32:1291-1315. doi:10.1007/s10706-014-9806-z
 559 Lasota T, Telec Z, Trawiński B, Trawiński G (2012) Investigation of Rotation Forest
 560 Ensemble Method Using Genetic Fuzzy Systems for a Regression Problem. In: Pan J-S,
 561 Chen S-M, Nguyen NT (eds), *Intelligent Information and Database Systems*. Springer
 562 Berlin Heidelberg, Berlin, Heidelberg. pp 393-402
 563 Lee S, Lee M-J, Lee S (2018) Spatial prediction of urban landslide susceptibility based on
 564 topographic factors using boosted trees. *Environmental Earth Sciences* 77:656.
 565 doi:10.1007/s12665-018-7778-7
 566 Li R, Wang N (2019) Landslide Susceptibility Mapping for the Muchuan County (China): A
 567 Comparison Between Bivariate Statistical Models (WoE, EBF, and IoE) and Their
 568 Ensembles with Logistic Regression. *Symmetry* 11:762-785. doi:10.3390/sym11060762
 569 Losasso L, Rinaldi C, Alberico D, Sdao F (2017) Landslide Risk Analysis Along Strategic
 570 Touristic Roads in Basilicata (Southern Italy) Using the Modified RHRS 2.0 Method. In:
 571 Gervasi O, Murgante B, Misra S, Borruso G, Torre CM, Rocha AMAC, Taniar D,
 572 Apduhan BO, Stankova E, Cuzzocrea A (eds), *Computational Science and Its*
 573 *Applications – ICCSA 2017*. Springer International Publishing, Cham. pp 761-776
 574 Ma J, Tang H, Liu X, Hu X, Sun M, Song Y (2017) Establishment of a deformation
 575 forecasting model for a step-like landslide based on decision tree C5.0 and two-step
 576 cluster algorithms: a case study in the Three Gorges Reservoir area, China. *Landslides*
 577 14:1275-1281. doi:10.1007/s10346-017-0804-0
 578 Mandal S, Mondal S (2019) Artificial Neural Network (ANN) Model and Landslide
 579 Susceptibility. In: Mandal S, Mondal S (eds). *Statistical Approaches for Landslide*
 580 *Susceptibility Assessment and Prediction*. Springer International Publishing, Cham. pp
 581 123-133
 582 Mantas CJ, Castellano JG, Moral-García S, Abellán J (2018) A comparison of random forest
 583 based algorithms: random credal random forest versus oblique random forest. *Soft*
 584 *Computing*. doi:10.1007/s00500-018-3628-5
 585 Marino P, Peres DJ, Cancelliere A, Greco R, Bogaard TA (2020) Soil moisture information
 586 can improve shallow landslide forecasting using the hydrometeorological threshold
 587 approach. *Landslides*. doi:10.1007/s10346-020-01420-8
 588 Merghadi A, Yunus AP, Dou J, Whiteley J, ThaiPham B, Bui DT, Avtar R, Abderrahmane B
 589 (2020) Machine learning methods for landslide susceptibility studies: A comparative
 590 overview of algorithm performance. *Earth-Science Reviews* 207:103225.
 591 doi:<https://doi.org/10.1016/j.earscirev.2020.103225>
 592 Mohammady M, Pourghasemi HR, Pradhan B (2012) Landslide susceptibility mapping at
 593 Golestan Province, Iran: A comparison between frequency ratio, Dempster–Shafer, and
 594 weights-of-evidence models. *Journal of Asian Earth Sciences* 61:221-236.
 595 doi:<https://doi.org/10.1016/j.jseaes.2012.10.005>
 596 Mondal S, Mandal S (2018) Landslide susceptibility mapping of Darjeeling Himalaya, India
 597 using index of entropy (IOE) model. *Applied Geomatics*.
 598 doi:10.1007/s12518-018-0248-9

