Geomechanics and Engineering, *Vol. 25, No. 1 (2021) 1-16*
DOI: https://doi.org/10.12989/gae.2021.25.1.001

Landslide susceptibility assessment using feature selection-based machine learning models

Lei-Lei Liu[1a], Can Yang[1b] and Xiao-Mi Wang*[2]

[1]*Key Laboratory of Metallogenic Prediction of Nonferrous Metals and Geological Environment Monitoring, Ministry of Education, School of Geosciences and Info-Physics, Central South University, Changsha 410083, P.R. China*
[2]*School of Resources and Environmental Science, Hunan Normal University, Changsha 410083, P.R. China*

(Received October 28, 2020, Revised January 18, 2021, Accepted February 6, 2021)

Abstract. Machine learning models have been widely used for landslide susceptibility assessment (LSA) in recent years. The large number of inputs or conditioning factors for these models, however, can reduce the computation efficiency and increase the difficulty in collecting data. Feature selection is a good tool to address this problem by selecting the most important features among all factors to reduce the size of the input variables. However, two important questions need to be solved: (1) how do feature selection methods affect the performance of machine learning models? and (2) which feature selection method is the most suitable for a given machine learning model? This paper aims to address these two questions by comparing the predictive performance of 13 feature selection-based machine learning (FS-ML) models and 5 ordinary machine learning models on LSA. First, five commonly used machine learning models (i.e., logistic regression, support vector machine, artificial neural network, Gaussian process and random forest) and six typical feature selection methods in the literature are adopted to constitute the proposed models. Then, fifteen conditioning factors are chosen as input variables and 1,017 landslides are used as recorded data. Next, feature selection methods are used to obtain the importance of the conditioning factors to create feature subsets, based on which 13 FS-ML models are constructed. For each of the machine learning models, a best optimized FS-ML model is selected according to the area under curve value. Finally, five optimal FS-ML models are obtained and applied to the LSA of the studied area. The predictive abilities of the FS-ML models on LSA are verified and compared through the receive operating characteristic curve and statistical indicators such as sensitivity, specificity and accuracy. The results showed that different feature selection methods have different effects on the performance of LSA machine learning models. FS-ML models generally outperform the ordinary machine learning models. The best FS-ML model is the recursive feature elimination (RFE) optimized RF, and RFE is an optimal method for feature selection.

Keywords: landslide; susceptibility assessment; machine learning; feature selection; Geographic Information System (GIS)

1. Introduction

Landslide is a complex natural phenomenon (Boulfoul *et al.* 2020, Liu *et al.* 2020, Xing *et al.* 2019). It has the characteristics of wide distribution, high frequency of occurrence, fast movement, and among others (Liu and Chen 2015, Lombardi *et al.* 2017, Shou and Lin 2020). Such characteristics frequently threaten the lives and property of people and the ecological environments in disaster areas. Landslide susceptibility assessment (LSA) can effectively predict the spatial probability of regional landslides occurrence. The landslide susceptibility map (LSM) obtained from the LSA results hence provides useful information for regional landslide disaster prevention and mitigation. It is thus of great significance to conduct LSA to reduce the landslide disaster risk.

To date, many efforts have been made to propose quantitative prediction models for LSA based on the Geographic Information System (GIS) platform. According to Reichenbach *et al.* (2018), models for LSA can be divided into five categories, namely the geomorphological mapping (Paola *et al.* 2004), spatial analysis of landslide inventories (Degraff and Canuti 1988), heuristic models (Reichenbach *et al.* 2018), physical models (Balzano *et al.* 2019; Cheng *et al.* 2018) and statistical models (Liu *et al.* 2018; Sheil Brian *et al.* 2020). Among these models, machine learning models, which belong to the statistical models, have been gaining increasing attention in LSA. An important reason is that machine learning models can accurately represent the nonlinear relationship between the landslide susceptibility and conditioning factors. Another is that machine learning models do not require the conditioning factors to be normally distributed, which are especially suitable for the susceptibility analysis of landslides in regional scales (Bui *et al.* 2016). The commonly used machine learning models for LSA include support vector machine (SVM) (Hong *et al.* 2018b), artificial neural network (ANN) (Chen *et al.* 2017b), fuzzy

*Corresponding author, Ph.D.
E-mail: xiaomiw@yeah.net
[a]Ph.D.
E-mail: csulll@foxmail.com
[b]Master Student
E-mail: 195011074@csu.edu.cn

Copyright © 2021 Techno-Press, Ltd.
 ISSN: 2005-307X (Print), 2092-6219 (Online)

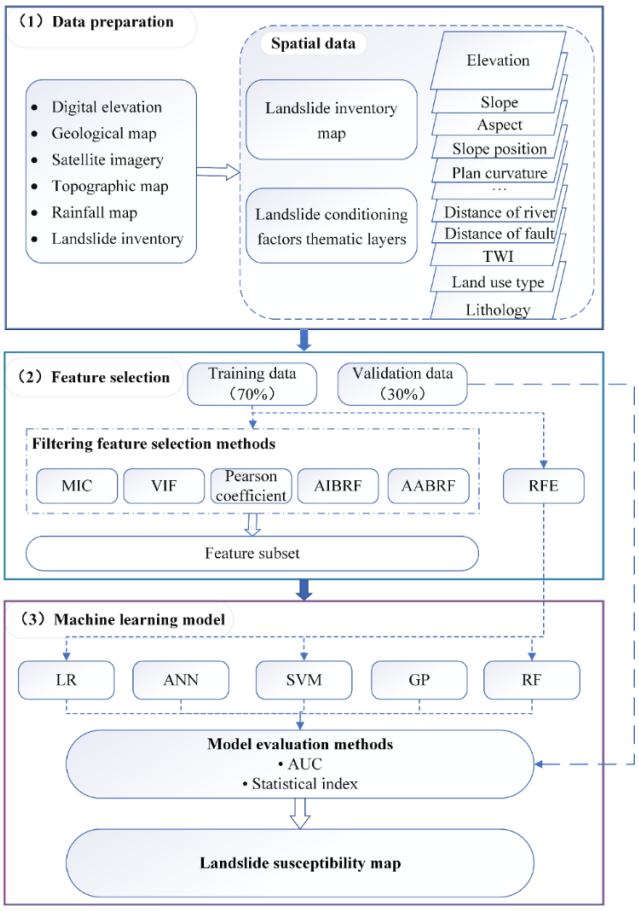

Fig. 1 Flowchart of landslide susceptibility modeling

inference (Chen *et al.* 2017a), decision tree (DT) (Hong *et al.* 2018a) and random forest (RF) (Catani *et al.* 2013). The applications of these models have greatly facilitated the development of LSA.

Since many conditioning factors are influencing the susceptibility, the data availability and number of the factors can significantly affect the performance of the machine learning models. Generally speaking, for a specific region, there are about 2 to 22 regulatory factors of LSA, but the number of factors that have appeared can be as high as 596 (Reichenbach *et al.* 2018). Considering all factors in the machine learning model will not only lead to difficulties in data collection but also affect the computational efficiency and accuracy. By contrast, if only a few factors are considered, the accuracy of the susceptibility model will be reduced. Therefore, to optimize the machine learning models, feature selection methods are often utilized to select the few but relatively important features (or factors) because the factors are often somewhat correlated with each other (Kavzoglu and Mather 2010). For example, Vasu *et al.* (2016) used a hybrid feature selection algorithm integrating the extreme learning machine for landslide susceptibility modeling of Mt. Woomyeon, South Korea. Lagomarsino *et al.* (2017) presented a tool for classification and regression based on the RF technique, which performs automatically the feature selection based on a quantitative criterion. Sun (2020) applied the RF model and the recursive feature elimination (RFE) method to carry out the analysis of landslide susceptibility mapping. Pham *et al.* (2020) proposed using correlation diagnosis for feature selection, and only eight factors were selected for the modeling. It is generally found from these studies that both the accuracy and efficiency of the feature selection-based machine learning (FS-ML) models are increased; and the increase in prediction accuracy may further have a great impact on the landslide susceptibility zoning results, even if the increase is slight (Bui *et al.* 2016). However, in most of the previous studies, only a single feature selection method is used to process the input data (Amato *et al.* 2019, Sun *et al.* 2020). Since there are many choices of feature selection methods, it is still not clear which feature selection method is the most suitable for a given machine learning model. In addition, how different feature selection methods affect the performance of different machine learning models is still an open question.

To summarize, this paper aims at addressing the above two issues by performing a systematic comparative study on various FS-ML models. As a result, a more accurate and reliable LSA can be achieved via selecting an optimal FS-ML model. Thirteen FS-ML models will be considered, which are constructed by coupling five commonly used machine learning models (i.e., Logistic regression (LR), SVM, ANN, Gaussian process (GP) and RF) and typical feature selection methods. To the best of our knowledge, such a systematic comparative study on FS-ML is limited in the literature. Four counties in central Hunan Province, China are chosen as the study area, because landslides frequently occur there. Then, various FS-ML models are constructed based on recorded landslides and associated conditioning factors. The LSA results are evaluated by the area under curve (AUC) value and statistical indices such as sensitivity, specificity and accuracy.

2. Methodology

The purpose of this study is to explore and compare the effectiveness of six feature selection methods in optimizing the performance of five commonly used machine learning models for a reasonable and reliable LSA. The flowchart for achieving this purpose is shown in Fig. 1. It can be seen from the figure that the LSA with FS-ML in this study mainly includes three steps: data preparation, feature selection, and landslide susceptibility model establishment and evaluation. The first step is about collecting the data of landslides and the associated conditioning factors, which includes compiling a landslide inventory database and constructing training and validation data sets. Then, in the second step, feature selection methods, such as maximum information coefficient (MIC), multicollinearity diagnosis, average impurity and average accuracy based on random forest (AIBRF and AABRF) and RFE, are applied to the above data to obtain the feature subsets of the conditioning factors. It should be noted that the prerequisite of using RFE with machine learning models is that the model itself can evaluate the feature importance. Hence, RFE is only used for LR, SVM and RF to form RFE-based FS-ML models in the current study. Finally, with the feature subsets

obtained above, the afore-mentioned five machine learning models are used to establish the landslide susceptibility prediction models using the corresponding training data set (70% of the landslide inventory data). The performance of the models is evaluated based on the validation data set (30% of the landslide inventory data) by using the statistical indices (e.g., sensitivity, specificity and accuracy) and receive operating characteristic (ROC) curve. In the following subsections, the methodologies constituting the above steps of LSA with FS-ML are described in detail.

