



Landslide susceptibility mapping using an ensemble model of Bagging scheme and random subspace–based naïve Bayes tree in Zigui County of the Three Gorges Reservoir Area, China

Xudong Hu¹ · Cheng Huang² · Hongbo Mei¹ · Han Zhang¹

Received: 23 December 2020 / Accepted: 5 May 2021 / Published online: 13 May 2021
© Springer-Verlag GmbH Germany, part of Springer Nature 2021

Abstract

A novel machine learning ensemble model that is a hybridization of Bagging and random subspace–based naïve Bayes tree (RSNBtree), named as BRSNBtree, was used to prepare a landslide susceptibility map for Zigui County of the Three Gorges Reservoir Area, China. The proposed method is implemented by using the Bagging scheme to integrate the base-level RSNBtree model. To predict landslide susceptibility for the study area, a spatial database consisted of 807 landslides and 11 conditioning factors has been prepared. Evaluation of conditioning factors was conducted using the Pearson correlation coefficient and Relief-F method. The results indicate that all factors except the topographic wetness index can be accepted as modeling inputs. Particularly, the distance to rivers is the most important factor in landslide susceptibility prediction. The performance of landslide models was evaluated using statistical indices and areas under the receiver operating characteristic curve (AUC). The support vector machines (SVM) and random forest (RF) were adopted for the comparison with our methods. Results show that the BRSNBtree (AUC = 0.968) achieves the highest prediction performance, which successfully refines the RSNBtree (AUC = 0.938) and outperforms the RF (AUC = 0.949) and SVM (AUC = 0.895). Therefore, the proposed BRSNBtree presents advantages in targeting landslide susceptible areas and provides a promising method for landslide susceptibility assessment. The developed susceptibility maps could facilitate effective landslide risk management for this landslide-prone area.

Keywords Landslides · Bagging · Random subspace · Naïve Bayes tree · Zigui

Introduction

Landslides are a common type of natural disaster that causes irrecoverable losses of life and enormous damages throughout the world each year. Many developing countries including China

are greatly affected by landslides due to their limited resources and particular geographical conditions (Alcántara-Ayala 2002; Ma and Xu 2019). Particularly, landslides and other geo-hazards frequently occur in the geologically complex Three Gorges Reservoir Area along the Yangtze River, China; the region is especially prone to slope failure during the rainy season (Zhang et al. 2017). For example, in July 2–4, 1996, a heavy rainfall event that delivered an average of 457.9 mm, which triggered 237 landslides in Zigui County of the Three Gorges Reservoir Area, caused 11 fatalities and economic losses of 80 million RMB, and threatened 2.86 million lives. Moreover, there is a growing tendency for landslide frequency with time in this region owing to the effect of stream erosion, engineering construction, and increasing land demand (Tang et al. 2019). This calls for exploring and applying effective measures to target landslide-prone areas (Criss et al. 2020). Currently, landslide susceptibility assessment based on “susceptibility index” calculations or susceptibility maps has been widely adopted as an efficient tool to predict and prevent landslides (Li et al. 2019).

✉ Hongbo Mei
hbmei@cug.edu.cn

Xudong Hu
xudonghu@cug.edu.cn

Cheng Huang
hc9977@qq.com

Han Zhang
wsedrc@cug.edu.cn

¹ School of Earth Resources, China University of Geosciences, Wuhan 430074, China

² Yunnan Geological Environmental Monitoring Institute, Kunming 650000, China

Over the last few years, the fast development of remote sensing and geographic information system techniques has greatly facilitated the conduction of landslide susceptibility assessment (Tehrany et al. 2013). The quality of any given assessment depends mainly on the data and algorithm used in the modeling process. Although numerous methods have been proposed for landslide susceptibility assessment, reviews on previous landslide studies show that traditional methods such as knowledge-driven methods and statistical methods are being gradually replaced by machine learning methods (Pham et al. 2016). Machine learning methods are highly recommended because this method is good at dealing with non-linear real-world problems, such as gully erosion assessment (Pourghasemi et al. 2020), flood frequency analysis (Pezhman et al. 2020), landslide detection (Ghorbanzadeh et al. 2019), and landslide modeling (Liu et al. 2021). Relevant work includes logistic regression (LR) (Riegel et al. 2020), maximum entropy (Pandey et al. 2020), classification and regression tree (CART) (Ghasemian et al. 2020), reduced error pruning tree (REPT) (Pham et al. 2019), Bayesian network (Lee et al. 2020), and naïve Bayes (NB) (Pham et al. 2015). In addition, black-box models such as support vector machines (SVM) (Nhu et al. 2020b) and artificial neural networks (ANN) (Lucchese et al. 2021) are likely to produce state-of-the-art modeling results and have been widely employed for landslide susceptibility assessment. Yet, to date, the prediction capability of machine learning methods for landslide modeling still suffers from limitations due to complex landslide instability mechanisms and dynamic mechanical properties (Merghadi et al. 2018; Yao et al. 2020). Additionally, the fitting function of a single machine learning algorithm is based on a sole hypothesis, which may not match the true hypothesis of the problems concerned (Rokach 2009). To put it in another way, the performance of a machine learning algorithm is largely determined by the property of a specific problem. Machine learning methods must be constantly evaluated as the case alters in order to derive reliable conclusions for landslide modeling.

Ensemble methods have recently received much attention due to their capability of improving the predictive accuracy of landslide models (Tien Bui et al. 2016; Pham and Prakash 2017; Chen et al. 2018a, 2018b; Hu et al. 2020, 2021; Razavi-Termeh et al. 2021). The ensemble method adopts a certain ensemble strategy to combine homogeneous or heterogeneous component algorithms. This method can expand the hypothesis fitting function and effectively reduce learning errors against the single algorithms, and therefore exhibits excellent ability in predicting landslide susceptibility. Multiple types of integrated models have been investigated and employed in landslide studies, with the Bagging (Breiman 1996) and Boosting (Freund and Schapire 1997), random subspace (RS) (Ho 1998), and random forest (RF) (Breiman 2001) being the most popular ensemble learning

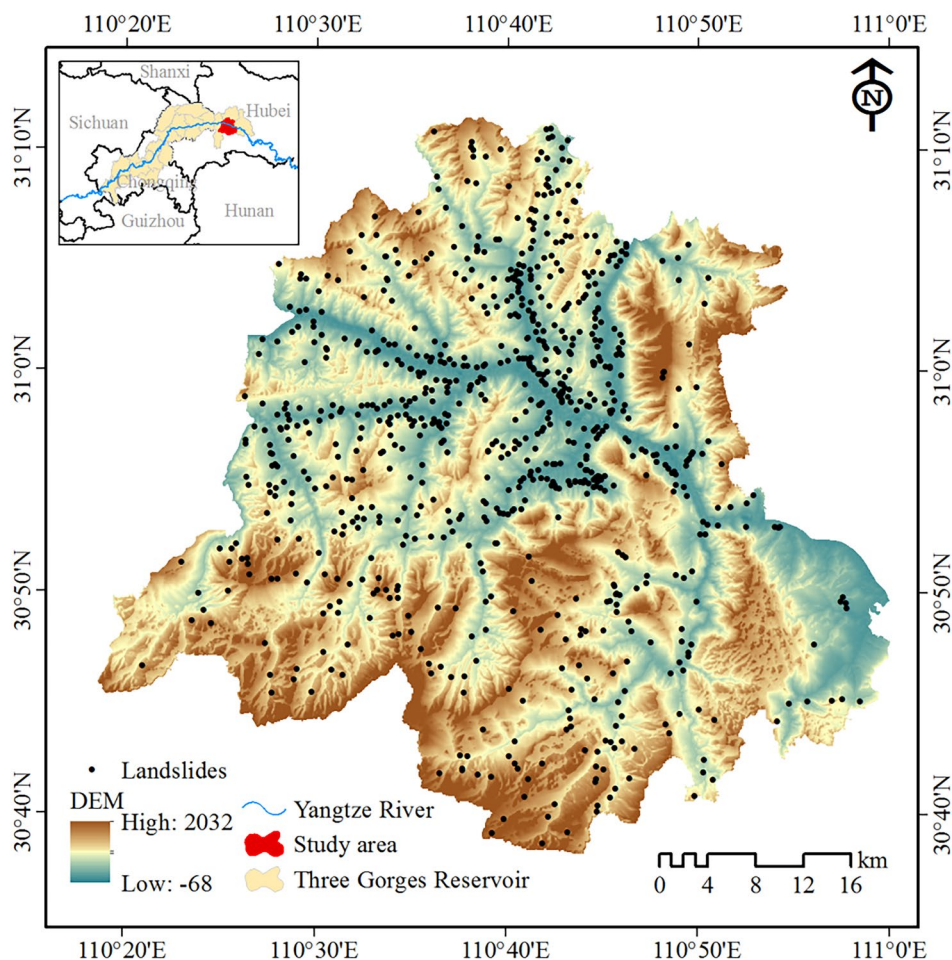
techniques. In particular, RS is a simple and useful ensemble scheme that produces subsets from original data to train and combine base-learners. Pham et al. (2020) have investigated the effect of RS over different decision trees (DT) such as the best first DT, functional tree, J48 DT, naïve Bayes Tree (NBtree), and REPT. Their results show that RS is capable to refine these single algorithms in relation to landslide modeling. More importantly, Shirzadi et al. (2017) reported that RS-based NBtree (RSNBtree) is an efficient method and significantly increases the prediction performance of the landslide model. However, the RSNBtree is still rarely applied in landslide susceptibility prediction and needs to be further investigated. Additionally, Bagging is one of the most earliest and well-known ensemble techniques. Bagging-based ensemble methods (e.g., J48 DT, ANN, SVM, logistic model tree, alternating decision tree, RF) have been widely developed and achieved impressive results in landslide studies (Pham et al. 2017; Hong et al. 2018; Truong et al. 2018; Dou et al. 2019; Nhu et al. 2020a; Wu et al. 2020). Therefore, it is feasible to explore improved RSNBtree using the Bagging ensemble in landslide prediction.