- Moral-García S, Mantas CJ, Castellano JG, Abellán J (2019) Ensemble of classifier chains and Credal C4.5 for solving multi-label classification. *Progress in Artificial Intelligence*. doi:10.1007/s13748-018-00171-x
- Myronidis D, Papageorgiou C, Theophanous S (2016) Landslide susceptibility mapping based on landslide history and analytic hierarchy process (AHP). *Natural Hazards* 81:245-263. doi:10.1007/s11069-015-2075-1
- Naghibi SA, Dolatkordestani M, Rezaei A, Amouzegari P, Heravi MT, Kalantar B, Pradhan B (2019) Application of rotation forest with decision trees as base classifier and a novel ensemble model in spatial modeling of groundwater potential. *Environmental Monitoring and Assessment* 191:248. doi:10.1007/s10661-019-7362-y
- Pardo C, Diez-Pastor JF, García-Osorio C, Rodríguez JJ (2013) Rotation Forests for regression. *Applied Mathematics and Computation* 219:9914-9924. doi:<https://doi.org/10.1016/j.amc.2013.03.139>
- Park I, Lee S (2014) Spatial prediction of landslide susceptibility using a decision tree approach: a case study of the Pyeongchang area, Korea. *International Journal of Remote Sensing* 35:6089-6112. doi:10.1080/01431161.2014.943326
- Park S, Choi C, Kim B, Kim J (2013) Landslide susceptibility mapping using frequency ratio, analytic hierarchy process, logistic regression, and artificial neural network methods at the Inje area, Korea. *Environmental Earth Sciences* 68:1443-1464. doi:10.1007/s12665-012-1842-5
- Peng J, Wang G, Wang Q, Zhang F (2017) Shear wave velocity imaging of landslide debris deposited on an erodible bed and possible movement mechanism for a loess landslide in Jingyang, Xi'an, China. *Landslides* 14:1503-1512. doi:10.1007/s10346-017-0827-6
- Persichillo MG, Bordoni M, Meisina C (2017) The role of land use changes in the distribution of shallow landslides. *Science of The Total Environment* 574:924-937. doi:<https://doi.org/10.1016/j.scitotenv.2016.09.125>
- Pfahring B, Holmes G, Kirkby R (2001) Optimizing the Induction of Alternating Decision Trees. In: Cheung D, Williams GJ, Li Q (eds), *Advances in Knowledge Discovery and Data Mining*. Springer Berlin Heidelberg, Berlin, Heidelberg. pp 477-487
- Pham BT, Prakash I, Singh SK, Shirzadi A, Shahabi H, Bui DT (2019) Landslide susceptibility modeling using Reduced Error Pruning Trees and different ensemble techniques: Hybrid machine learning approaches. *Catena* 175:203-218.
- Pham BT, Tien Bui D, Dholakia MB, Prakash I, Pham HV (2016) A Comparative Study of Least Square Support Vector Machines and Multiclass Alternating Decision Trees for Spatial Prediction of Rainfall-Induced Landslides in a Tropical Cyclones Area. *Geotechnical and Geological Engineering* 34:1807-1824. doi:10.1007/s10706-016-9990-0
- Pham BT, Tien Bui D, Prakash I (2017) Landslide Susceptibility Assessment Using Bagging Ensemble Based Alternating Decision Trees, Logistic Regression and J48 Decision Trees Methods: A Comparative Study. *Geotechnical and Geological Engineering* 35:2597-2611. doi:10.1007/s10706-017-0264-2
- Pham BT, Tien Bui D, Prakash I (2018) Bagging based Support Vector Machines for spatial prediction of landslides. *Environmental Earth Sciences* 77:146. doi:10.1007/s12665-018-7268-y