2.1 Machine learning models

2.1.1 Logistic regression

LR is a specific type of generalized linear model. It is frequently used for binary classification problems with probability theory. Consider, for example, the LSA in this study. Within the framework of the LR model, the occurrence of landslides is considered as the binary dependent variable, which takes the values of 0 and 1 to represent the absence and presence of landslides, respectively (Yalcin *et al.* 2011). Mathematically, LR relates the occurrence probability P of landslides to a 'logit' function with the assumption that landslide occurrence is dependent on one or more landslide conditioning factors. The relationship is described as

$$P = \frac{\exp(b_0 + b_1x_1 + b_2x_2 + \cdots + b_nx_n)}{1 + \exp(b_0 + b_1x_1 + b_2x_2 + \cdots + b_nx_n)} \quad (1)$$

where P is the probability of landslide occurrence, varying between 0 and 1; $x_1, x_2, \ldots, x_n$ represent the landslide conditioning factors; $b_1, b_2, \ldots, b_n$ are the model coefficients; and b_0 is a model constant (Lombardo and Mai, 2018). It is obvious that Eq. (1) is a parametric model without "hyper-parameters" to be tuned, thereby making LR especially suitable for LSA. However, to obtain accurate LSA results with LR, the following requirements should be satisfied (Merghadi *et al.* 2020): (1) landslide occurrence (or dependent variable in LR) is in binary form; (2) minimum duplicates are attained in the input data and the data size is large; (3) little or no multicollinearity exists among landslide conditioning factors; and (4) conditioning factors and log of odds are in linear form.

2.1.2 Support vector machine

SVM is a promising classification method, which was proposed by Cortes and Vapnik (1995) based on the concept of structural risk minimization and statistical learning theory. The main idea is to map nonlinearly separable data to a high-dimensional feature space through a nonlinear mapping and to find the optimal classification hyperplane that maximally splits two sample classes in this feature space. SVM is especially suitable for binary classification problems. Consider, for example, the LSA problem with positive and negative landslide inventory data samples. The principle of SVM is briefly introduced as follows. Suppose the training data set is $(\boldsymbol{x}_i, \boldsymbol{y}_i)$ $(i = 1, 2, .., n)$, where $\boldsymbol{x}_i$ is the landslide conditioning factor vector and $\boldsymbol{y}_i = \pm 1$ corresponds to the absence (-1) or presence (+1) of landslide for the ith sample. Through a nonlinear mapping function $\varphi(\boldsymbol{x})$, the original data is then transferred into a feature space, where an optimal hyperplane of $\boldsymbol{\omega}^* \cdot \varphi(\boldsymbol{x}) + \mathrm{b}^* = 0$ can be found to maximize the classification interval between two sample classes. Subsequently, the following classification function $f(\boldsymbol{x})$ in Eq. (2) can be used to predict the classification type (e.g., the absence or presence of landslides) at unsampled locations as

$$f(\boldsymbol{x}) = \mathrm{sign}(\boldsymbol{\omega}^* \cdot \varphi(\boldsymbol{x}) + \mathrm{b}^*) \quad (2)$$

Details for solving for the optimal parameters $\boldsymbol{\omega}^*$ and b^* and hyper-parameters settings can be found in the literature (Merghadi *et al.* 2020).

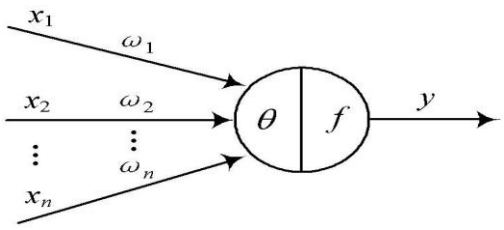

Fig. 2 Artificial neuron networks model

2.1.3 Artificial neural networks

ANN (Chen *et al.* 2017b) is a general-purpose nonlinear function approximator that is widely used in pattern recognition and classification. ANN consists of several basic units (neurons), which can realize the nonlinear calculation of the input data. Fig. 2 is a structural model of a single neuron. In the figure, $\boldsymbol{x} = [x_1, x_2, \ldots, x_n]^T$ indicates the landslide conditioning factors, ω_i is the weight of the corresponding factor, θ is the threshold of the neuron, and f is the transfer function of the neuron. In this study, the transfer function is the sigmoid function as

$$f(z) = \frac{1}{1 + e^{-z}} \quad (3)$$

where $z = \sum_i^n \omega_i x_i - \theta$ is the sum of neuron inputs. Then, the output y of the neuron is expressed as

$$y = f\left(\sum_i^n \omega_i x_i - \theta\right) \quad (4)$$

2.1.4 Gaussian process

GP (Rasmussen and Nickisch 2010) is a statistically significant machine learning method based on the Bayesian framework. GP has good self-adaptability to complex problems such as small samples, high dimensionality, and nonlinearity. The classification process of GP is described as follows. Suppose the training dataset is $\mathbf{S} = \{(\boldsymbol{x}_i, y_i) | i = 1, 2, 3, \ldots, n\}$, where n denotes the number of training data or landslide samples and $\mathbf{X} = [\boldsymbol{x}_i, \ldots, \boldsymbol{x}_n]^T$ and $\boldsymbol{y} = [y_i, \ldots, y_n]^T$ are respectively the input dataset and output dataset. The class labels in $\boldsymbol{y}$ are independent and identically distributed. In the binary classification problem, the probability that the sample $\boldsymbol{x}_i$ belongs to the class labels y_i can be expressed as

$$p(y_i|f_i) = \mathrm{Sig}(y_i f_i) \quad (5)$$

where $f_i = f(\boldsymbol{x}_i)$, and $\mathrm{Sig}(\cdot)$ is a function to convert the output value into a probability to obtain the probability that the test sample belongs to a certain class. $\mathrm{Sig}(\cdot)$ is usually termed as the response function, from which the likelihood function can be obtained as

$$p(\boldsymbol{y}|\boldsymbol{f}) = \prod_{i=1}^{n} p\left(y_i|f(\boldsymbol{x}_i)\right) = \prod_{i=1}^{n} \mathrm{Sig}(y_i f_i) \quad (6)$$

The posterior probability can be obtained from the Bayesian formula as

$$p(\boldsymbol{f}|\mathbf{X}, \boldsymbol{y}) = \frac{p(\boldsymbol{y}|\boldsymbol{f})p(\boldsymbol{f}|\mathbf{X})}{p(\boldsymbol{y}|\mathbf{X})} \quad (7)$$

In summary, for the specified data $\mathbf{X}_*$, the posterior probability of the corresponding $\boldsymbol{f}_*$ is

$$p(\boldsymbol{f}_*|\mathbf{X}, \boldsymbol{y}, \mathbf{X}_*) = \int p(\boldsymbol{f}_*|\boldsymbol{f}, \mathbf{X}, \mathbf{X}_*)\, p(\boldsymbol{f}|\mathbf{X}, \boldsymbol{y}) d\boldsymbol{f} \quad (8)$$

Further deducing the classification prediction probability corresponding to f_* is

$$p(\boldsymbol{y}_*|\mathbf{X}, \mathbf{X}_*, \boldsymbol{y}) = \int \mathrm{Sig}(\boldsymbol{y}_* \boldsymbol{f}_*) p(f_*|\mathbf{X}, \mathbf{X}_*, \boldsymbol{y})\, d\boldsymbol{f}_* \quad (9)$$

In the GP model, $p(\boldsymbol{y}_*|\mathbf{X}, \mathbf{X}_*, \boldsymbol{y})$ = 0.5 is usually used as the threshold. When $p > 0.5$, $y_* = +1$, representing landslide occurrence; otherwise $y_* = -1$, representing landslide absence. For more details on the GP model, the reader is referred to as Skolidis and Sanguinetti (2011).

2.1.5 Random forest

RF refers to a classification model that uses multiple decision trees as classifiers to train and predict samples. It was first proposed by Breiman (2001). The model combines the "Bootstrap aggregating" idea and the "random subspace" method. Since the formation of multiple decision trees uses a random method, the RF model is also called a random decision tree. As a highly flexible machine learning method, RF has been widely used in the assessment of geological disaster susceptibility. The basic principle of RF is to repeatedly use resampling technology (or bootstrapping) to randomly select a certain number of samples from the original training sample set with replacements to generate a new training sample set. An independent decision tree classifier combines these classifiers to form an RF model. For prediction data, the classification results should be determined according to the final voting results of these decision tree classifiers, and its essence belongs to an improved algorithm of the decision tree model.

It should be noted that the above machine learning models in the following applications are achieved by using scikit-learn (http://scikit-learn.org), an open-source framework developed specifically for machine learning applications in Python language. Additionally, Bayesian optimization (https://github.com/fmfn/Bayesian Optimization) is utilized to determine the hyper-parameters of the models. As a reference, the hyper-parameters for each of the five models that are best built after feature selections are shown in Table 1.

Table 1 Hyper-parameters of machine learning models best built after feature selection

Model	Hyper-parameter
LR	['C'=0.004 , 'solver'= sag , 'penalty' =L2]
SVM	['kernel'= rbf, 'C' =100, 'gamma'=10^(-1.924)]
ANN	['hidden_layer_sizes'=133, 'solver'=lbfgs, 'alpha'=0.816]
GP	['optimizer'=fmin_l_bfgs_b, 'max_iter_predictint'=100]
RF	['max_depth'=15, 'max_features'=0.2479918, 'min_samples_split'=2, 'n_estimators'= 400]

2.2 Feature selection

In the modeling of landslide susceptibility, the selection of landslide conditioning factors is an important step, because there may be some noise factors that can reduce the predictive ability of the model. In this study, filtering feature selection methods, such as MIC, multicollinearity diagnosis, AIBRF and AABRF, are used to evaluate the importance of each conditioning factor to create a subset of features accordingly. In addition, a wrapped feature selection method, RFE, is also adopted to examine its suitability for optimizing machine learning models. Note that since the major aim of the study is to select important factors, rather than to create new factors, methods that introduce some principle factors to represent all factors, such as principle component analysis, are not considered herein (Wold *et al.* 1987).

2.2.1 Maximum information coefficient

MIC is a new correlation measure proposed by Reshef *et al.* (2011). This method is improved based on mutual information, so it is more universal and fairer than mutual information. The formula for calculating the MIC of landslide condition factors is as follows:

$$MIC(D) = \max_{XY<B(n)} \frac{I(D, X, Y)}{\log(\min(X, Y))} \quad (10)$$

where B is a function of the sample size n, e.g., $B=n^{0.6}$; $I(D, X, Y)$ refers to the largest mutual information value falling into the grid area D. The value range of MIC is [0, 1]. When MIC=0, it means that the X and Y variables are independent of each other. By contrast, the X and Y variables are somewhat related to each other when MIC=1.