In this paper, a novel machine learning ensemble method for modeling landslide susceptibility is proposed. This method is noted as the BRSNBtree, aimed to refine the RSNBtree via a Bagging scheme. We firstly implement a RSNBtree model based on the RS technique and the NBtree model. Then, the RSNBtree is treated as a component of the Bagging framework to perform ensemble prediction. We investigate the potential of the introduced method for real-world landslide spatial prediction in Zigui County of the Three Gorges Reservoir Area, China, and used it to prepare a landslide susceptibility map. The main difference between the present study and the formerly revealed studies is that two types of ensemble techniques are jointly used to create the composite machine learning model for landslide spatial prediction. Moreover, landslide modeling using ensemble learning was rarely examined in Zigui County. Developing and applying improved landslide models is needed to conduct an accurate and reliable landslide susceptibility assessment for this landslide-prone region. The results from this study should facilitate effective landslide risk management for Zigui area.

Study area

The study area is located in the east of the Three Gorges Reservoir Area in Hubei Province, China (Fig. 1), covering an area of 2273 km². Its latitude and longitude range from 30°38' to 31°11' N, and 110° 18' to 111° 00' E, respectively. Topographically, the area is separated from north to south by the Yangtze River, with the elevation increasing from northeast to southwest. The region is characterized by a subtropical monsoon climate, with an annual temperature from 17 to 19 °C (Li et al. 2019). The average

Fig. 1 Location of the study area and landslide inventory



annual precipitation of the area stands at 1490 mm, and heavy rainfall usually occurs from June to September and accounts for approximately 70% of the total annual rainfall. The max daily precipitation can reach 358 mm in the rain season. The stream system mainly consists of the Yangtze River and its tributaries, and the stability of slopes suffers from the effects of river scouring and erosion. Geologically, strata from Silurian to Quaternary crop out across the area. Triassic limestone and Jurassic sandstone and mudstone are extensively distributed, especially in the area along the river. Few faults appear in the study area, and the principal structure is represented by the Zigui syncline.

Data preparation

Data source

In the present study, the data used for landslide susceptibility modeling are acquired from available sources, including field survey, a digital elevation model (DEM) with a resolution of 30 m, Landsat 8 OLI remote sensing

images with a resolution of 30 m, the Second Detailed Land Investigation Nationwide (SDLIN), precipitation station, and a brief engineering rock group (ERG) map at a scale of 1:650,000. Historical landslide records and the ERG map were collected from field surveys supported by the Wuhan Center of Geological Survey. DEM data and remote sensing imageries were available on the website of <http://www.gscloud.cn/>. SDLIN data was provided by the Department of natural resources of Hubei Province. All data were converted into raster format with a pixel size of 30 m (consistent with the resolution of the DEM data).

Spatial database

Preparation of landslide inventory maps is the pre-requisite of the landslide susceptibility prediction. Landslide inventory maps depict the spatial distribution pattern of past landslides and can boost our knowledge of the relationship between landslide occurrence and landslide-related factors (Tsangaratos and Ilia 2016). In Zigui County, a total of 807 landslide locations have been recorded through the field

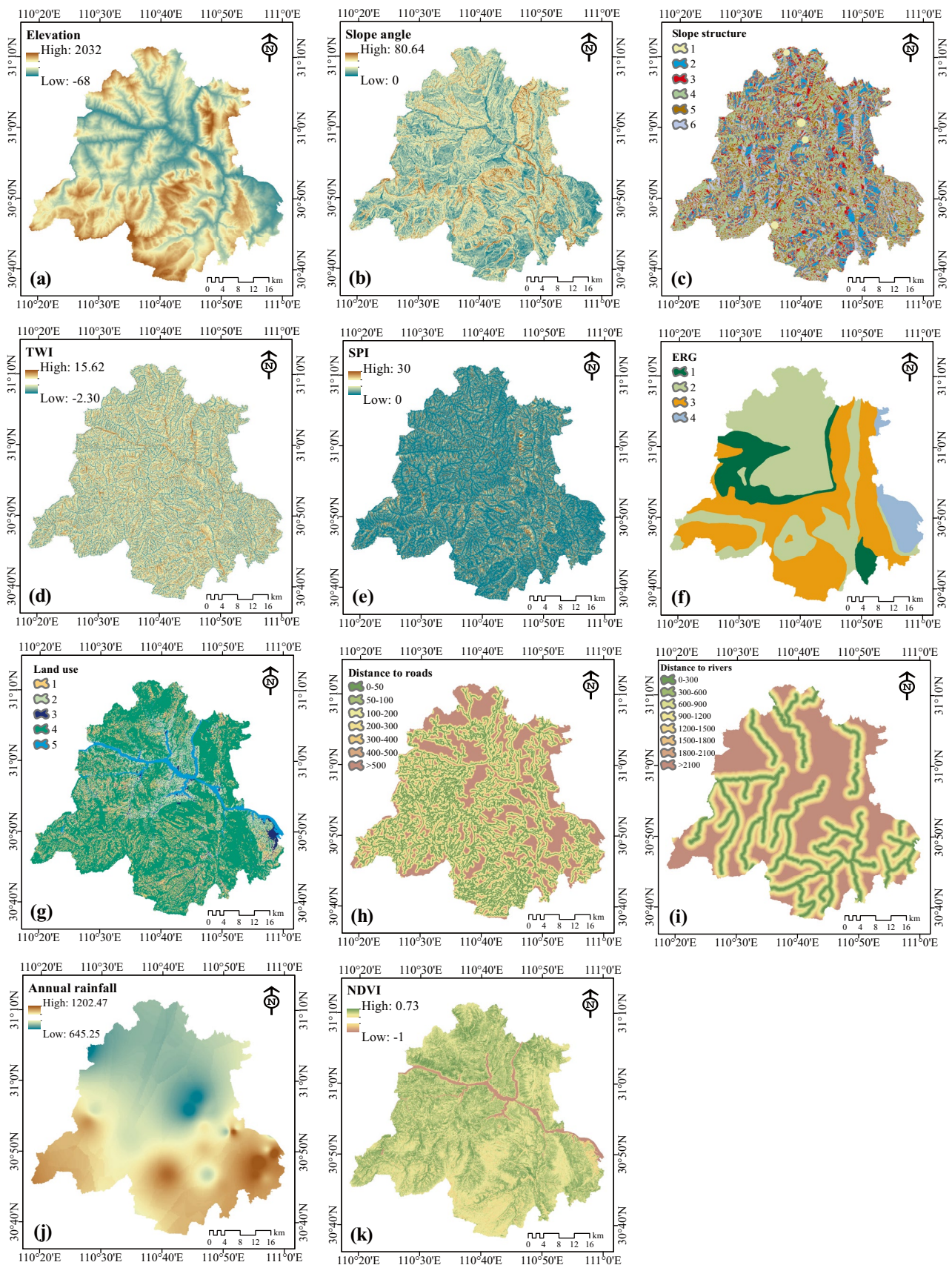


Fig. 2 Landslide conditioning factors. (a) Elevation. (b) Slope angle. (c) Slope structure. (d) TWI. (e) SPI. (f) ERG. (g) Land use. (h) Distance to roads. (i) Distance to rivers. (j) Annual rainfall. (k) NDVI

surveys, with areas and volumes ranging from 1.0×10^3 to 2.9×10^6 m² and 5.4×10^3 to 2.6×10^8 m³. Landslide analysis reveals that the small-size landslide ($< 10 \times 10^4$ m³), medium-size landslide (10×10^4 m³– 100×10^4 m³), large-size landslide (100×10^4 m³– 1000×10^4 m³), and giant-size landslide ($> 1000 \times 10^4$ m³), respectively, account for 19.7%, 45.6%, 29.4%, and 7.3% of the total landslides. The dominant types are loess landslides and loess-bedrock landslides nearby riverbanks. In the case where landslide susceptibility mapping is performed on large scale, landslide inventory is usually prepared in point data given the advantages in improving mapping efficiency, avoiding uncertainty in depicting landslide boundaries, reducing spatial autocorrelation between landslide samples, and providing equal treatment to landslides with different sizes (Petschko et al. 2014; Goetz et al. 2015; Rahali 2017; Chen et al. 2018a). Therefore, the centroids of those landslide locations were imported into ArcGIS 10.2 to generate a landslide inventory map for the study area (Fig. 1). Landslide spatial prediction based on the machine learning technique is a binary classification problem. Therefore, the landslide occurrence can be treated as the target variable, and “1” and “0” are respectively assigned for landslide units and the non-landslide units (Bennett et al. 2016). All landslides were randomly separated into two parts, with 565 landslides (70%) for model training and 242 landslides (30%) for model validation. As for non-landslide units, they were randomly generated using the “Create Random Points” tool in the ArcGIS 10.2.

Previous researches have revealed that the landslide occurrence is related to environmental features as well as anthropogenic activities (Pourghasemi and Kerle 2016). Considering the data availability, the topographical, geological, and climatic conditions of the study area, eleven landslide-related factors including elevation, slope angle, slope structure, topographic wetness index (TWI), stream power index (SPI), engineering rock group (ERG), land use, distance to roads, distance to rivers, annual rainfall, and normalized difference vegetation index (NDVI) were selected for modeling landslide susceptibility for the study area.