- Polykretis C, Chalkias C, Ferentinou M (2017) Adaptive neuro-fuzzy inference system (ANFIS) modeling for landslide susceptibility assessment in a Mediterranean hilly area. *Bulletin of Engineering Geology and the Environment*. doi:10.1007/s10064-017-1125-1
- Polykretis C, Ferentinou M, Chalkias C (2015) A comparative study of landslide susceptibility mapping using landslide susceptibility index and artificial neural networks in the Krios River and Krathis River catchments (northern Peloponnesus, Greece). *Bulletin of Engineering Geology and the Environment* 74:27-45. doi:10.1007/s10064-014-0607-7
- Pourghasemi HR, Kerle N (2016) Random forests and evidential belief function-based landslide susceptibility assessment in Western Mazandaran Province, Iran. *Environmental Earth Sciences* 75:185. doi:10.1007/s12665-015-4950-1
- Pourghasemi HR, Pradhan B, Gokceoglu C (2012) Application of fuzzy logic and analytical hierarchy process (AHP) to landslide susceptibility mapping at Haraz watershed, Iran. *Natural Hazards* 63:965-996. doi:10.1007/s11069-012-0217-2
- Pradhan B (2010) Landslide susceptibility mapping of a catchment area using frequency ratio, fuzzy logic and multivariate logistic regression approaches. *Journal of the Indian Society of Remote Sensing* 38:301-320. doi:10.1007/s12524-010-0020-z
- Pradhan B (2013) A comparative study on the predictive ability of the decision tree, support vector machine and neuro-fuzzy models in landslide susceptibility mapping using GIS. *Computers & Geosciences* 51:350-365. doi:<https://doi.org/10.1016/j.cageo.2012.08.023>
- Pradhan B, Lee S (2010) Landslide susceptibility assessment and factor effect analysis: backpropagation artificial neural networks and their comparison with frequency ratio and bivariate logistic regression modelling. *Environmental Modelling & Software* 25:747-759. doi:<https://doi.org/10.1016/j.envsoft.2009.10.016>
- Raja NB, Çiçek I, Türkoğlu N, Aydın O, Kawasaki A (2017) Landslide susceptibility mapping of the Sera River Basin using logistic regression model. *Natural Hazards* 85:1323-1346. doi:10.1007/s11069-016-2591-7
- Ramakrishnan D, Singh TN, Verma AK, Gulati A, Tiwari KC (2013) Soft computing and GIS for landslide susceptibility assessment in Tawaghat area, Kumaon Himalaya, India. *Natural Hazards* 65:315-330. doi:10.1007/s11069-012-0365-4
- Razavizadeh S, Solaimani K, Massironi M, Kavian A (2017) Mapping landslide susceptibility with frequency ratio, statistical index, and weights of evidence models: a case study in northern Iran. *Environmental Earth Sciences* 76:499. doi:10.1007/s12665-017-6839-7
- Ren F, Wu X, Zhang K, Niu R (2015) Application of wavelet analysis and a particle swarm-optimized support vector machine to predict the displacement of the Shuping landslide in the Three Gorges, China. *Environmental Earth Sciences* 73:4791-4804. doi:10.1007/s12665-014-3764-x
- Rodriguez JJ, Kuncheva LI, Alonso CJ (2006) Rotation Forest: A New Classifier Ensemble Method. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 28:1619-1630. doi:10.1109/TPAMI.2006.211
- Sahana M, Sajjad H (2017) Evaluating effectiveness of frequency ratio, fuzzy logic and logistic regression models in assessing landslide susceptibility: a case from Rudraprayag district, India. *Journal of Mountain Science* 14:2150-2167. doi:10.1007/s11629-017-4404-1