2.2.2 Multicollinearity diagnosis

To avoid multicollinearity and minimize bias in the model results, researchers usually use the Pearson correlation matrix (Reichenbach *et al.* 2018), variance inflation factor (VIF) and tolerance index (TI) (Pourghasemi *et al.* 2020). Generally, Pearson correlation, with a value between -1 and 1, is widely used to measure the correlation between two variables as

$$\rho_{x,y} = \frac{cov(x, y)}{\sigma_x \sigma_y} = \frac{E\left[(x - \mu_x)(y - \mu_y)\right]}{\sigma_x \sigma_y} \quad (11)$$

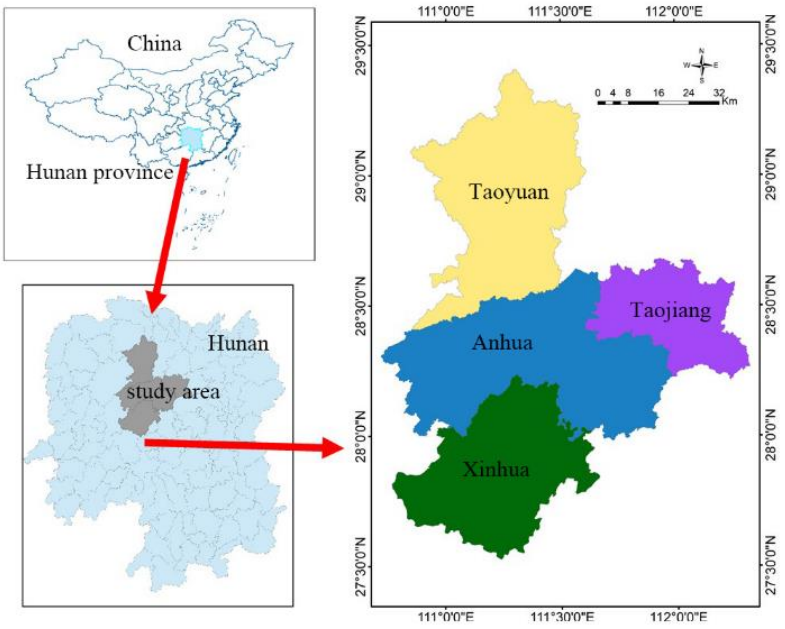

Fig. 3 Location of the study area

where $cov(x, y)$ is the covariance between two variables; σ_x and σ_y are standard deviations of x and y, respectively.

VIF focuses on the standard error variations of landslide conditioning factors. The closer the VIF is to 1, the lighter the multicollinearity is, and vice versa. If the VIF value is larger than 5, it indicates that there is a multicollinearity problem in the conditioning factors. Compared with VIF, TI is a similar measure but scaled inversely, with TI < 0.2 indicating high multicollinearity. The expression of VIF is written as

$$\mathrm{VIF} = \left(\frac{1}{1 - R^2}\right) \tag{12}$$

where R^2 is the coefficient of determination.

2.2.3 Average impurity and average accuracy based on random forest

The RF provides two methods for feature selection: average impurity reduction and average accuracy reduction (Reif *et al.* 2006). AIBRF indicates the average reduction in error of each feature. AABRF measures the impact of order changes on the accuracy of the model when the order of eigenvalues of each feature is disturbed. For unimportant features, the order of disruption will not affect the accuracy of the model too much, but for important features, the order of disruption will reduce the accuracy of the model.

2.2.4 Recursive feature elimination

RFE (Guyon *et al.* 2002) is performed on the prediction model with weights in the features. Through recursive methods, the size of the feature set is continuously reduced to select the required features. The implementation process is as follows: (1) use these original features to train the prediction model to obtain the weights of the features; (2) delete the feature with the minimum weight; (3) construct the LSA model based on the updated feature set in the last step and obtain the AUC value of the model; and (4) continue steps (1) to (3) until the model AUC reaches its maximum or the required number of features. The order for eliminating the features in the above process is the ranking of the features. Since GP and ANN do not weigh each feature, RFE is only used in the remaining three models.

2.3 Model evaluation methods

2.3.1 The receiver operating characteristic curve

The ROC curve defines the performance of the two-classifier system as the change in its recognition threshold (Wang *et al.* 2015). The ROC curve expresses sensitivity as a function of the false positive rate (1-specificity). It can be generated by plotting the cumulative distribution function of the sensitivity on the *y*-axis and the false positive rate on the *x*-axis. It has been widely used as a standard tool for evaluating the overall performance of LSA models (Hong *et al.* 2018c). The AUC value is a quantitative indicator of the model quality, which is divided into poor (0.5-0.6), average (0.6-0.7), good (0.7-0.8), very good (0.8-0.9), and excellent (0.9-1) (Samia *et al.* 2020). The higher the AUC value is, the better the model would be. The AUC value of 1 indicates that the model is perfect (Youssef *et al.* 2015).

2.3.2 Statistical index

In the assessment of landslide susceptibility, most researchers believe that scientific methods should be used to evaluate the performance of landslide susceptibility prediction models. However, there is no clear agreement on which methods should be used. In this study, sensitivity, specificity, and accuracy are used to evaluate the performance of the landslide susceptibility prediction model. The precise definition of these statistical measures has been elaborated in many landslide studies (Bui *et al.*

2016). They can be calculated using the following equations as

$$\text{sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{13}$$

$$\text{specificity} = \frac{\text{TN}}{\text{FP} + \text{TN}} \tag{14}$$

$$\text{accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{15}$$

where the false positive (FP) class and false negative (FN) class are the numbers of pixels that are misclassified, while the true positive (TP) class and true negative (TN) class are the numbers of pixels that are correctly classified.

3. Study area and data

3.1 Study area

The study area is located in the central area of Hunan Province, China (hereinafter abbreviately referred to as central Hunan), covering Taoyuan County, Anhua County, Taojiang County, and Xinhua County, as shown in Fig. 3. From the figure, it is seen that the area spans east longitude 110°43′07″-112°19′5″and north latitude 27°31′06″-29°24′03″. The area is about 15212.51 km^2, accounting for about 7.18% of the total area of Hunan province. The area has a subtropical monsoon climate. The main lithologic units in the study area include limestone, marl, granite, conglomerate, marble, kaolin, potash feldspar, sand, sandstone, slate, clay, quartz. The topography of the study area is mainly the result of geological structure and weathering and erosion mechanisms. The highest observation altitude is 1585 m, and the lowest observation altitude is -135 m. Areas with an altitude less than 200 m account for about 50% of the study area, and only 3% of the areas are above 1200 m. Areas with slopes greater than 46° account for approximately 2% of the total area, while areas with slopes less than 15° account for approximately 51%. The study area has abundant precipitation, developed water systems, and interlaced rivers and streams. The average annual rainfall from 1981 to 2010 is 1556 mm (http://data.cma.cn) and the average annual temperature is 16.8°C. The rainy season is from March to August, and April to June accounts for 62.6% of the total rainfall in the rainy season and 46.7% of the total rainfall in the whole year. The average rainfall in May and June is 200 mm and 350 mm, respectively. The land-use types in the study area are forest land (about 47%), agricultural land (3.9%), residential (7.2%), bare land (0.1%), grassland (39.5%) and water body (2%). According to the statistics of the local government in the study area, a total of 2115 people were affected by landslides, and the loss of property and infrastructure is estimated to be about 30 million yuan. However, to date, there have been few studies on predicting the spatial distribution of landslides and preventing their damage.

3.2 Data

3.2.1 Landslide inventory map

Obtaining historical landslide data is one of the important steps of landslide susceptibility evaluation. In this study, 1017 landslide disaster points collected by the Hunan Provincial Geological Monitoring Station based on field surveys from 2015 to 2018 were used. Of the 1,017 landslides, 739 were caused by rainfall, 260 by slope cutting and 18 by a combination of rainfall and slope cutting. Based on the collected historical landslide data, the landslides in the study area are mainly small, and 85% of the landslides are within 30 m in length and width. Hence, a landslide can be abstracted as a pixel in the 30 m raster data, as shown in Fig. 4.

3.2.2 Landslide conditioning factors

The development of landslide disasters is generally affected by many conditioning factors, for example, geological conditions, geomorphic conditions, hydrological conditions, and human engineering activities (Hong *et al.* 2016). These conditions have different primary and secondary status and functions in different regions. According to previous studies (Moore *et al.* 1991; Weiss, 2001) and the landslide characteristics in the study area, a total of 15 landslide conditioning factors are considered for LSA in this study, including elevation, slope degree, aspect, slope position, micro-geomorphology, plane curvature, profile curvature, topographic wetness index, lithology, distance from fault (D.F. fault), vegetation coverage, land use type, distance from river (D.F. river), average precipitation and distance from road (D.F. road). The generation of the thematic layers of these factors and the corresponding data sources are described as follows.

First, based on the ASTER global digital elevation model with a spatial resolution of 30 m in ArcGIS 10.4 environment, the spatial analysis tools are used to generate elevation (Fig. 4(a)), slope (Fig. 4(b)), aspect (Fig. 4(c)), and plane curvature (Fig. 4(f)), profile curvature (Fig. 4(g)) and topographic wetness index map (Fig. 4(h)); the topography tools are used to generate slope position (Fig. 4(d)) and micro-geomorphology index map (Fig. 4(e)). Then, with the 1: 250,000 scale geological map, the lithology layer is divided into 8 categories, and the lithology map (Fig. 4(i)) and fault distance map (Fig. 4(j)) are compiled. In addition, the NDVI map (Fig. 4(k)) is obtained by using the ArcGIS band synthesis function based on the Landsat8 satellite image with a resolution of 30 meters (http://www.gscloud.cn). The land-use type map (Fig. 4(l)) comes from the national geographic information resource directory service system (http://www.webmap.cn). Moreover, based on the topographic map with a scale of 1: 250,000, the distance maps from rivers and roads are constructed (Fig. 4(m)-(n)) by using the buffer tool of the ArcGIS software. Finally, since the landslide disasters in the study area all occurred during the flood season, the annual average flood season precipitation from 146 rainfall stations from 2015 to 2018 was used to construct the precipitation map (Fig. 4(o)).