Specifically, five topographical factors such as elevation, slope angle, slope structure, TWI, and SPI (Fig. 2a–e) were extracted from the DEM data. The elevation is one of the most commonly used factors in landslide susceptibility studies, which is closely linked to human activities, rainfall, vegetation cover, climate, and other conditions, and has an indirect effect on landslide occurrence (Hong et al. 2016). The slope angle impacts the stability of the slope as it controls both the shear force and the water velocity on the slope

(Fernández and Lutz 2010). The increasing slope angle will promote slope instability. The slope structure is a spatial combination concerning the slope aspect and strata tendency, which determines the type and degree of slope deformation. The slope structure is regarded as an essential factor for studying the landslide distribution and development in the Three Gorges Reservoir Area. The slope structure classes of the study area include (1) horizontal slope, (2) consequent slope, (3) consequent-diagonal slope, (4) transverse slope, (5) reverse-diagonal slope, and (6) reverse slope. The TWI is an important factor for modeling landslide susceptibility and can be used to evaluate the conditions of soil and runoff volume. The SPI indicates the erosion power of a stream; the slope failure is particularly likely to occur in the area where stream erosion power is remarkable.

A strong correlation can be found between the geological lithology and the landslide vulnerability (Juliev et al. 2019). Different lithological units have different physical hardness, interlayer structures, and weathering resistance, which affects landslide occurrence possibility. In this study, the ERG was used as the geological factor to analyze the landslide susceptibility, which includes four categories: (1) group of the loose rock and soil, (2) group of layered clastic, (3) group of layered carbonatite, and (4) group of massive crystalline rocks (Fig. 2f).

Based on the SDLIN data, three essential factors such as the land use (Fig. 2g), distance to roads (Fig. 2h), and distance to rivers (Fig. 2i) were prepared. Land-use category is suggested to be an important indicator in the detection of landslides (Nsengiyumva et al. 2019). Through in situ surveys, five land-use types were identified for the study area: (1) farmland, (2) orchard, (3) residential area, (4) forest, and (5) water. Detailed descriptions for these land-use categories could be found in He et al. (2008) and Chen and Wang (2010). Road construction as a typical anthropogenic intervention feature usually causes increased strain behind the slope and decreased slope toe support (Regmi et al. 2014). The runoff of rivers directly reflects water erosion power to slopes and is associated with the slope failure (Preuth et al. 2010). In this study, distance to roads and distance to rivers were respectively used to examine the impact of road construction and stream network on the landslide occurrence.

The role of the precipitation can never be neglected in landslide prediction (Grelle et al. 2013). The precipitation data were obtained from the precipitation station in Zigui County, and then the annual rainfall map was constructed by using the Kriging interpolation in ArcGIS software (Fig. 2j).

Vegetation affects the stability of slope materials by controlling the root function, rainfall infiltration, and soil erosion (Jia et al. 2014). NDVI is a commonly used index to indicate the vegetation coverage. The NDVI used in this study was extracted from Landsat 8 OLI imagery (Fig. 2k).

Methodology

The flowchart of landslide susceptibility mapping using the introduced ensemble model is displayed in Fig. 3, mainly including four steps namely the construction of spatial database, landslide modeling, model evaluation, and development of landslide susceptibility maps. ArcGIS 10.2 and ENVI 5.3 associated with IBM SPSS Statistics 22 were used for data processing, while the landslide modeling was conducted in R 3.5.3 software.

Evaluation of landslide conditioning factors

It is essential to evaluate landslide conditioning factors for landslide spatial prediction. In this study, the evaluation program of landslide conditioning factors is involved in importance level analysis and correlation analysis (Wu et al. 2020). The importance level of factors affects the quality of landslide modeling because factors with noise or negative predictive capability may destroy the modeling results. Additionally, highly correlated factors are also not beneficial to model’s performance. In this study, the Relief-F (ReF) algorithm, known as an effective feature selection method, was adopted to measure the importance level of the eleven landslide conditioning factors. Additionally, the correlation between these factors was evaluated using the Pearson correlation coefficient (PCC)

The ReF was first introduced by Kira and Rendell (Kira and Rendell 1992). The ReF measures the quality of variables by distinguishing between instances from different classes. For a certain landslide sample R , its k -nearest neighbor landslide and non-landslide samples are respectively noted as H and M .

Given the number of iteration m , the ReF updates the weights of each variable as following:

$$W(A) = W(A) - \frac{\sum diff(A, R, H_j)}{mk} + \frac{\sum \frac{p(C)}{1-p(Class(R))} \sum diff(A, R, M_j(C))}{mk} \tag{1}$$

where $p(C)$ is the possibility of a class, $p(Class(R))$ is the possibility of a class within samples, $diff$ indicates the difference between samples in a variable A , and $M_j(C)$ notes the j th nearest neighbor sample.

The PCC is used to indicate a negative or positive correlation between variables, defined as below:

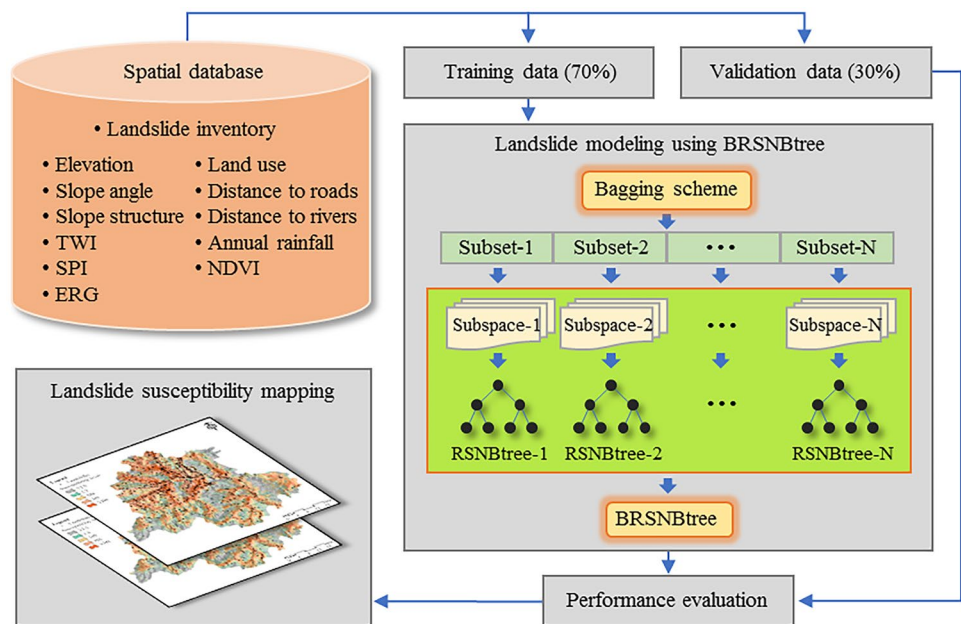
$$PCC_{ij} = \frac{cov(v_i, v_j)}{\sqrt{var(v_i) \times var(v_j)}} \tag{2}$$

where cov and var denote the covariance and variance of variables, respectively.

Landslide modeling methods

In this study, a novel ensemble model, the BRSNBtree, was proposed for landslide susceptibility mapping. Inspired by typical ensemble models, our method is comprised of a base-learner and ensemble strategy. The base-learner used in this study is served by a RSNBtree model. As for the ensemble strategy, the Bagging was adopted as the meta-learner because of its efficiency in ensemble prediction. The structure of the BRSNBtree has been displayed in Fig. 3.

Fig. 3 Flowchart of landslide susceptibility mapping using the BRSNBtree model



Random subspace based naïve Bayes tree

Firstly, we prepare a random subspace-based naïve Bayes tree (RSNBtree) model based on the single NBtree and the RS technique. The NBtree belongs to the DT intelligence algorithm, consisted of the C4.5 tree and NB (Kohavi 1996). Therefore, its architecture and learning procedure are similar to other types of DTs except that some leaf nodes that predict a single class are replaced by NB categorizers (Chen et al. 2017). The construction of a NBtree starts at the root node, sorts and splits the attribute, and then moves down the tree branch corresponding to the “purity” of the attribute. Above steps will be repeated several rounds until a terminal node is met.

For NBtree, the information gain ratio (IGR) is adopted as the splitting criteria to measure the “purity” (Quinlan 1993). IGR measure is based on the concept of entropy and defined as below:

$$\text{Entropy}(D) = - \sum_{i=1}^m \frac{|D_i|}{|D|} \log_2 \frac{|D_i|}{|D|} \quad (3)$$

where $|D|$ is the total number of cases in the dataset D , $|D_i|$ denotes the number of the cases that belongs to the class D_i , and m is the number of classes.

Then, the IGR of an attribute A can be calculated as following formula:

$$\text{IGR}(A) = \frac{\text{Entropy}(D) - \text{Entropy}_A(D)}{\text{SplitInfo}(A)} \quad (4)$$

where $\text{Entropy}_A(D)$ represents the entropy of the D after being segmented on attribute A and $\text{SplitInfo}(A)$ is the normalized factor.

If a node is confused by splitting criteria and not able to make predictions, a NB classifier can be used to decide and select the class that maximizes the posterior probability. The NB assumes that all attributes related to the target class are conditionally independent. The classification of the NB is performed as below:

$$c = P(c_i) \prod_{j=1}^n P(A_j|c_i) / \sum_{i=1}^m \left[P(c_i) \prod_{j=1}^n P(A_j|c_i) \right] \quad (5)$$

where n is the total number of attributes and P represents the possibility of a class that occurs.