- Sahu S, Mehtre BM (2015) Network intrusion detection system using J48 Decision Tree. 2015 International Conference on Advances in Computing, Communications and Informatics (ICACCI). pp 2023-2026
- Sathyadevan S, Nair RR (2015) Comparative Analysis of Decision Tree Algorithms: ID3, C4.5 and Random Forest. In: Jain LC, Behera HS, Mandal JK, Mohapatra DP (eds), Computational Intelligence in Data Mining - Volume 1. Springer India, New Delhi. pp 549-562
- Sheng T, Chen Q (2017) An Altitude Based Landslide and Debris Flow Detection Method for a Single Mountain Remote Sensing Image. In: Zhao Y, Kong X, Taubman D (eds), Image and Graphics. Springer International Publishing, Cham. pp 601-610
- Sok HK, Ooi MP-L, Kuang YC (2015) Sparse alternating decision tree. Pattern Recognition Letters 60-61:57-64. doi:<https://doi.org/10.1016/j.patrec.2015.03.002>
- Sridhar B, Rao PJ, Narasimha Rao G, Duvvuru R, Anusha C, Sanyasi Naidu D, Srinivas E, Sridevi T, Madhuri M, Padmini Y (2019) Identification of Landslide Hazard Zones Along the Bheemili Beach Road, Visakhapatnam District, A.P. In: Rao PJ, Rao KN, Kubo S (eds), Proceedings of International Conference on Remote Sensing for Disaster Management. Springer International Publishing, Cham. pp 515-522
- Sun W, Tian Y, Mu X, Zhai J, Gao P, Zhao G (2017) Loess Landslide Inventory Map Based on GF-1 Satellite Imagery. Remote Sensing 9. doi:10.3390/rs9040314
- Tehrany MS, Kumar L (2018) The application of a Dempster-Shafer-based evidential belief function in flood susceptibility mapping and comparison with frequency ratio and logistic regression methods. Environmental Earth Sciences 77:490. doi:10.1007/s12665-018-7667-0
- Thomas MA, Mirus BB, Collins BD, Lu N, Godt JW (2018) Variability in soil-water retention properties and implications for physics-based simulation of landslide early warning criteria. Landslides 15:1265-1277. doi:10.1007/s10346-018-0950-z
- Tien Bui D, Ho TC, Revhaug I, Pradhan B, Nguyen DB (2014a) Landslide Susceptibility Mapping Along the National Road 32 of Vietnam Using GIS-Based J48 Decision Tree Classifier and Its Ensembles. In: Buchroithner M, Prechtel N, Burghardt D (eds). Cartography from Pole to Pole: Selected Contributions to the XXVIth International Conference of the ICA, Dresden 2013. Springer Berlin Heidelberg, Berlin, Heidelberg. pp 303-317
- Tien Bui D, Pradhan B, Revhaug I, Trung Tran C (2014b) A Comparative Assessment Between the Application of Fuzzy Unordered Rules Induction Algorithm and J48 Decision Tree Models in Spatial Prediction of Shallow Landslides at Lang Son City, Vietnam. In: Srivastava PK, Mukherjee S, Gupta M, Islam T (eds). Remote Sensing Applications in Environmental Research. Springer International Publishing, Cham. pp 87-111
- Tsangaratos P, Ilia I (2017) Chapter 24 - Applying Machine Learning Algorithms in Landslide Susceptibility Assessments. In: Samui P, Sekhar S, Balas VE (eds). Handbook of Neural Computation. Academic Press. pp 433-457
- Tsige M, Ruiz J, del Río IA, Jiménez-Díaz A (2016) Modeling of Landslides in Valles Marineris, Mars, and Implications for Initiation Mechanism. Earth, Moon, and Planets 118:15-26. doi:10.1007/s11038-016-9488-z