Furthermore, it is worthwhile to point out that, since

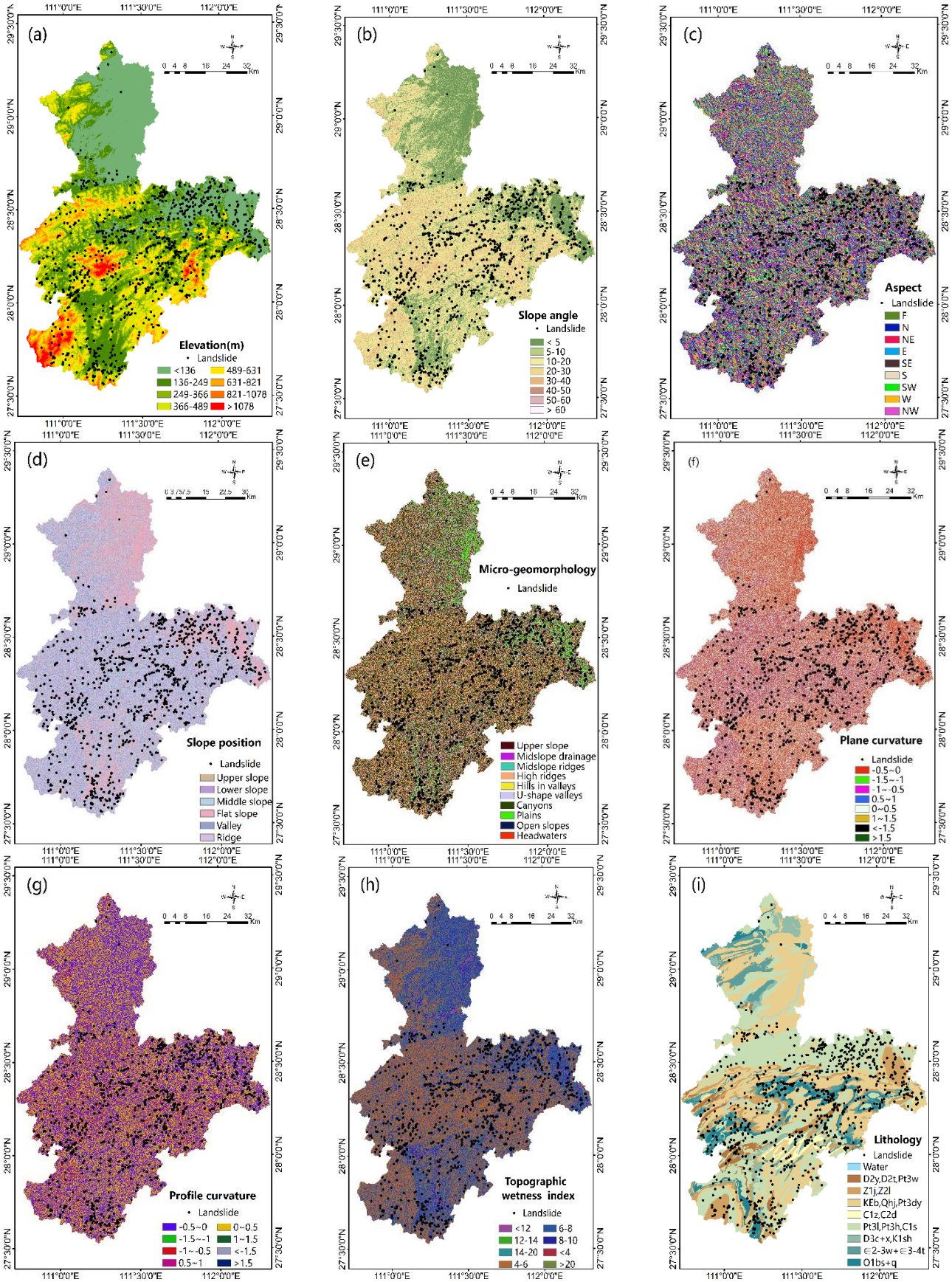

Fig. 4 Landslide conditioning factor maps: (a) Elevation, (b) Slope, (c) Aspect, (d) Slope position, (e) Microrelief, (f) Plan curvature, (g) Profile curvature, (h) Topographic wetness index, (i) Topography, (j) Distance from fault, (k) NDVI, (l) Land-use type, (m) Distance from river, (n) Perennial mean precipitation and (o) Distance from road

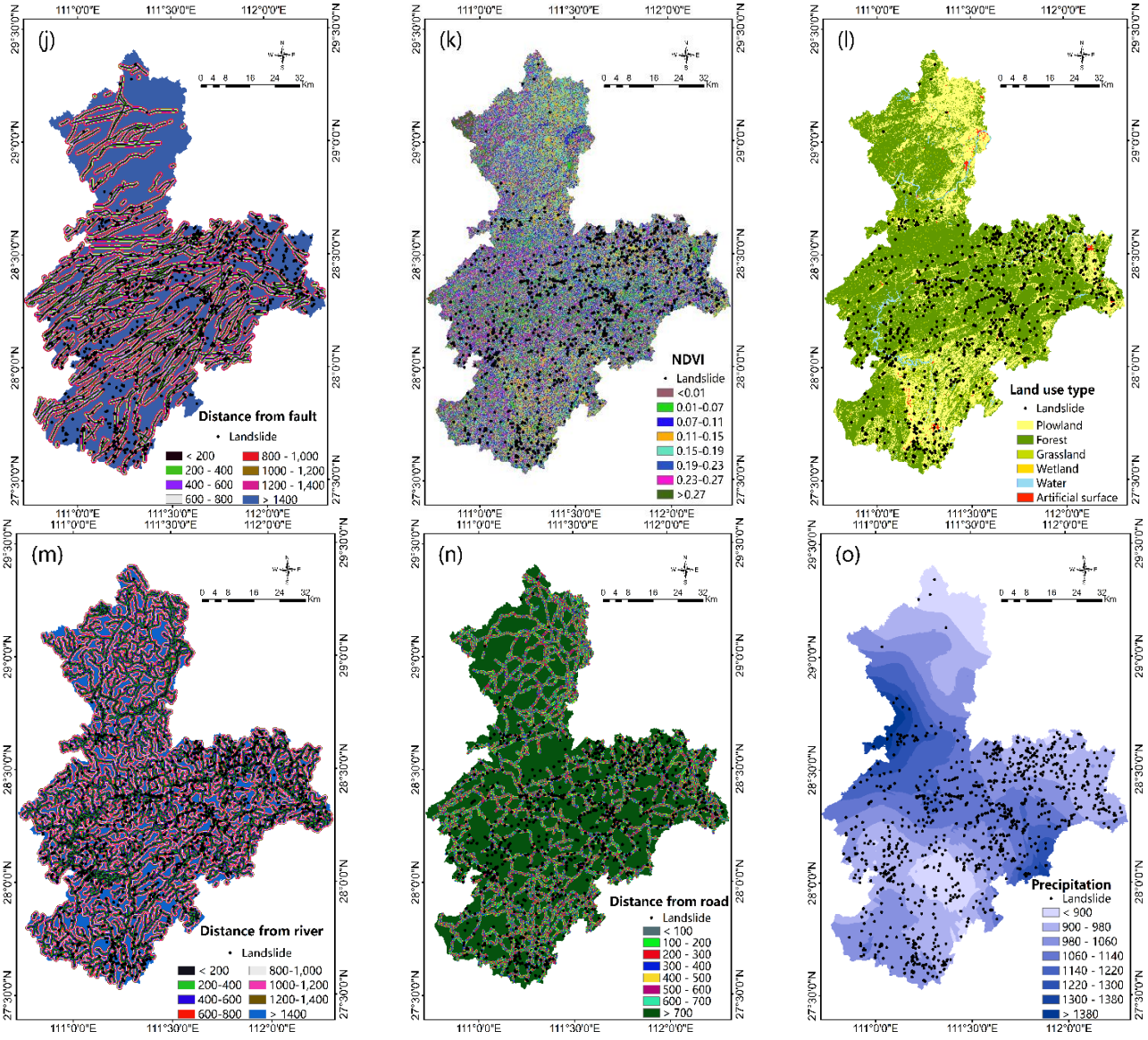

Fig. 4 (Continued)

statistical machine learning models are used to predict the susceptibility in this study, other physical factors that affect the development of landslides are not considered, such as the material interface behaviour (Li *et al.* 2016). Hence, if physical models are adopted for LSA, it is suggested to incorporate such factors into the model. However, it is generally difficult to obtain such physical data in a large study area, and physical models are very time-consuming.

3.2.3 Training and validation data

The collected landslide data are all positive samples. However, for machine learning model construction, negative samples are necessary. In general, negative samples are randomly sampled from the interpreted conditioning factor maps, but there is no agreement on determining the number of negative samples in the literature. Hence, a parametric study on the effect of the ratio of landslide to non-landslide grid cells on the machine learning model was conducted in advance to obtain a relatively higher prediction accuracy. Three kinds of ratio values of 1:1, 1:5 and 1:10 were considered. The results showed that the model worked best when the sample ratio is 1:1. Therefore, in this study, the same number (1017) of non-landslide samples as the landslide samples were randomly selected in the non-landslide area. They were combined with the 1017 recorded landslide samples to form the sample set for machine learning training and validation. Specifically, according to the literature (Chen *et al.* 2018), 70% of the sample set is for training the machine models, while the remaining 30% is used for model validation purposes. Finally, it should be noted that in the process of model establishment and verification, the landslide sample is assigned a value of 1 and the non-landslide point is assigned a value of 0 for supervised machine learning.