The NBtree is used to integrate with the RS method in order to construct a RSNBtree model. To be specific, the RS technique randomly samples the sub-feature space from the original dataset to generate a subset with q dimensionality ($q < n$). A base-learner is subsequently applied to each of these subsets considering the NBtree algorithm. After several sampling rounds (e.g., $t \in [1, 2, \dots, T]$), a series of subsets and NBtrees will be created in parallel and each NBtree is specialized in the corresponding subset. The final

decision for predicting landslide susceptibility is obtained by majority voting all NBtrees:

$$c(x) = \text{argmax} \sum_{t=1}^T 1(\text{NBtree}_t(x) = y) \quad (6)$$

Bagging ensemble

Bagging is probably one of the most well-known ensemble methods, which was firstly introduced by Breiman (1996). The Bagging mingles classifications from casually produced training sets in parallel using a bootstrap resampling scheme. Therefore, Bagging is considered to be an efficient ensemble to improve unstable or poor estimation. In Bagging, each randomly sampled subset is used to build a predictor. As the sampling procedure proceeds (T rounds), these predictors will be aggregated to form a new bagging ensemble model based on majority voting. The advantage of the bagging method is that it can perform self-evaluation using out-of-bag samples.

In this study, the RSNBtree is treated as the base-learner of the Bagging algorithm to construct a BRSNBtree model. To implement this novel ensemble method, we keep the iteration rounds at 20, and set the size of subspaces (a percentage of remained features) and the size of subsets (a percentage of sampled data) as 75% and 80%.

Results

Selection of landslide conditioning factors

The importance evaluation result is given in Fig. 4. Among the eleven landslide conditioning factors, the distance to rivers has the highest importance level (ReF = 51.03%), then followed by distance to roads (21.89%), annual rainfall (7.00%), elevation (4.33%), ERG (3.78%), land use (2.91%), NDVI (2.61%), slope angle (2.48%), slope structure (1.65%), SPI (1.18%), and TWI (0.85%), respectively. It turns out that each factor has a positive contribution to landslide spatial prediction. The result of the PCC analysis is shown in Table 1. We can observe that correlation between TWI and SPI has the highest correlation with a value of 0.71, which indicates a strong correlation and reaches the critical threshold (0.7) according to Martín et al. (2012). By comparison, the remaining pairs present a weak correlation. Considering the ReF evaluation results in combination with the PCC results, the TWI was removed from the initial dataset because of its incompetence and high correlation, while the remaining ten factors were used as input to perform landslide susceptibility modeling.

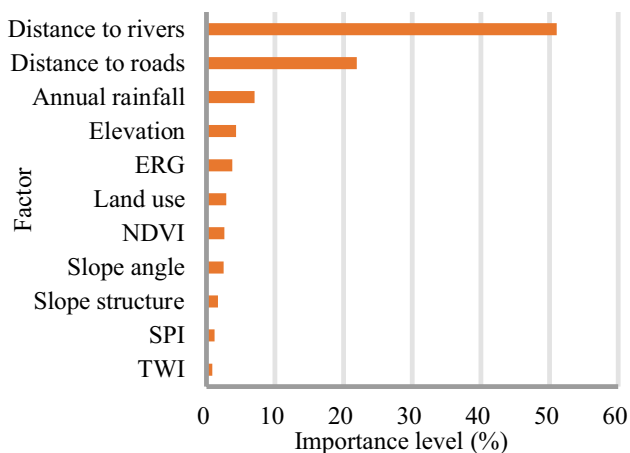


Fig. 4 Importance analysis of conditioning factors using the Ref method

Landslide susceptibility mapping

The RSNBtree model and the BRSNBtree model were constructed and applied to predict landslide susceptibility for Zigui County. The selected landslide-related factors were fed into the modeling to estimate the landslide susceptibility index (LSI) that indicates the possibility of landslide occurrence for each unit. Landslide susceptibility maps of the study area were then depicted in ArcGIS 10.2 software and reclassified into five levels using the geometrical interval classification method: very low susceptibility (VLS), low susceptibility (LS), moderate susceptibility (MS), high susceptibility (HS), and very high susceptibility (VHS) (Fig. 5). The area percentage of each susceptibility level has been summarized in Fig. 6. It can be observed that two landslide models present similar mapping patterns

and areal distributions. By using the BRSNBtree model, 21.92%, 12.55%, 25.53%, 27.48%, and 12.51% of the area were respectively classified in the groups of VLS, LS, MS, HS, and VHS. Most of the landslides (76.46%) are located in extremely and highly vulnerable area. About 15.86%, 4.09%, and 3.59% of the total landslides fall into MS, LS, and VLS, respectively. The result demonstrates that the landslide susceptibility map constructed by the BRSNBtree model has good spatial prediction accuracy. Regarding the RSNBtree, 21.45% of the areas have VLS level and 15.84% have LS level, and areas with MS, HS, and VHS were modeled to account for 26.77%, 24.47%, and 11.47% of the study area, respectively. Similar results between the RSNBtree and BRSNBtree connote that the use of the Bagging ensemble scheme will not significantly influence the areal composition of landslide susceptibility levels of the RSNBtree. However, some differences are also can be found among the two models. For instance, more landslides were assigned to the MS level by using the RSNBtree model when compared with the BRSNBtree model. This leads to relatively fewer landslides (7%) to fall into the HS and VHS area when the RSNBtree model was used to predict landslide susceptibility for the study area.

Inspection of the pattern of landslide susceptibility shows that areas with VHS or HS tend to be distributed along the Yangtze River and its tributaries, which further confirms that rivers play a significant role in effecting slope stability in Zigui County. Additionally, VHS and HS levels can be observed in the areas far from rivers but nearby roads or with high annual rainfall.

Performance evaluation

In the present study, the statistical indices and the receiver operating characteristic curve (ROC) were adopted to

Table 1 Pearson correlation coefficient evaluation results

	Var1	Var2	Var3	Var4	Var5	Var6	Var7	Var8	Var9	Var10	Var11
Var1	1	0.07	0	-0.09	0.23	0.37	0.24	0	0.42	-0.1	0.1
Var2		1	0.08	0.08	0.25	0.09	0.03	0.29	-0.02	-0.13	0.2
Var3			1	0	0.03	-0.02	0.03	0.02	0.09	-0.01	0.04
Var4				1	0.04	-0.02	0.08	0.03	0.02	0.71	0.03
Var5					1	0.51	0.09	0.11	0.03	-0.03	0.3
Var6						1	0.13	0.06	0.05	-0.04	0.17
Var7							1	-0.1	0.35	0.03	0.11
Var8								1	-0.04	-0.09	-0.05
Var9									1	0.01	0.06
Var10										1	-0.03
Var11											1

Bold font means relatively high correlation

Var1 elevation, Var2 slope angle, Var3 slope structure, Var4 SPI; Var5, distance to roads, Var6 distance to rivers, Var7, ERG, Var8, NDVI, Var9 annual rainfall, Var10 TWI, Var11, Land use.

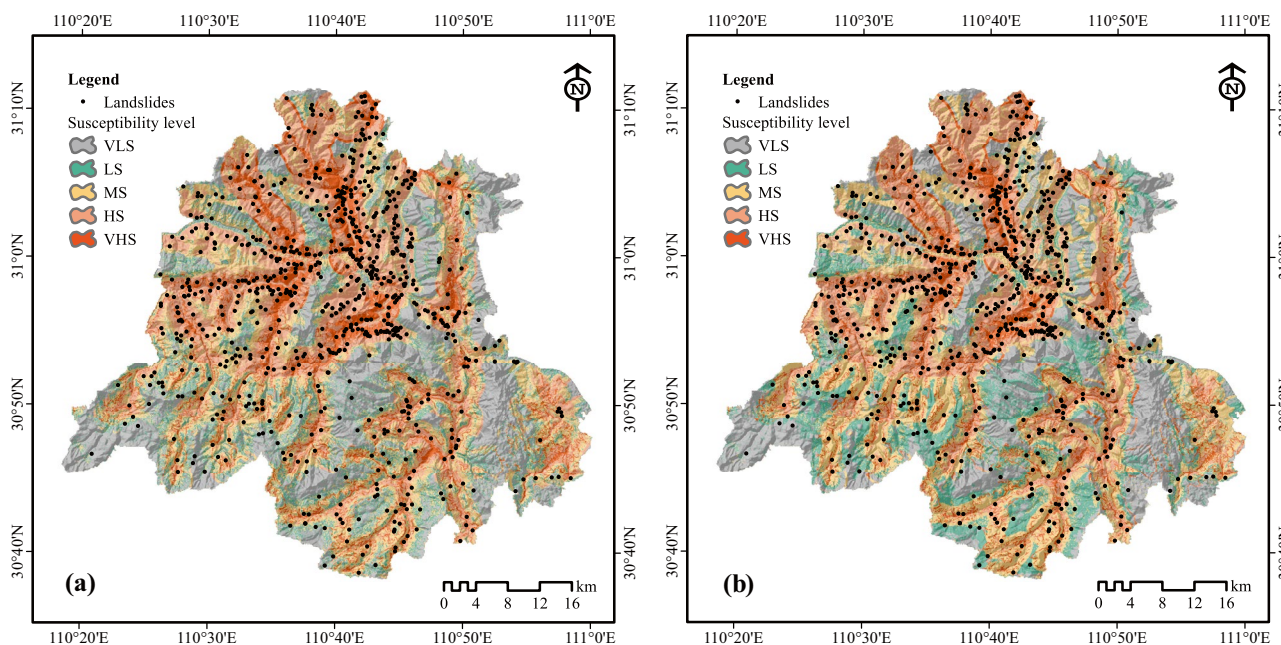
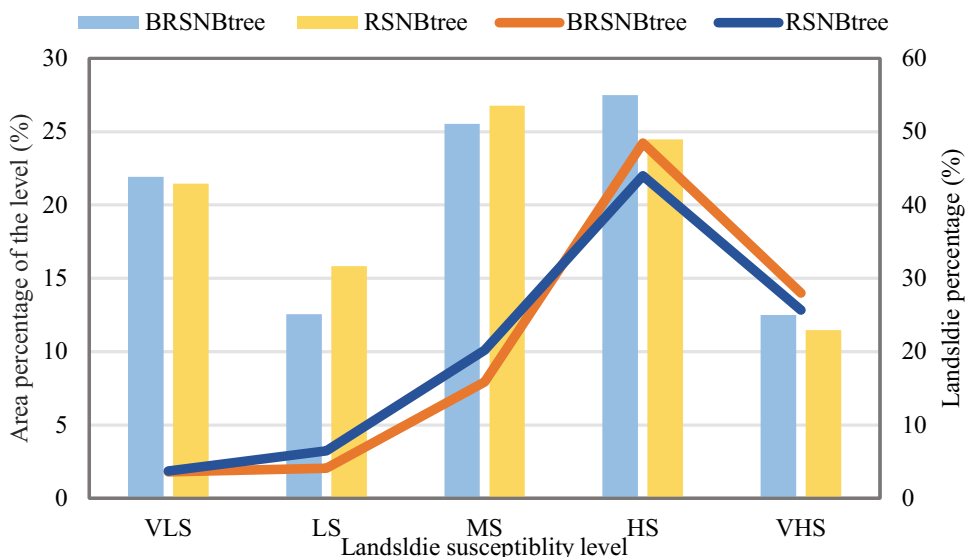