- Venkatesan M, Thangavelu A, Prabhavathy P (2013) An Improved Bayesian Classification Data Mining Method for Early Warning Landslide Susceptibility Model Using GIS. In: Bansal JC, Singh P, Deep K, Pant M, Nagar A (eds), Proceedings of Seventh International Conference on Bio-Inspired Computing: Theories and Applications (BIC-TA 2012). Springer India, India. pp 277-288
- Vens C (2013) Random Forest. In: Dubitzky W, Wolkenhauer O, Cho K-H, Yokota H (eds). Encyclopedia of Systems Biology. Springer New York, New York, NY. pp 1812-1813
- Wang Q, Li W, Chen W, Bai H (2015) GIS-based assessment of landslide susceptibility using certainty factor and index of entropy models for the Qianyang County of Baoji city, China. Journal of Earth System Science 124:1399-1415. doi:10.1007/s12040-015-0624-3
- 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. doi:<https://doi.org/10.1016/j.catena.2019.04.010>
- Wu C-H (2015) Landslide susceptibility mapping by using landslide ratio-based logistic regression: A case study in the southern Taiwan. Journal of Mountain Science 12:721-736. doi:10.1007/s11629-014-3416-3
- Wu C, Qiao J (2009) Relationship between landslides and lithology in the Three Gorges Reservoir area based on GIS and information value model. Frontiers of Forestry in China 4:165-170. doi:10.1007/s11461-009-0030-6
- Wu Q, Tang H, Ma X, Wu Y, Hu X, Wang L, Criss R, Yuan Y, Xu Y (2019) Identification of movement characteristics and causal factors of the Shuping landslide based on monitored displacements. Bulletin of Engineering Geology and the Environment 78:2093-2106. doi:10.1007/s10064-018-1237-2
- Xie J, Uchimura T, Chen P, Liu J, Xie C, Shen Q (2019) A relationship between displacement and tilting angle of the slope surface in shallow landslides. Landslides 16:1243-1251. doi:10.1007/s10346-019-01135-5
- Xie Z, Chen G, Meng X, Zhang Y, Qiao L, Tan L (2017) A comparative study of landslide susceptibility mapping using weight of evidence, logistic regression and support vector machine and evaluated by SBAS-InSAR monitoring: Zhouqu to Wudu segment in Bailong River Basin, China. Environmental Earth Sciences 76:313. doi:10.1007/s12665-017-6640-7
- Xu Q, Liu H, Ran J, Li W, Sun X (2016) Field monitoring of groundwater responses to heavy rainfalls and the early warning of the Kualiangzi landslide in Sichuan Basin, southwestern China. Landslides 13:1555-1570. doi:10.1007/s10346-016-0717-3
- Yin Y, Zheng W, Liu Y, Zhang J, Li X (2010) Erratum to: Integration of GPS with InSAR to monitoring of the Jiaju landslide in Sichuan, China. Landslides 7:391-391. doi:10.1007/s10346-010-0232-x
- Youssef AM, Pourghasemi HR, Pourtaghi ZS, Al-Katheeri MM (2016) Erratum to: Landslide susceptibility mapping using random forest, boosted regression tree, classification and regression tree, and general linear models and comparison of their performance at Wadi Tayyah Basin, Asir Region, Saudi Arabia. Landslides 13:1315-1318. doi:10.1007/s10346-015-0667-1
- Zhang C-X, Zhang J-S (2008) RotBoost: A technique for combining Rotation Forest and AdaBoost. Pattern Recognition Letters 29:1524-1536.

doi:<https://doi.org/10.1016/j.patrec.2008.03.006>

Zhang Z, Yang F, Chen H, Wu Y, Li T, Li W, Wang Q, Liu P (2016) GIS-based landslide susceptibility analysis using frequency ratio and evidential belief function models. Environmental Earth Sciences 75:948. doi:10.1007/s12665-016-5732-0

Zhao T, Dai F, Xu N-w (2017) Coupled DEM-CFD investigation on the formation of landslide dams in narrow rivers. Landslides 14:189-201. doi:10.1007/s10346-015-0675-1

Zhou C, Yin K, Cao Y, Ahmed B, Li Y, Catani F, Pourghasemi HR (2018) Landslide susceptibility modeling applying machine learning methods: A case study from Longju in the Three Gorges Reservoir area, China. Computers & Geosciences 112:23-37. doi:<https://doi.org/10.1016/j.cageo.2017.11.019>

Zhuang J, Peng J, Wang G, Javed I, Wang Y, Li W (2018) Distribution and characteristics of landslide in Loess Plateau: A case study in Shaanxi province. Engineering Geology 236:89-96. doi:<https://doi.org/10.1016/j.enggeo.2017.03.001>

Figures

Figure 1

Study area. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

Figure 2

Landslide conditioning factors: (a) plan curvature, (b) profile curvature, (c) slope angle, (d) slope aspect, (e) elevation, (f) TWI, (g) land use, (h) NDVI, (i) soil, (j) lithology, (k) distance to roads, Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

Figure 3

Landslide susceptibility maps: (a) J48 model, (b) ADTree model, (c) RaF model, (d) RF-J48 model, (e) RF-ADTree model, and (f) RF-RaF model. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

Figure 4

Percentages of landslide susceptibility grades.

Figure 5

ROC curves of the six models using training dataset.

Figure 6

ROC curves of the six models using validation dataset.