4. Results

4.1 LSA results based on ordinary machine learning models

Before constructing FS-ML models for LSA, this part

Table 2 Performance results of ordinary machine learning models

Model	AUC	TN	FP	FN	TP	Sensitivity	Specificity	Accuracy
LR	0.706	182	125	96	207	0.683	0.593	0.638
SVM	0.693	174	133	86	217	0.716	0.567	0.641
ANN	0.680	185	122	108	195	0.644	0.603	0.623
GP	0.718	182	125	93	210	0.693	0.593	0.643
RF	0.742	188	119	87	216	0.713	0.612	0.662

Table 3 Feature importance results based on MIC, VIF, AIBRF and AABRF

Features	MIC	VIF	AIBRF	AABRF
Micro-geomorphology	0.015	3.416	0.053	-0.0003
Land use	0.028	1.059	0.049	0.028
Profile curvature	0.006	1.820	0.035	-0.0006
Aspect	0.006	1.014	0.089	0.021
Slope position	0.010	4.559	0.058	0.002
Slope degree	0.013	1.781	0.059	0.012
Plane curvature	0.003	1.805	0.034	-0.0014
Elevation	0.025	1.490	0.071	0.023
NDVI	0.018	1.193	0.079	0.023
TWI	0.018	1.708	0.055	0.008
D.F. river	0.031	1.180	0.090	0.034
D.F. road	0.040	1.181	0.087	0.052
D.F. fault	0.008	1.043	0.081	0.023
Lithology	0.017	1.031	0.075	0.020
Precipitation	0.017	1.044	0.087	0.064

first uses ordinary machine learning models (i.e., the proposed machine learning models are constructed without feature selection) for the susceptibility analysis, which is considered as the benchmark of the study. Table 2 gives the performance results of the considered machine learning models without feature selection. It shows that RF outperforms its peers in terms of both AUC value and statistics (e.g., AUC=0.742, Accuracy=0.662). ANN has the worst performance of all models with the AUC=0.680 and Accuracy=0.623. The descending order of the model performance based on AUC and accuracy is RF, GP, LR, SVM, and ANN, which is generally consistent with previous findings (Pourghasemi and Rahmati 2018).

4.2 LSA results based on FS-ML models

4.2.1 Feature selection results

Table 3 shows the feature selection results of the aforementioned 15 independent features (conditioning factors) using MIC, VIF, AIBRF and AABRF. It can be seen from the table that different feature selection methods produce different feature ranking results. The reason is that the internal mechanism of each feature selection method is different (Micheletti *et al.* 2014). The MIC shows that D.F. road (MIC=0.04) and D.F. river (MIC=0.04) are the most important features for landslide prediction herein, followed

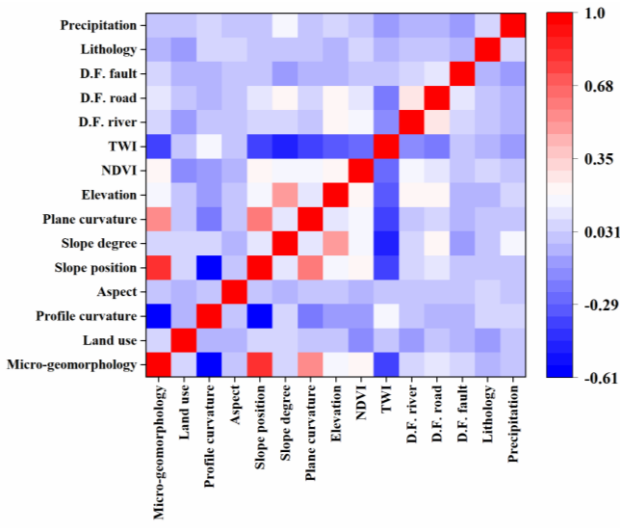

Fig. 5 The Pearson correlation coefficient matrix

by the land use (MIC=0.028), altitude (MIC=0.025), NDVI (MIC=0.018), TWI (MIC=0.018), precipitation (MIC=0.017), lithology (MIC=0.017), micro-geomorphology (MIC=0.015), slope degree (MIC=0.013), slope position (MIC=0.01), D.F. fault (MIC=0.008), aspect (MIC=0.006), profile curvature (MIC=0.006), plane curvature (MIC=0.003). In addition, the results in the second column show that the VIF value of slope position are obviously larger than those of the other features.

The AIBRF results show that the feature importance of the D.F. river is the highest among all features with a value of 0.09 (Pourghasemi *et al.* 2013), whereas plan curvature (AIBRF=0.034) and profile curvature (AIBRF=0.035) are considered of low importance. The features in the middle are aspect (AIBRF=0.089), D.F. road (AIBRF=0.087), precipitation (AIBRF=0.087), D.F. fault (AIBRF=0.081), NDVI (AIBRF=0.079), lithology (AIBRF=0.075), elevation (AIBRF=0.071), slope degree (AIBRF=0.059), slope position (AIBRF=0.058), TWI (AIBRF=0.055), micro-geomorphology (AIBRF=0.053), and land use (AIBRF=0.049)

In AABRF results, micro-geomorphology (AABRF=-0.0003), profile curvature (AABRF=-0.0006) and plane curvature (AABRF=-0.0003) are negative, suggesting that the three features are useless for landslide susceptibility prediction. Similar as AIBRF, AABRF also considers precipitation (AABRF=0.064), D.F. road (AABRF=-0.052), D.F. river (AABRF=0.034) are top three important features of landslide. After the D.F. river, the importance rank is the land use (AABRF=0.028), D.F. fault (AABRF=0.023), NDVI (AABRF=0.023), elevation (AABRF=0.023), aspect (AABRF=0.021), lithology (AABRF=0.02), slope degree (AABRF=0.012), TWI (AABRF=0.008), and slope position (AABRF=0.002).

Fig. 5 is the output results of the Pearson correlation matrix among different features. According to the correlation analysis, most features are positively or negatively correlated with each other, only the correlation coefficient between micro-geomorphology and slope position is greater than the threshold of 0.7. The slope

Table 4 Performance results of filter-based FS-ML models

Model	Subsets	AUC	TN	FP	FN	TP	Sensitivity	Specificity	Accuracy
LR	A	0.707	182	125	94	209	0.690	0.593	0.641
	B	0.706	182	125	92	211	0.696	0.593	0.644
SVM	A	0.696	178	129	91	212	0.700	0.580	0.639
	B	0.702	178	129	83	220	0.726	0.580	0.652
ANN	A	0.649	174	133	110	193	0.637	0.567	0.602
	B	0.681	192	115	116	187	0.617	0.625	0.621
GP	A	0.719	184	123	93	210	0.693	0.599	0.646
	B	0.715	172	135	86	217	0.716	0.560	0.638
RF	A	0.741	184	123	84	219	0.723	0.599	0.661
	B	0.742	189	118	84	219	0.723	0.616	0.669

Table 5 Performance results of RFE-based FL-ML models

Model	AUC	TN	FP	FN	TP	Sensitivity	Specificity	Accuracy
LR	0.718	187	120	92	211	0.696	0.609	0.652
SVM	0.717	188	120	93	211	0.694	0.610	0.652
RF	0.744	192	115	86	217	0.716	0.625	0.670

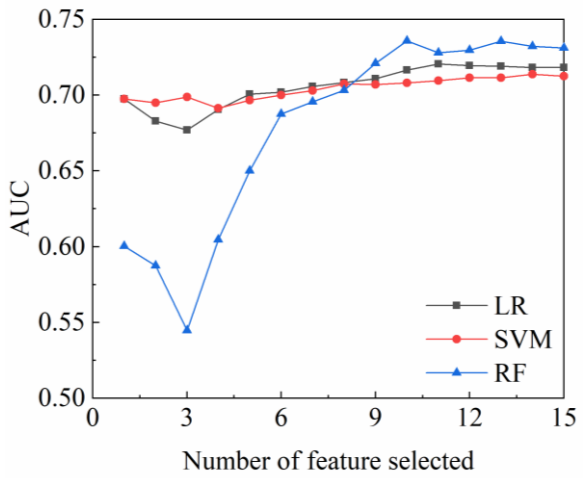

Fig. 6 Variations of AUC values of RFE-based FS-ML models with the number of features selected

position also has relatively strong correlations with the profile curvature (-0.61) and plane curvature (0.62). These results imply that one of the features of micro-geomorphology and slope position should be removed from the feature set. Based on the sum of the absolute values of the correlations between each of the two candidate features and the other features, the slope position is finally not considered in the following machine learning models.

In summary, it is generally found that the plane curvature has little effect on the landslides in the study area and this result is similar to the results of Hong *et al.* (2017). D.F. river, D.F. road and land use are considered to be the most important factors, which may be due to the existence of more landslides in or near populated areas, transport routes, terraces, mining areas, deforestation, wasteland and other areas related to human activities. The results of feature selection tend to select the factors related to the engineering activities, such as land-use type and D.F. road. It is believed that these factors have greater impacts on the occurrence of landslides. Note that since MIC and AIBRF do not remove features, finally only two feature subsets A and B are obtained. Group A is based on VIF, where slope position was deleted because its values were obviously larger than other features (Hong *et al.* 2018b). Group B is based on AABRF, the AABRF value of micro-geomorphology, profile curvature and plane curvature were negative, which need to be removed to avoid a bad impact on the model (Chen *et al.* 2018). These two feature subsets are then used in the following machine learning models for LSA.

4.2.2 LSA results

In this subsection, based on the above-mentioned feature selection results and two feature subsets, five machine learning methods are used to establish 10 landslide susceptibility evaluation models with 10-fold cross-validation. The performance results of the 10 FS-ML models are shown in Table 4. Similar to Table 1, the performance of RF is the best, followed by the GP, LR, SVM, and ANN. Meanwhile, RFE is coupled with LR, SVM and RF, and three extra FS-ML models are established. Fig. 6 plots the variation of AUC score against the number of features for the three models. It is seen that when RF has only a few features, the AUC value is low, but the accuracy is significantly improved after the number of features is increased. Similar observations can also be found on the other two curves, although they are less obvious than the RF model curve. It is also observed that when the number of features is 11, 13, 14, the AUC score of RF, LR and SVM reaches their corresponding maximum, respectively. The optimized AUC for RF, LR and SVM by RFE is 0.744, 0.718, and 0.717, respectively. Compared with LR and SVM, RF has a higher AUC value with fewer

Table 6 Landslide distributions in various susceptibility zones

Model	Class	Area (km^2)	Percentage of domain	Number of landslides	Percentage of landslides	Frequency
LR	Very low	1200.83	8.0%	86	8.5%	0.072
	Low	2689.407	17.9%	68	6.7%	0.025
	Medium	4494.363	29.8%	241	23.7%	0.054
	High	4283.685	28.4%	319	31.4%	0.074
	Very high	2392	15.9%	303	29.8%	0.127
SVM	Very low	1904.117	12.6%	30	2.9%	0.016
	Low	4118.57	27.3%	152	14.9%	0.037
	Medium	4772.588	31.7%	311	30.6%	0.065
	High	2612.681	17.3%	275	27.0%	0.105
	Very high	1652.33	11.0%	249	24.5%	0.151
ANN	Very low	1782.705	11.8%	95	9.3%	0.053
	Low	3226.806	21.4%	106	10.4%	0.033
	Medium	4728.263	31.4%	279	27.4%	0.059
	High	3779.509	25.1%	324	31.9%	0.086
	Very high	1542.983	10.2%	213	20.9%	0.138
GP	Very low	1884.063	12.5%	36	3.5%	0.019
	Low	3565.166	23.7%	143	14.1%	0.040
	Medium	5994.716	39.8%	433	42.6%	0.072
	High	2436.877	16.2%	235	23.1%	0.096
	Very high	1179.465	7.8%	170	16.7%	0.144
RF	Very low	3122.354	20.7%	74	7.3%	0.024
	Low	3852.329	25.6%	154	15.1%	0.040
	Medium	3199.067	21.2%	216	21.2%	0.068
	High	2770.717	18.4%	257	25.2%	0.093
	Very high	2115.781	14.1%	316	31.1%	0.149