Fig. 5 Landslide susceptibility map of the study area: (a) the BRSNBtree model and (b) the RSNBtree model

evaluate the model’s performance; they are commonly used evaluation criteria in landslide studies (Jiao et al. 2019). Statistical measures used in this study include the accuracy (ACC) root-mean-squared error (RMSE), kappa statistic (K), and F-measure (F). In addition, the area under the ROC (AUC) is also an important performance metric. Those metrics are defined based on the true positive (TP), true negative (TN), false positive (FP), and false negative (FN), and the detailed description and calculation can refer to Dou et al. (2019) and Wu et al. (2020).

To verify the proposed landslide modeling method, two successful algorithms such as SVM and RF were further implemented for the comparison. The SVM is developed from the concept of structural risk minimization and tries to design a separating hyperplane to maximize the margin of different classes (Vapnik 1995). RF achieves a powerful ensemble version of the CART algorithm by resampling original dataset with replacement and randomly modifying the predictive variables (Breiman 2001). SVM and RF have been widely applied for landslide susceptibility assessment

Fig. 6 Distribution of landslide susceptibility levels



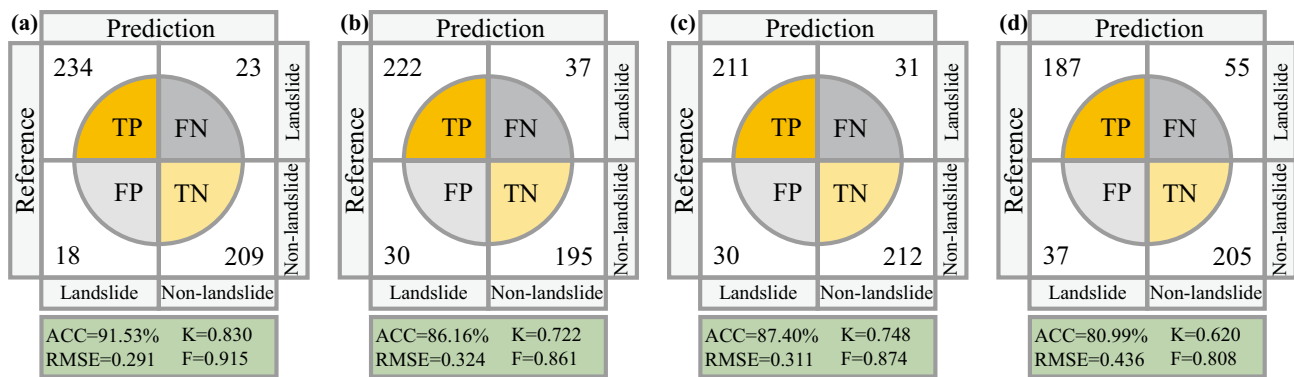


Fig. 7 Model evaluation using statistical measures: (a) the BRSNBtree model; (b) the RSNBtree model; (c) the RF model; (d) the SVM model

and used as benchmark methods. In this work, the radial-basis-function SVM with the penalty coefficient of 0.4 was utilized. As for the RF algorithm, numbers of trees and random predictive variables were respectively set as 500 and 4.

Performance evaluation results were shown in Fig. 7. Our method holds the best prediction capability among all models, with the ACC, K, RMSE, and F values of 91.53%, 0.83, 0.291, and 0.915. The RF comes second, which gained ACC, K, RMSE, and F with values of 87.40%, 0.748, 0.311, and 0.874, respectively. Additionally, RSNBtree achieved an acceptable performance with ACC, K, RMSE, and F values of 86.16%, 0.722, 0.324, and 0.861. By contrast, the SVM had slightly lower values for all the abovementioned metrics. The ACC, K, RMSE, and F values for the SVM model were 80.99%, 0.620, 0.436, and 0.808, respectively. The overall performance of various models is presented using the AUC measure (Fig. 8). All four landslide models have acceptable performance, yet the BRSNBtree achieved more preferable results than the remaining models because the BRSNBtree model yielded the highest AUC (0.968), followed by the RF (0.949), the RSNBtree (0.936), and the SVM (0.895). Overall, the evaluation based on various performance metrics denotes that the proposed BRSNBtree method is capable to refine the RSNBtree and outperforms the RF and SVM. In general, the model with better performance is far reliable in terms of assessing landslide susceptibility. The BRSNBtree produces the best landslide susceptibility modeling result and is recommended to detect landslide susceptible areas for the Zigui County.

Discussion

There have been continuous efforts in exploring efficient means for landslide spatial prediction at the Three Gorges Reservoir Area (Bi et al. 2012; Chen et al. 2016). Our case study area Zigui County is one of the most important areas of detection and prevention of landslide disasters in the

east of the Three Gorges Reservoir Area. Machine learning models have been widely applied in landslide risk studies. However, complex and non-linear relationship between landslide occurrence and affecting factors varies from regions to regions, which brings difficulty to landslide prediction tasks. The objective of this study is to develop a novel machine learning ensemble method for improving landslide susceptibility mapping for the landslide-prone Zigui area. Moreover, reviews on landslide studies show that the ensemble modeling was still rarely applied in this region. We expect the proposed ensemble method should increase the accuracy of landslide spatial prediction and facilitate landslide prevention for the study area.

Conducting feature selection is highly recommended before landslide modeling as the quality of a model is affected by the data used (Pham et al. 2016; Dou et al.

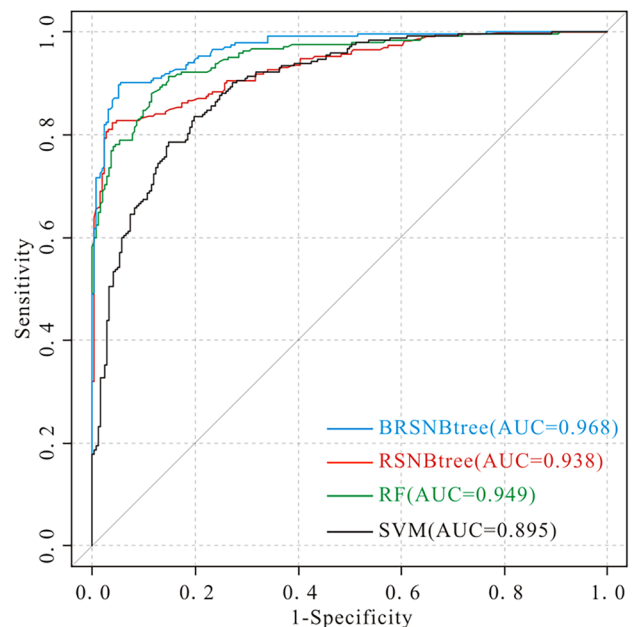


Fig. 8 ROC analysis of the various landslide models

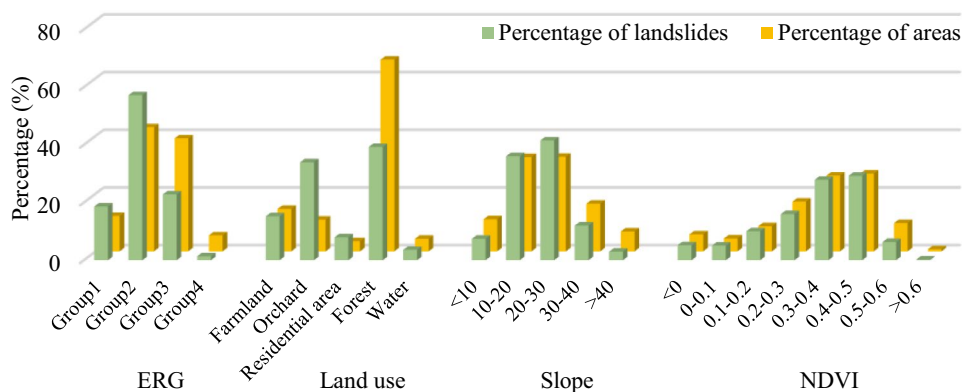
2019). In this study, the ReF and PCC methods were jointly used to perform feature selection, including importance level test and correlation analysis. According to the results from the ReF and PCC evaluations, 10 of 11 conditioning factors were finally remained as the model's input. Note that the distance to rivers gained the highest importance among the selected factors. This result quite matches the actual conditions of Zigui County and agrees with the landslide studies on other segments of the Three Gorges Reservoir Area—highly susceptible regions tend to be distributed along riverbanks (Chen et al. 2014, 2016; Zhang et al. 2017; Yu and Gao 2020). Particularly, Chen et al. (2012) evaluated landslide affecting factors in the Zigui segment of the Three Gorges Reservoir Area based on the likelihood ratio method and found that distances between 0 and 200 m away from rivers are most closely related to landslide occurrences. Our study area belongs to the upper reaches of the Yangtze River, with characters of developed stream systems and active fluctuation of water level. Slope toes nearby rivers suffer from water erosion and soaking environment, which leads to changes in soil moisture and rock hardness thereby particularly conducive to slope failure. Moreover, the high susceptibility level also can be observed in areas with high road density and low elevation. The cooperation of these negative factors also accompanied by rainfall further promotes slope sliding.