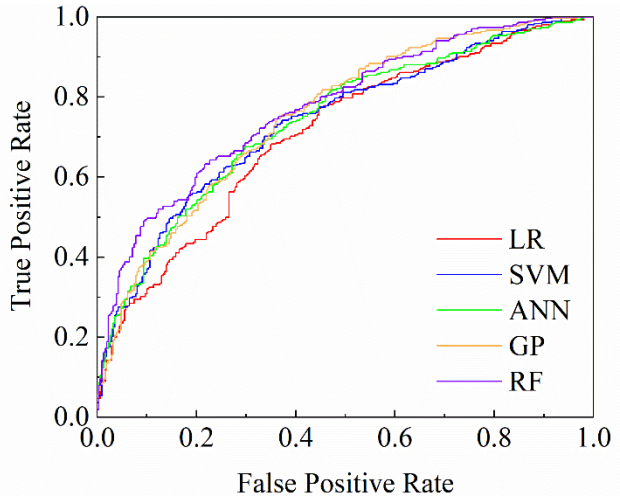

Fig. 7 ROC of five optimal LSA models corresponding to five machine learning models

features. Specifically, micro-geomorphology, slope position, slope aspect and plane curvature are not considered in LR; plane curvature is deleted from the original feature set for SVM; plane curvature and profile curvature are excluded when building the RF model. In all three models, RFE removes plane curvature from the original feature set, which is similar to some filtering feature selection methods, as given in Table 3. For convenience, Table 5 lists the final performance results of the RFE-based FL-ML models. Again, the performance of LR and SVM is similar, whereas RF is much better. The performance of the RF model was improved when decreased a little after feature selection.

4.3 Landslide susceptibility maps for the study area

To produce a relatively accurate and reliable LSM for the study area, this part presents the LSMs of the study area by using the optimal LSA models corresponding to five commonly used machine learning models. It should be noted that since LR, SVM, ANN, GP and RF models are improved after feature selection, the FS-ML corresponding to these models are used for mapping. The five optimal LSA modeling results are shown in Fig. 7: LR (AUC=0.718), SVM (AUC=0.717), ANN (AUC=0.681), GP (AUC=0.719), and RF (AUC=0.744). The overall optimal model is RF with RFE (AUC=0.744). To produce an LSM, a landslide susceptibility index is first generated for all pixels in the study area. Then, each pixel is assigned a unique susceptibility index. The natural breakpoint

method is subsequently used to reclassify these susceptibility indices (Irigaray *et al.* 2007). Based on the reclassification of the landslide susceptibility index, the map with five susceptibility levels of very low, low, medium, high, and very high can be constructed (Chen *et al.* 2017c). The LSMs generated by the five optimal models are shown in Fig. 8. For susceptibility levels of very low, low, medium, high, and very high, the corresponding proportions for different models are as follows: LR (8.0%, 17.9%, 29.8%, 28.4%, 15.9%), SVM (12.6%, 27.3%, 31.7%, 17.3%, 11.0%), ANN (11.8 %, 21.4%, 31.4%, 25.1%, 10.2%), GP (12.5%, 23.7%, 39.8%, 16.2%, 7.8%) and RF (20.7%, 25.6%, 21.2%, 18.4%, 14.1) %). Among all five subfigures, the very low, low, and medium susceptibility regions of the GP model (Fig. 8(d)) account for the largest proportion, indicating that GP is less conservative than the other four models. By contrast, the high and very high ratios of the LR model (Fig. 8(a)) is the largest among all models. Moreover, it can be seen from Fig. 8 that, in general, high susceptibility areas are generally located along roads and river networks, especially areas with heavy rainfall, which are mainly distributed in the northern areas of Xinhua County and Taojiang County, whereas the very low susceptibility areas are mainly concentrated in Taoyuan County Northern region.

To verify the overall method and the effectiveness of the final susceptibility map, the landslide susceptibility map and the landslide inventory map are superimposed to obtain the landslide statistics in the study area. The results are tabulated in Table 6. As shown in Table 6, all frequency values of the five models all increase with the increase of the susceptibility level, indicating that the susceptibility assessment of these five models is more successful and conforms to the objective laws. In the RF landslide susceptibility map, out of 1017 landslides observed, 74 landslides (7.3%) belong to very low susceptibility areas, 154 landslides (15.1%) belong to low susceptibility areas, and 216 landslides (21.2%) belong to the medium susceptibility area, 257 landslides (25.2%) belong to the high susceptibility area, and 356 landslides (31.1%) belong to the very high susceptibility areas. The results showed that 56.3% of the landslides occurred in the high and the very high susceptibility areas, accounting for about 32.5% of the total area. This simple verification method is based on spatial cross-checking of surveying and mapping results, and it is a common indicator of the credibility of landslide susceptibility maps.

5. Discussions

5.1 Analysis of FS-ML models

The current study aims at evaluating the effect of different feature selection methods on different machine learning models to find the most suitable feature selection method for each machine learning model. From the modeling results, feature selection affects to some extent the performance of the model in different ways (Micheletti *et al.* 2014). Selecting the wrong feature selection method may lead to the poor performance of a model. Different machine learning models should use different feature selection methods because different machine learning algorithms have different learning mechanisms and calculation principles (Merghadi *et al.* 2020). The same feature selection method can make some models optimized, but also can make other models work worse in prediction. It is suggested in the future when a machine learning method is used for landslide susceptibility assessment, multiple feature selection methods should be applied to the considered model to select the most appropriate one.

The relationship between landslide and its conditioning factors generally varies from area to area, but they also may have some similarities in different areas. For example, in this area, the contribution of vegetation activity is found relevant to landslides, which is consistent with the conclusions of other studies (Akgun *et al.* 2012). In addition, feature selection results also show the importance of lithology and elevation on landslides susceptibility study which was confirmed by previous studies (Hong *et al.* 2018c, Wang *et al.* 2017).

The model comparison reveals that the RF model generally performs better than the other models, which is consistent with previous studies (Chen *et al.* 2017c, Zhang *et al.* 2017). The reasons are three-fold. First, the RF method is not designed for linear features, and it is an ensemble learning algorithm, which belongs to the bagging type. By combining several weak classifiers, the final results are voted or averaged so that the results of the overall model have high accuracy and generalization performance (Breiman 2001). Second, RF is a non-parametric technique based on a large number of classification trees, which will not cause the risk of overfitting (for example, each tree is a completely independent random experiment), whereas LR is based on the logarithmic transformation of the generalized linear model. Third, RF can handle many independent variables, including both numerical and categorical variables. These variables do not need to be rescaled, transformed, or modified. It can resist outliers in the predictor and automatically handle any missing values.

Overall, the combinations of machine learning methods with different feature subsets have shown well performance. However, the disadvantage is that FS-ML models increase the complexity of the LSA algorithm and the difficulty of feature interpretation. The best AUC value from the proposed FS-ML models in this study is 0.744, which is not as large as those in the literature, although it is acceptable for LSA (Youssef *et al.* 2015). There are two possible reasons: First, the study area is very large and the landslide types are complex (Reichenbach *et al.* 2018); Second, the effect of features may change as the geographical location changes (Yang *et al.* 2019). Nevertheless, the LSA model may be further improved by considering spatial heterogeneity and distinguishing landslide types (Samia *et al.* 2020).

5.2 Analysis of landslide susceptibility map of the study area

From the best landslide susceptibility zoning map, it

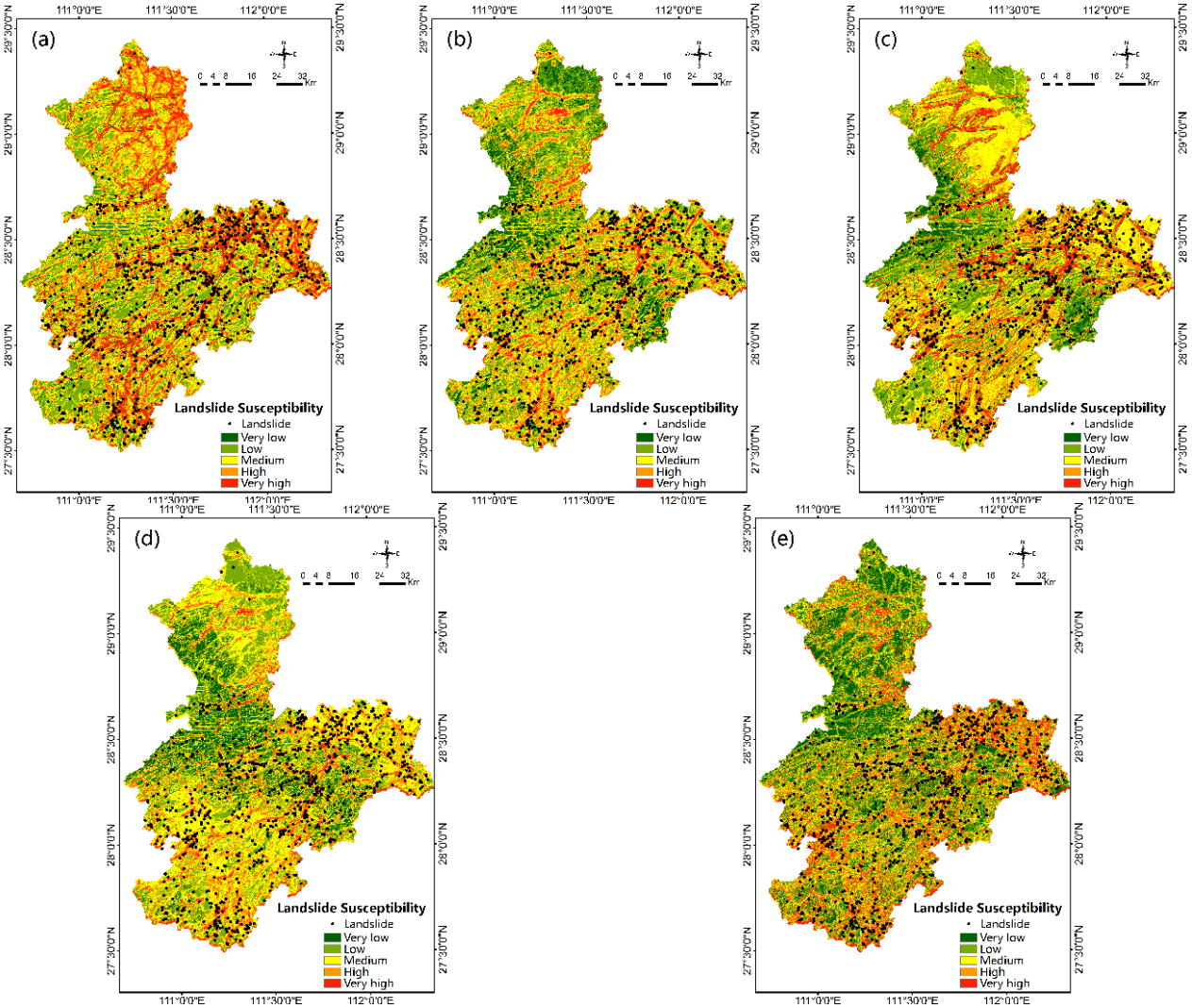

Fig. 8 Landslide susceptibility maps of the study area: (a) Logistic regression, (b) Support vector machine, (c) Artificial neural network, (d) Gaussian process and (e) Random forest

reflects the following two characteristics, indicating that the distribution characteristics of landslide susceptibility in the central Hunan area are consistent with the geographical environment.

First, the very high and high prone areas are mainly distributed along the Zijiang river and its tributaries. The closer to the river, the higher the susceptibility index. It is mainly due to the cutting effect of the water system on the terrain. In the case of rainfall, the slope body is easy to form a water surface and a water path, which increases the surface water of the slope body. At the same time, the NDVI map shows that the vegetation coverage in these areas is low, which indirectly reflects that the human engineering activities in the area are frequent, and the slopes are seriously damaged by humans.

Second, the very low and low-prone areas are mainly distributed in areas close to roads and far away from the artificial surface, mainly because human engineering activities have become a huge external force to change the geological environment. Frequent human activities will greatly accelerate the process of slope deformation and destruction.

However, the LSA implemented in this study did not consider time variation. The obtained landslide susceptibility map is usable for the coming new cases as long as the features of the study area do not have drastic changes (Samia *et al.* 2020).

6. Conclusions

This study has compared systematically 13 different FS-ML models for a more reasonable and reliable LSA in practical engineering. The purpose is twofold, which aims at exploring how feature selection methods affect the performance of machine learning models and finding out which feature selection method is optimal for a given LSA machine learning model. A case study in central Hunan, China has shown that:

- The AUC values of the FS-ML models are generally larger than the original models without feature selection, which means feature selection is important for an accurate LSA.

• A same conditioning factor may contribute differently to different machine learning models, and different feature selection methods have different effects on machine learning models. The optimal feature selection methods for different machine learning models are problem-specific.

• Regarding the effect of the feature selection method on improving machine learning model accuracy, RFE has the most outstanding optimization effect on LR, SVM and RF. AABFR is suitable for GP, while VIF is a better way to optimize ANN. Additionally, considering that RFE-RF model has the highest AUC value and both RFE and RF are not too complicated to implement, the RFE-RF model is considered best for practical LSA in the study area.

• In terms of computational complexity and computational efficiency, RF-based and LR-based FS-ML models are the best, followed by SVM-based model. By contrast, ANN-based and GP-based models perform relatively worse, because their model structures are relatively complex, and the computational efficiency is low.

To conclude, since landslides in different areas may have different characteristics and only one study area was considered in this study, the above conclusions might be case-correlated. Therefore, it is recommended that researchers try different feature selection methods to optimize their own machine learning models to avoid accidental errors in practical LSA.

Acknowledgments

The research described in this paper was financially supported by the National Natural Science Foundation of China (Project No. 41902291), the Natural Science Foundation of Hunan Province, China (Project Nos. 2020JJ5704 and 2020JJ5015), the Hunan Provincial Innovation Foundation for Postgraduate (Project No. CX20200236) and the Fundamental Research Funds for Central South University (Project No. 1053320192194).

References

Akgun, A., Sezer, E.A., Nefeslioglu, H.A., Gokceoglu, C. and Pradhan, B. (2012), "An easy-to-use MATLAB program (MamLand) for the assessment of landslide susceptibility using a Mamdani fuzzy algorithm", *Land Degrad. Develop.*, **38**(1), 23-34. https://doi.org/10.1016/j.cageo.2011.04.012.

Amato, G., Eisank, C., Castro Camilo, D. and Lombardo, L. (2019), "Accounting for covariate distributions in slope-unit-based landslide susceptibility models. A case study in the alpine environment", *Eng. Geol.*, **260**(3), 105237. https://doi.org/10.1016/j.enggeo.2019.105237.

Balzano, B., Tarantino, A., Nicotera, M.V., Forte, G., de Falco, M. and Santo, A. (2019), "Building physically based models for assessing rainfall-induced shallow landslide hazard at catchment scale: Case study of the Sorrento Peninsula (Italy)", *Can. Geotech. J.*, **56**(9), 1291-1303. https://doi.org/10.1139/cgj-2017-0611.

Boulfoul, K., Hammoud, F. and Abbeche, K. (2020), "Numerical study on the optimal position of a pile for stabilization purpose of a slope", *Geomech. Eng.*, **21**(5), 401-411. https://doi.org/10.12989/gae.2020.21.5.401.

Breiman, L. (2001), "Random forests", *Machine Learn.*, **45**(1), 5-32. https://doi.org/10.1023/a:1010933404324.

Bui, D.T., Tuan, T.A., Klempe, H., Pradhan, B. and Revhaug, I. (2016), "Spatial prediction models for shallow landslide hazards: A comparative assessment of the efficacy of support vector machines, artificial neural networks, kernel logistic regression, and logistic model tree", *Landslides*, **13**(2), 361-378. https://doi.org/10.1007/s10346-015-0557-6.

Catani, F., Lagomarsino, D., Segoni, S. and Tofani, V. (2013), "Landslide susceptibility estimation by random forests technique: Sensitivity and scaling issues", *Nat. Hazards Earth Syst. Sci.*, **13**(11), 2815-2831. https://doi.org/10.5194/nhess-13-2815-2013.

Chen, W., Panahi, M. and Pourghasemi, H.R. (2017a), "Performance evaluation of GIS-based new ensemble data mining techniques of adaptive neuro-fuzzy inference system (ANFIS) with genetic algorithm (GA), differential evolution (DE), and particle swarm optimization (PSO) for landslide spatial modelling", *Catena*, **157**, 310-324. https://doi.org/10.1016/j.catena.2017.05.034.

Chen, W., Pourghasemi, H.R. and Zhao, Z. (2017b), "A GIS-based comparative study of Dempster-Shafer, logistic regression and artificial neural network models for landslide susceptibility mapping", *Geocarto Int.*, **32**(4), 367-385. https://doi.org/10.1080/10106049.2016.1140824.

Chen, W., Xie, X., Peng, J., Shahabi, H., Hong, H., Bui, D.T., Duan, Z., Li, S. and Zhu, A.X. (2018), "GIS-based landslide susceptibility evaluation using a novel hybrid integration approach of bivariate statistical based random forest method", *Catena*, **164**, 135-149. https://doi.org/10.1016/j.catena.2018.01.012.

Chen, W., Xie, X., Wang, J., Pradhan, B., Hong, H., Bui, D.T., Duan, Z. and Ma, J. (2017c), "A comparative study of logistic model tree, random forest, and classification and regression tree models for spatial prediction of landslide susceptibility", *Catena*, **151**, 147-160. https://doi.org/10.1016/j.catena.2016.11.032.

Cheng, W.C., Ni, J.C., Arulrajah, A. and Huang, H.W. (2018), "A simple approach for characterising tunnel bore conditions based upon pipe jacking data", *Tunn. Undergr. Sp. Tech.*, **71**, 494-504. https://doi.org/10.1016/j.tust.2017.10.002.

Cortes, C. and Vapnik, V. (1995), "Support-vector networks", *Machine Learning*, **20**(3), 273-297. https://doi.org/10.1007/BF00994018.

Degraff, J.V. and Canuti, P. (1988), "Using isopleth mapping to evaluate landslide activity in relation to agricultural practices", *B. Eng. Geol. Environ.*, **38**(1), 61-71.

Guyon, I., Weston, J., Barnhill, S. and Vapnik, V. (2002), "Gene selection for cancer classification using support vector machines", *Machine Learning*, **46**(1-3), 389-422. https://doi.org/10.1023/A:1012487302797.

Hong, H., Ilia, I., Tsangaratos, P., Chen, W. and Xu, C. (2017), "A hybrid fuzzy weight of evidence method in landslide susceptibility analysis on the Wuyuan area, China", *Geomorphology*, **290**, 1-16. https://doi.org/10.1016/j.geomorph.2017.04.002.

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

Hong, H., Pourghasemi, H.R. and Pourtaghi, Z.S. (2016), "Landslide susceptibility assessment in Lianhua County (China): A comparison between a random forest data mining technique and bivariate and multivariate statistical models",

Geomorphology, **259**, 105-118. https://doi.org/10.1016/j.geomorph.2016.02.012.

Hong, H., Pradhan, B., Sameen, M.I., Kalantar, B., Zhu, A. and Chen, W. (2018b), "Improving the accuracy of landslide susceptibility model using a novel region-partitioning approach", *Landslides*, **15**(4), 753-772. https://doi.org/10.1007/s10346-017-0906-8.

Hong, H., Tsangaratos, P., Ilia, I., Liu, J., Zhu, A.X. and Chen, W. (2018c), "Application of fuzzy weight of evidence and data mining techniques in construction of flood susceptibility map of Poyang County, China", *Sci. Total Environ.*, **625**, 575-588. https://doi.org/10.1016/j.scitotenv.2017.12.256.

Irigaray, C., Fernández, T., El Hamdouni, R. and Chacón, J. (2007), "Evaluation and validation of landslide-susceptibility maps obtained by a GIS matrix method: Examples from the Betic Cordillera (southern Spain)", *Nat. Hazards*, **41**(1), 61-79. https://doi.org/10.1007/s11069-006-9027-8.