In contrast, some factors such as slope, ERG, land use, and NDVI have a small impact on landslide occurrence. We summarized the area percentage of factor's subclasses and corresponding percentage of landslides for the four factors to illustrate the relationship between variable distribution and landslide occurrence from the statistic perspective (Fig. 9). The statistic result regarding ERG shows that 80% of landslides are located in classolite and carbonatite that together occupy areas with 82.2% of the whole region. For land use, 39.3% of landslides occur in the forest that accounts for 66.5% of the total areas of the study area. In terms of the slope, most landslides (77.6%) are allocated to the class of

10–30°, which covers 65.4% of the total study area. As for the NDVI, 57.1% of the total landslides belong to the class of 0.3–0.5, the area of which is nearly half of the whole region. The results connote that most of the landslides tend to be widely spread across a large region associated with a single subclass or a narrow range of a factor, which has little indicative significance to landslide occurrence especially in the case where slope failure is influenced by various factors. In such areas, the factors with high discriminability might be more helpful for landslide susceptibility prediction. Therefore, the underperformance of the slope, ERG, land use, and NDVI might be attributed to the outweighing of other stronger factors (e.g., distance to rivers, distance to roads, and annual rainfall). Similar results can be inspected in a landslide study at the Zigui–Badong segment of the Three Gorges Reservoir Area (Zhang et al. 2017), where some factors such as the EGR, slope, and vegetation coverage indices are far less important in landslide susceptibility modeling.

For a specific study area, the option of modeling method is also important in analyzing landslide susceptibility. In this study, we introduced a novel method referred to as the BRSNBtree, and investigated its performance for landslide prediction in the Zigui area. The proposed BRSNBtree performs very well and significantly improves RSNBtree. Notably, our method achieved a superior predictive capability than RF and SVM. Given the advantageous properties of the BRSNBtree model, (1) machine learning ensemble method can optimize the fitting function of algorithms and decrease classification errors of landslide models (Hu et al. 2020), (2) RS algorithm is capable to avoid over-fitting issues (Onan 2015), and (3) Bagging is confident in reducing the variance of base learning algorithms (Breiman 1996). Inspection of previous landslide studies shows that RS and Bagging are effective in ensemble prediction. The RS has been successfully employed to refine the single algorithms such as SVM (Tien Bui et al. 2019), ANN (Pham et al. 2017), and REPT (Pham et al. 2019), in different landslide modeling tasks. Pham et al. (2018) reported that the CART integrated with

Fig. 9 Percentage of areas and corresponding percentage of landslides regarding the ERG, land use, slope, and NDVI



RS method performed better than the SVM, NBtree, and LR. Arabameri et al. (2021) concluded that the RS was more useful than the Bagging in ensemble modeling when Credal-C4.5 was treated as the component model. Particularly, RSNBtree was suggested to be a promising landslide susceptibility modeling method (Shirzadi et al. 2017). Our study reveals that the prediction capability of the RSNBtree can be further improved by using the Bagging scheme when landslide modeling is the case. This finding is related to previous studies (Hong et al. 2018; Pham et al. 2017; Dou et al. 2019), all of whom state that ensemble learning can reinforce the performance of base-classifiers. Chen et al. (2018c) successfully coupled the Bagging with the kernel logistic regression (KLR) and concluded that such a combination can overcome over-fitting and variance problems of data. Pham and Prakash (2017) proved that the hybridization of Bagging ensemble and NBTree could achieve a high classification accuracy and outperform the NBTree with rotation forest ensemble, single NBTree, and the SVM when performing landslide modeling. Pham et al. (2017) found that Bagging and RS had an equal effect on the ANN, and the ANN with the Bagging or BS was superior to that with the Boosting. Likewise, Nhu et al. (2020a) proved that the accuracy of the RF can be significantly improved by 9.5% by using RS or Bagging techniques for predicting landslide susceptibility. Hu et al. (2021) reported that the use of different base-learners could influence the results of ensemble prediction, and the Bagging-based ANN generated more stable and accurate modeling results than the Bagging-based C4.5. These ensemble models achieved encouraging performance for landslide susceptibility modeling, but the joint use of different ensemble strategies for landslide models is rarely explored. The introduced BRSNBtree model is benefited from the cooperation effects of RS and Bagging ensemble, which achieves further improvement for landslide prediction. Therefore, the main advantage of BRSNBtree is to decrease the model's variance and mitigate over-fitting problems (Breiman 1996; Onan 2015). Landslide susceptibility mapping results also show advantages of our method because the HS and VHS areas modeled from the BRSNBtree can capture more landslides than those from the RSNBtree. The results mean that the landslide susceptibility assessment conducted by the BRSNBtree model may be more practical for landslide prediction and prevention. Nevertheless, a possible limitation of ensemble modeling is that improvement in performance requires higher calculation cost. For example, the BRSNBtree model (76.8 s) takes more times than the RSNBtree (7.4 s), RF (5.1 s), and SVM (0.17 s) in modeling process.

Finally, to verify the reliability of landslide susceptibility maps developed by the BRSNBtree model, we further calculated a reliability index namely the landslide density (LD). It is defined as a proportion of the number of landslides and

areas of corresponding susceptibility level (Jiao et al. 2019). The LD values of our method are respectively 0.06, 0.11, 0.22, 0.62, and 0.79 with the susceptibility level ranging from the VLS to VHS. It can be seen that the value of LD increases as the susceptibility level improves, indicating a reliable susceptibility assessment (Pham et al. 2016). More importantly, the LD value regarding the very high susceptibility level in this study reaches 0.79, better than the study of Liu et al. (2014) (the corresponding LD is 0.21), which further confirms the effectiveness of our method.

However, some limitations should be noted. The scale of the ERG map used in this study is not in line with other factor layers. Finding geology data with an appropriate scale for landslide susceptibility analysis is a crucial issue. Nevertheless, it may be difficult to obtain more detailed ERG data at present because extensive field survey is time-consuming and labor-intensive. For landslide modeling, Catani et al. (2013) examined model's performance under different survey scales of landslide-related factors and different resolutions of mapping units. The performance of landslide models is affected by the scale but layers with smaller scales (or coarser resolution) might not necessarily lead to a decreased prediction accuracy. In the future, we will attempt to collect available geology data with larger scales and focus on data scaling issues for landslide susceptibility mapping at the case area.

Conclusion

A novel machine learning ensemble model that integrates RSNBtree with the Bagging technique was applied to assess landslide susceptibility at the Zigui County of the Three Gorges Reservoir Region. Based on the feature selection method, 10 landslide conditioning factors were fed into landslide models for predicting landslide susceptibility. The distance to rivers, distance to roads, and annual rainfall have the greatest effect on landslide occurrence in the study area. Particularly, it can be observed from landslide susceptibility distribution patterns that vulnerable areas are highly associated with stream systems. The result emphasizes the significant role of hydrology in promoting slope instability.

Model performance evaluation results indicate that the RSNBtree can be further reinforced through the Bagging scheme. Benefited from cooperation effects of RS and Bagging techniques, the developed BRSNBtree model achieved the highest prediction capability for landslide susceptibility modeling and outperformed the SVM and RF. Hence, the BRSNBtree provides a promising and better way to target landslide-prone areas for Zigui County. Landslide susceptibility maps developed from this study would help local managers to get better knowledge of the state of sliding risk and facilitate landslide mitigation and management. Also, this study demonstrated the superiority of ensemble methods in landslide susceptibility assessment.

Acknowledgements The authors are grateful to the Wuhan Center of Geological Survey and the Department of natural resources of Hubei Province for providing the data used in this study. In addition, we are deeply grateful to Prof. Bob Criss (Department of Earth and Planetary Sciences, Washington University) for giving careful suggestions to our work, which help to improve the quality of this paper.

Funding This research was supported by the Project “Construction of Geological Hazard Risk Identification and Risk Release System in the Three Gorges Reservoir Area” (No. 0001212012AC50001).

References

- Alcántara-Ayala I (2002) Geomorphology, natural hazards, vulnerability and prevention of natural disasters in developing countries. *Geomorphology* 47(2–4):107–124. [https://doi.org/10.1016/s0169-555x\(02\)00083-1](https://doi.org/10.1016/s0169-555x(02)00083-1)
- Arabameri A, Pal SC, Rezaie F, Chakraborty R, Saha A, Blaschke T et al (2021) Decision tree based ensemble machine learning approaches for landslide susceptibility mapping. *Geocarto Int.* <https://doi.org/10.1080/10106049.2021.1892210>
- Bennett GL, Miller SR, Roering JJ, Schmidt DA (2016) Landslides, threshold slopes, and the survival of relict terrain in the wake of the Mendocino Triple Junction. *Geology* 44(5):363–366. <https://doi.org/10.1130/G37530.1>
- Bi R, Ehret D, Xiang W, Rohn J, Schleier M, Jiang J (2012) Landslide reliability analysis based on transfer coefficient method: a case study from Three Gorges Reservoir. *Journal of Earth Science* 23(2):187–198. <https://doi.org/10.1007/s12583-012-0244-7>
- Breiman L (1996) Bagging predictors. *Mach Learn* 24(2):123–140. <https://doi.org/10.1007/BF00058655>
- Breiman L (2001) Random forest. *Mach Learn* 45(1):5–32
- Catani F, Lagomarsino D, Segoni S, Tofani V (2013) Landslide susceptibility estimation by random forests technique: sensitivity and scaling issues. *Nat Hazard* 13(11):2815–2831. <https://doi.org/10.5194/nhess-13-2815-2013>
- Chen T, Niu R, Du B, Wang Y (2014) Landslide spatial susceptibility mapping by using GIS and remote sensing techniques: a case study in Zigui County, the Three Georges reservoir. *China Environmental Earth Sciences* 73(9):5571–5583. <https://doi.org/10.1007/s12665-014-3811-7>
- Chen T, Niu R, Jia X (2016) A comparison of information value and logistic regression models in landslide susceptibility mapping by using GIS. *Environ Earth Sci* 75(10). <https://doi.org/10.1007/s12665-016-5317-y>
- Chen W, Li X, Wang Y, Liu S (2012) Landslide susceptibility mapping using LiDAR and DMC data: a case study in the Three Gorges area. *China Environmental Earth Sciences* 70(2):673–685. <https://doi.org/10.1007/s12665-012-2151-8>
- Chen W, Shirzadi A, Shahabi H, Ahmad BB, Zhang S, Hong H et al (2017) A novel hybrid artificial intelligence approach based on the rotation forest ensemble and naive Bayes tree classifiers for a landslide susceptibility assessment in Langao County, China. *Geomat Nat Haz Risk* 8(2):1955–1977. <https://doi.org/10.1080/19475705.2017.1401560>
- Chen W, Peng J, Hong H, Shahabi H, Pradhan B, Liu J et al (2018a) Landslide susceptibility modelling using GIS-based machine learning techniques for Chongren County, Jiangxi Province, China. *Sci Total Environ* 626:1121–1135. <https://doi.org/10.1016/j.scitotenv.2018.01.124>
- Chen W, Shahabi H, Shirzadi A, Hong H, Akgun A, Tian Y et al (2018b) Novel hybrid artificial intelligence approach of bivariate statistical-methods-based kernel logistic regression classifier for landslide susceptibility modeling. *Bull Eng Geol Env* 78(6):4397–4419. <https://doi.org/10.1007/s10064-018-1401-8>
- Chen W, Shahabi H, Zhang S, Khosravi K, Shirzadi A, Chapi K et al (2018c) Landslide susceptibility modeling based on GIS and novel Bagging-based kernel logistic regression. *Appl Sci* 8(12). <https://doi.org/10.3390/app8122540>
- Chen Z, Wang J (2010) Land use and land cover change detection using satellite remote sensing techniques in the mountainous Three Gorges Area. *China International Journal of Remote Sensing* 31(6):1519–1542. <https://doi.org/10.1080/01431160903475381>
- Criss RE, Yao W, Li C, Tang H (2020) A predictive, two-parameter model for the movement of reservoir landslides. *Journal of Earth Science*. <https://doi.org/10.1007/s12583-020-1331-9>
- Dou J, Yunus AP, Bui DT, Merghadi A, Sahana M, Zhu Z et al (2019) Improved landslide assessment using support vector machine with bagging, boosting, and stacking ensemble machine learning framework in a mountainous watershed, Japan. *Landslides* 1–18. <https://doi.org/10.1007/s10346-019-01286-5>
- Fernández DS, Lutz MA (2010) Urban flood hazard zoning in Tucumán Province, Argentina, using GIS and multicriteria decision analysis. *Eng Geol* 111(1–4):90–98
- Freund Y, Schapire RE (1997) A decision-theoretic generalization of on-line learning and an application to boosting. *J Comput Syst Sci* 55:119–139. <https://doi.org/10.1006/jcss.1997.1504>
- Ghasemian B, Asl DT, Pham BT, Avand M, Nguyen HD, Janizadeh S (2020) Shallow landslide susceptibility mapping: a comparison between classification and regression tree and reduced error pruning tree algorithms. *Vietnam Journal of Earth Sciences* 42(3):208–227. <https://doi.org/10.15625/0866-7187/42/3/14952>
- Ghorbanzadeh O, Blaschke T, Gholamnia K, Meena S, Tiede D, Aryal J (2019) Evaluation of different machine learning methods and deep-learning convolutional neural networks for landslide detection. *Remote Sensing* 11(2). <https://doi.org/10.3390/rs11020196>
- Goetz JN, Brenning A, Petschko H, Leopold P (2015) Evaluating machine learning and statistical prediction techniques for landslide susceptibility modeling. *Comput Geosci* 81:1–11. <https://doi.org/10.1016/j.cageo.2015.04.007>
- Grelle G, Soriano M, Revellino P, Guerriero L, Anderson MG, Diambra A et al (2013) Space–time prediction of rainfall-induced shallow landslides through a combined probabilistic/deterministic approach, optimized for initial water table conditions. *Bull Eng Geol Env* 73(3):877–890. <https://doi.org/10.1007/s10064-013-0546-8>
- He Y, Liu Y, Zhao H, Chen X (2008) The analysis of land use spatial patterns responded to different geomorphology type: two cases studies in Hubei province, China. Paper presented at the Geoinformatics 2008 and Joint Conference on GIS and Built Environment: Advanced Spatial Data Models and Analyses Guangzhou, China, Jun 28–29, 2008
- Ho TK (1998) The random subspace method for constructing decision forests. *IEEE Trans Pattern Anal Mach Intell* 20(8):832–844. <https://doi.org/10.1109/34.709601>
- Hong H, Pourghasemi HR, Pourtaghi ZS (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, Liu J, Bui DT, Pradhan B, Acharya TD, Pham BT et al (2018) 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>
- Hu X, Zhang H, Mei H, Xiao D, Li Y, Li M (2020) Landslide susceptibility mapping using the Stacking ensemble machine learning method in Lushui. *Southwest China Applied Sciences* 10(11):4016–4037. <https://doi.org/10.3390/app10114016>