Kavzoglu, T. and Mather, P.M. (2010), "The role of feature selection in artificial neural network applications", *Int. J. Remote Sensing*, **23**(15), 2919-2937. https://doi.org/10.1080/01431160110107743.

Lagomarsino, D., Tofani, V., Segoni, S., Catani, F. and Casagli, N. (2017), "A tool for classification and regression Using random forest methodology: Applications to landslide susceptibility mapping and soil thickness modeling", *Environ. Model. Asses.*, **22**(3), 201-214. https://doi.org/10.1007/s10666-016-9538-y.

Li, C., Yao, D., Wang, Z., Liu, C.C., Wuliji, N., Yang, L., Li, L. and Amini, F. (2016), "Model test on rainfall-induced loess-mudstone interfacial landslides in Qingshuihe, China", *Environ. Earth Sci.*, **75**(9), 835. https://doi.org/10.1007/s12665-016-5658-6.

Liu, D. and Chen, X. (2015), "Shearing characteristics of slip zone soils and strain localization analysis of a landslide", *Geomech. Eng.*, **8**(1), 33-52. https://doi.org/10.12989/gae.2015.8.1.033.

Liu, L.L., Cheng, Y.M., Pan, Q.J. and Dias, D. (2020), "Incorporating stratigraphic boundary uncertainty into reliability analysis of slopes in spatially variable soils using one-dimensional conditional Markov chain model", *Comput. Geotech.*, **118**, 103321. https://doi.org/10.1016/j.compgeo.2019.103321.

Liu, L.L., Deng, Z.P., Zhang, S.H. and Cheng, Y.M. (2018), "Simplified framework for system reliability analysis of slopes in spatially variable soils", *Eng. Geol.*, **239**, 330-343. https://doi.org/10.1016/j.enggeo.2018.04.009.

Lombardi, M., Cardarilli, M. and Raspa, G. (2017), "Spatial variability analysis of soil strength to slope stability assessment", *Geomech. Eng.*, **12**(3), 483-503. https://doi.org/10.12989/gae.2017.12.3.483.

Lombardo, L. and Mai, P.M. (2018), "Presenting logistic regression-based landslide susceptibility results", *Eng. Geol.*, **244**, 14-24. https://doi.org/10.1016/j.enggeo.2018.07.019.

Merghadi, A., Yunus, A.P., Dou, J., Whiteley, J., ThaiPham, B., Bui, D.T., Avtar, R. and Abderrahmane, B. (2020), "Machine learning methods for landslide susceptibility studies: A comparative overview of algorithm performance", *Earth-Sci. Rev.*, **207**, 103225. https://doi.org/10.1016/j.earscirev.2020.103225.

Micheletti, N., Foresti, L., Robert, S., Leuenberger, M., Pedrazzini, A., Jaboyedoff, M. and Kanevski, M. (2014), "Machine learning feature selection methods for landslide susceptibility mapping", *Math. Geosci.*, **46**(1), 33-57. https://doi.org/10.1007/s11004-013-9511-0.

Moore, I., Grayson, R. and Ladson, T. (1991), "Digital Terrain Modeling: A review of hydrological, geomorphological, and biological applications", *Hydrol. Process.*, **5**, 3-30. https://doi.org/10.1002/hyp.3360050103.

Paola, R., Galli, M., Cardinali, M., Guzzetti, F. and Ardizzone, F., (2004), *Geomorphological Mapping to Assess Landslide Risk: Concepts*, *Methods and Applications in the Umbria Region of Central Italy*, in *Landslide Hazard and Risk*, Hoboken, New Jersey, U.S.A.

Pham, B.T., Avand, M., Janizadeh, S., Phong, T.V., Al-Ansari, N., Ho, L.S., Das, S., Le, H.V., Amini, A., Bozchaloei, S.K., Jafari, F. and Prakash, I. (2020), "GIS based hybrid computational approaches for flash flood susceptibility assessment", *Water*, **12**(3), 683. https://doi.org/10.3390/w12030683.

Pourghasemi, H.R. and Rahmati, O. (2018), "Prediction of the landslide susceptibility: Which algorithm, which precision?", *Catena*, **162**, 177-192. https://doi.org/10.1016/j.catena.2017.11.022.

Pourghasemi, H.R., Kornejady, A., Kerle, N. and Shabani, F. (2020), "Investigating the effects of different landslide positioning techniques, landslide partitioning approaches, and presence-absence balances on landslide susceptibility mapping", *Catena*, **187**, 104364. https://doi.org/10.1016/j.catena.2019.104364.

Pourghasemi, H.R., Pradhan, B., Gokceoglu, C., Mohammadi, M. and Moradi, H.R. (2013), "Application of weights-of-evidence and certainty factor models and their comparison in landslide susceptibility mapping at Haraz watershed, Iran", *Arab. J. Geosci.*, **6**(7), 2351-2365. https://doi.org/10.1007/s12517-012-0532-7.

Rasmussen, C.E. and Nickisch, H. (2010), "Gaussian processes for machine learning (GPML) toolbox", *J. Mach. Learn. Res.*, **11**(6), 3011-3015. https://doi.org/10.1115/1.4002474.

Reichenbach, P., Rossi, M., Malamud, B.D., Mihir, M. and Guzzetti, F. (2018), "A review of statistically-based landslide susceptibility models", *Earth-Sci. Rev.*, **180**, 60-91. https://doi.org/10.1016/j.earscirev.2018.03.001.

Reif, D.M., Motsinger, A.A., Mckinney, B.A., Jr, J.E.C. and Moore, J.H. (2006), "Feature selection using a random forests classifier for the integrated analysis of multiple data types" *Proceedings of the IEEE Symposium on Computational Intelligence in Bioinformatics & Computational Biology*, Toronto, Canada, September.

Reshef, D.N., Reshef, Y.A., Finucane, H.K., Grossman, S.R., McVean, G., Turnbaugh, P.J., Lander, E.S., Mitzenmacher, M. and Sabeti, P.C. (2011), "Detecting novel associations in large data sets", *Science*, **334**(6062), 1518-1524. https://doi.org/10.1126/science.1205438.

Samia, J., Temme, A., Bregt, A., Wallinga, J., Guzzetti, F. and Ardizzone, F. (2020), "Dynamic path-dependent landslide susceptibility modelling", *Nat. Hazards Earth Syst. Sci.*, **20**(1), 271-285. https://doi.org/10.5194/nhess-20-271-2020.

Sheil, B.B., Suryasentana, S.K. and Cheng, W.C. (2020), "Assessment of anomaly detection methods applied to microtunneling", *J. Geotech. Geoenviron. Eng.*, **146**(9), 04020094. https://doi.org/10.1061/(ASCE)GT.1943-5606.0002326.

Shou, K.J. and Lin, J.F. (2020), "Evaluation of the extreme rainfall predictions and their impact on landslide susceptibility in a sub-catchment scale", *Eng. Geol.*, **265**, 105434. https://doi.org/10.1016/j.enggeo.2019.105434.

Skolidis, G. and Sanguinetti, G. (2011), "Bayesian nultitask classification with Gaussian process priors", *IEEE T. Neur. Networks*, **22**(12), 2011-2021. https://doi.org/10.1109/tnn.2011.2168568.

Sun, D., Wen, H., Wang, D. and Xu, J. (2020), "A random forest model of landslide susceptibility mapping based on hyperparameter optimization using Bayes algorithm", *Geomorphology*, **362**, 107201. https://doi.org/10.1016/j.geomorph.2020.107201.

Vasu, N.N. and Lee, S.R. (2016), "A hybrid feature selection algorithm integrating an extreme learning machine for landslide

susceptibility modeling of Mt. Woomyeon, South Korea", *Geomorphology*, **263**, 50-70. https://doi.org/10.1016/j.geomorph.2016.03.023.

Wang, F., Xu, P., Wang, C., Wang, N. and Jiang, N. (2017), "Application of a GIS-based slope unit method for landslide susceptibility mapping along the Longzi river, Southeastern Tibetan Plateau, China", *ISPRS Int. J. Geo-Inform.*, **6**(6), 172. https://doi.org/10.3390/ijgi6060172.

Wang, L.J., Guo, M., Sawada, K., Lin, J. and Zhang, J. (2015), "Landslide susceptibility mapping in Mizunami City, Japan: A comparison between logistic regression, bivariate statistical analysis and multivariate adaptive regression spline models", *Catena*, **135**, 271-282. https://doi.org/10.1016/j.catena.2015.08.007.

Weiss, A. (2001), "Topographic position and landforms analysis", *Proceedings of the ESRI User Conference*, San Diego, California, U.S.A., July.

Wold, S., Esbensen, K. and Geladi, P. (1987), "Principal component analysis", *Chemometr. Intell. Lab.*, **2**(1-3), 37-52. https://doi.org/10.1016/0169-7439(87)80084-9.

Xing, H., Liu, L. and Luo, Y. (2019), "Water-induced changes in mechanical parameters of soil-rock mixture and their effect on talus slope stability", *Geomech. Eng.*, **18**(4), 353-362. https://doi.org/10.12989/gae.2019.18.4.353.

Yalcin, A., Reis, S., Aydinoglu, A.C. and Yomralioglu, T. (2011), "A GIS-based comparative study of frequency ratio, analytical hierarchy process, bivariate statistics and logistics regression methods for landslide susceptibility mapping in Trabzon, NE Turkey", *Catena*, **85**(3), 274-287. https://doi.org/https://doi.org/10.1016/j.catena.2011.01.014.

Yang, Y., Yang, J., Xu, C., Xu, C. and Song, C. (2019), "Local-scale landslide susceptibility mapping using the B-GeoSVC model", *Landslides*, **16**(7), 1301-1312. https://doi.org/10.1007/s10346-019-01174-y.

Youssef, A.M., Al-Kathery, M. and Pradhan, B. (2015), "Landslide susceptibility mapping at Al-Hasher area, Jizan (Saudi Arabia) using GIS-based frequency ratio and index of entropy models", *Geosci. J.*, **19**(1), 113-134. https://doi.org/10.1007/s12303-014-0032-8.

Zhang, K., Wu, X., Niu, R., Yang, K. and Zhao, L. (2017), "The assessment of landslide susceptibility mapping using random forest and decision tree methods in the Three Gorges Reservoir area, China", *Environ. Earth Sci.*, **76**(11), 405. https://doi.org/10.1007/s12665-017-6731-5.

IC