- Hu X, Mei H, Zhang H, Li Y, Li M (2021) Performance evaluation of ensemble learning techniques for landslide susceptibility mapping at the Jinping county. *Southwest China Natural Hazards* 105(2):1663–1689. <https://doi.org/10.1007/s11069-020-04371-4>
- Jia K, Liang S, Zhang L, Wei X, Yao Y, Xie X (2014) Forest cover classification using Landsat ETM+ data and time series MODIS NDVI data. *International Journal of Applied Earth Observations Geoinformation* 33(1):32–38
- Jiao Y, Zhao D, Ding Y, Liu Y, Xu Q, Qiu Y et al (2019) Performance evaluation for four GIS-based models purposed to predict and map landslide susceptibility: a case study at a World Heritage site in Southwest China. *CATENA* 183:104–221. <https://doi.org/10.1016/j.catena.2019.104221>
- Juliev M, Mergili M, Mondal I, Nurtaev B, Pulatov A, Hubl J (2019) Comparative analysis of statistical methods for landslide susceptibility mapping in the Bostanlik District, Uzbekistan. *Sci Total Environ* 653:801–814. <https://doi.org/10.1016/j.scitotenv.2018.10.431>
- Kira K, Rendell LA (1992) The feature selection problem: traditional methods and a new algorithm. In: *Proceedings of the 10th National Conference on Artificial Intelligence*, Atlanta, Germany, April 17, 1992, vol 4. AAAI Press, pp 129–134
- Kohavi R (1996) Scaling up the accuracy of naive-Bayes classifiers: a decision-tree hybrid. In: *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, Portland, August, 1996. AAAI Press, pp 202–207
- Lee S, Lee MJ, Jung HS, Lee S (2020) Landslide susceptibility mapping using Naïve Bayes and Bayesian network models in Umyeonsan. *Korea Geocarto International* 35(15):1665–1679. <https://doi.org/10.1080/10106049.2019.1585482>
- Li C, Fu Z, Wang Y, Tang H, Yan J, Gong W et al (2019) Susceptibility of reservoir-induced landslides and strategies for increasing the slope stability in the Three Gorges Reservoir Area: Zigui Basin as an example. *Eng Geol* 261(1):1–20. <https://doi.org/10.1016/j.enggeo.2019.105279>
- Liu Y, Fang J, Chen X, Chen Y (2014) Evaluation of landslide susceptibility in Zigui County based on certainty factor method. *J Nat Dis Sci* 23(6):209–217 (in Chinese with English Abstract). <https://doi.org/10.13577/j.jnd.2014.0626>
- Liu Z, Gilbert G, Cepeda JM, Lysdahl AOK, Piciullo L, Hefre H et al (2021) Modelling of shallow landslides with machine learning algorithms. *Geosci Front* 12(1):385–393. <https://doi.org/10.1016/j.gsf.2020.04.014>
- Lucchese LV, Oliveira GGD, Pedrollo OC (2021) Investigation of the influence of nonoccurrence sampling on landslide susceptibility assessment using Artificial Neural Networks. *CATENA* 198:105067–105078. <https://doi.org/10.1016/j.catena.2020.105067>
- Ma S, Xu C (2019) Applicability of two newmark models in the assessment of coseismic landslide hazard and estimation of slope-failure probability: an example of the 2008 wenchuan mw 7.9 earthquake affected area. *J Earth Sci* 30(5):1020–1030. <https://doi.org/10.1007/s12583-019-0874-0>
- Martín B, Alonso JC, Martín CA, Palacín C, Magaña M, Alonso J (2012) Influence of spatial heterogeneity and temporal variability in habitat selection: a case study on a great bustard metapopulation. *Ecol Model* 228:39–48. <https://doi.org/10.1016/j.ecolmodel.2011.12.024>
- Merghadi A, Abderrahmane B, Tien Bui D (2018) Landslide susceptibility assessment at Mila Basin (Algeria): a comparative assessment of prediction capability of advanced machine learning methods. *ISPRS Int J Geo Inf* 7(7):268. <https://doi.org/10.3390/ijgi7070268>
- Nhu VH, Shirzadi A, Shahabi H, Chen W, Clague JJ, Geertsema M (2020a) Shallow landslide susceptibility mapping by Random Forest base classifier and its ensembles in a semi-arid region of Iran. *Forests* 11(4):421–449. <https://doi.org/10.3390/f11040421>
- Nhu VH, Zandi D, Shirzadi A, Ansari NA, Singh SK, Shahabi HS et al (2020b) Comparison of support vector machine, Bayesian logistic regression, and alternating decision tree algorithms for shallow landslide susceptibility mapping along a mountainous road in the west of Iran. *Appl Sci* 10(15):5047–5074. <https://doi.org/10.3390/app10155047>
- Nsengiyumva JB, Luo G, Amanambu AC, Mind'je R, Habiyaemye G, Karamage F et al (2019) Comparing probabilistic and statistical methods in landslide susceptibility modeling in Rwanda/Centre-Eastern Africa. *Sci Total Environ* 659:1457–1472. <https://doi.org/10.1016/j.scitotenv.2018.12.248>
- Onan A (2015) Classifier and feature set ensembles for web page classification. *Journal of Information Ence* 42(2):150–165. <https://doi.org/10.1177/0165551515591724>
- Pandey VK, Pourghasemi HR, Sharma MC (2020) Landslide susceptibility mapping using maximum entropy and support vector machine models along the highway corridor. *Garhwal Himalaya Geocarto International* 35(2):168–187. <https://doi.org/10.1080/10106049.2018.1510038>
- Petschko H, Brenning A, Bell R, Goetz J, Glade T (2014) Assessing the quality of landslide susceptibility maps – case study Lower Austria. *Nat Hazard* 14(1):95–118. <https://doi.org/10.5194/nhess-14-95-2014>
- Pezhman AF, Mehdi V, Hadi KF, Elke H (2020) Regional flood frequency analysis through some machine learning models in semi-arid regions. *Water Resour Manage* 34:2887–2909. <https://doi.org/10.1007/s11269-020-02589-2>
- Pham BT, Bui DT, Pourghasemi HR, Indra P, Dholakia MB (2015) Landslide susceptibility assessment in the Uttarakhand area (India) using GIS: a comparison study of prediction capability of naïve bayes, multilayer perceptron neural networks, and functional trees methods. *Theoret Appl Climatol* 128(1–2):255–273. <https://doi.org/10.1007/s00704-015-1702-9>
- Pham BT, Pradhan B, Tien Bui D, Prakash I, Dholakia MB (2016) A comparative study of different machine learning methods for landslide susceptibility assessment: a case study of Uttarakhand area (India). *Environ Model Softw* 84:240–250. <https://doi.org/10.1016/j.envsoft.2016.07.005>
- Pham BT, Prakash I (2017) A novel hybrid model of Bagging-based Naïve Bayes Trees for landslide susceptibility assessment. *Bull Eng Geol Env* 78(3):1911–1925. <https://doi.org/10.1007/s10064-017-1202-5>
- Pham BT, Tien Bui D, Prakash I, Dholakia MB (2017) Hybrid integration of Multilayer Perceptron Neural Networks and machine learning ensembles for landslide susceptibility assessment at Himalayan area (India) using GIS. *CATENA* 149:52–63. <https://doi.org/10.1016/j.catena.2016.09.007>
- Pham BT, Prakash I, Tien Bui D (2018) Spatial prediction of landslides using a hybrid machine learning approach based on Random Subspace and Classification and Regression Trees. *Geomorphology* 303(15):256–270. <https://doi.org/10.1016/j.geomorph.2017.12.008>
- Pham BT, Prakash I, Singh SK, Shirzadi A, Shahabi H, Tran TTT et al (2019) Landslide susceptibility modeling using Reduced Error Pruning Trees and different ensemble techniques: Hybrid machine learning approaches. *CATENA* 175:203–218. <https://doi.org/10.1016/j.catena.2018.12.018>
- Pham BT, Phong TV, Nguyen-Thoi T, Parial KK, Singh S, Ly HB et al (2020) Ensemble modeling of landslide susceptibility using random subspace learner and different decision tree classifiers. *Geocarto Int* 1–23. <https://doi.org/10.1080/10106049.2020.1737972>
- Pourghasemi HR, Kerle N (2016) Random forests and evidential belief function-based landslide susceptibility assessment in Western Mazandaran Province. *Iran Environmental Earth Sciences* 75(3):185. <https://doi.org/10.1007/s12665-015-4950-1>
- Pourghasemi HR, Sadhasivam N, Kariminejad N, Collins AL (2020) Gully erosion spatial modelling: role of machine learning

- algorithms in selection of the best controlling factors and modeling process. *Geosci Front* 11(6):2207–2219. <https://doi.org/10.1016/j.gsf.2020.03.005>
- Preuth T, Glade T, Demoulin A (2010) Stability analysis of a human-influenced landslide in eastern Belgium. *Geomorphology* 120(1–2):38–47. <https://doi.org/10.1016/j.geomorph.2009.09.013>
- Quinlan JR (1993) C4. 5: programs for machine learning. Morgan Kaufmann San Mateo, California
- Rahali H (2017) Improving the reliability of landslide susceptibility mapping through spatial uncertainty analysis: a case study of Al Hoceima. *Northern Morocco Geocarto International* 34(1):43–77. <https://doi.org/10.1080/10106049.2017.1357767>
- Razavi-Termeh SV, Shirani K, Pasandi M (2021) Mapping of landslide susceptibility using the combination of neuro-fuzzy inference system (ANFIS), ant colony (ANFIS-ACOR), and differential evolution (ANFIS-DE) models. *Bull Eng Geol Env* 80:2045–2067. <https://doi.org/10.1007/s10064-020-02048-7>
- Regmi DA, Devkota CK, Yoshida K, Pradhan B, Pourghasemi RH (2014) Application of frequency ratio, statistical index, and weights-of-evidence models and their comparison in landslide susceptibility mapping in Central Nepal Himalaya. *Arab J Geosci* 7(2):725–742. <https://doi.org/10.1007/s12517-012-0807-z>
- Riegel RP, Alves DD, Schmidt BC, Oliveira GGd, Haetinger C, Osório DMM et al (2020) Assessment of susceptibility to landslides through geographic information systems and the logistic regression model. *Nat Hazards* 103:497–511. <https://doi.org/10.1007/s11069-020-03997-8>
- Rokach L (2009) Taxonomy for characterizing ensemble methods in classification tasks: a review and annotated bibliography. *Comput Stat Data Anal* 53(12):4046–4072. <https://doi.org/10.1016/j.csda.2009.07.017>
- Shirzadi A, Bui DT, Pham BT, Solaimani K, Chapi K, Kavian A et al (2017) Shallow landslide susceptibility assessment using a novel hybrid intelligence approach. *Environ Earth Sci* 76(2). <https://doi.org/10.1007/s12665-016-6374-y>
- Tang H, Wasowski J, Juang CH (2019) Geohazards in the Three Gorges Reservoir Area, China—lessons learned from decades of research. *Eng Geol* 261:1–16. <https://doi.org/10.1016/j.enggeo.2019.105267>
- Tehrany MS, Pradhan B, Jebur MN (2013) Spatial prediction of flood susceptible areas using rule based decision tree (DT) and a novel ensemble bivariate and multivariate statistical models in GIS. *J Hydrol* 504:69–79. <https://doi.org/10.1016/j.jhydrol.2013.09.034>
- Tien Bui D, Ho TC, Pradhan B, Pham BT, Nhu VH, Revhaug I (2016) GIS-based modeling of rainfall-induced landslides using data mining-based functional trees classifier with AdaBoost, Bagging, and MultiBoost ensemble frameworks. *Environmental Earth Sciences* 75(14):1–22. <https://doi.org/10.1007/s12665-016-5919-4>
- Tien Bui D, Shirzadi A, Shahabi H, Geertsema M, Omidvar E, Clague J et al (2019) New ensemble models for shallow landslide susceptibility modeling in a semi-arid watershed. *Forests* 10(9):743–765. <https://doi.org/10.3390/f10090743>
- Truong X, Mitamura M, Kono Y, Raghavan V, Yonezawa G, Truong X et al (2018) Enhancing prediction performance of landslide susceptibility model using hybrid machine learning approach of Bagging ensemble and Logistic Model Tree. *Appl Sci* 8(7). <https://doi.org/10.3390/app8071046>
- Tsangaratos P, Iliu I (2016) Comparison of a logistic regression and Naïve Bayes classifier in landslide susceptibility assessments: The influence of models complexity and training dataset size. *CATENA* 145:164–179. <https://doi.org/10.1016/j.catena.2016.06.004>
- Vapnik VN (1995) The nature of statistical learning theory. Springer, New York, USA. https://doi.org/10.1007/978-1-4757-3264-1_1
- Wu Y, Ke Y, Chen Z, Liang S, Zhao H, Hong H (2020) Application of alternating decision tree with AdaBoost and bagging ensembles for landslide susceptibility mapping. *CATENA* 187:1–17. <https://doi.org/10.1016/j.catena.2019.104396>
- Yao W, Li C, Zhan H, Zhou JQ, Criss RE, Xiong S et al (2020) Multi-scale study of physical and mechanical properties of sandstone in Three Gorges Reservoir Region subjected to cyclic wetting–drying of Yangtze River Water. *Rock Mech Rock Eng* 53(5):2215–2231. <https://doi.org/10.1007/s00603-019-02037-7>
- Yu X, Gao H (2020) A landslide susceptibility map based on spatial scale segmentation: a case study at Zigui-Badong in the Three Gorges Reservoir Area. *China Plos One* 15(3):1–20. <https://doi.org/10.1371/journal.pone.0229818>
- Zhang K, Wu X, Niu R, Yang K, 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). <https://doi.org/10.1007/s12665-017-6731-5>