Machine learning algorithms and techniques for landslide susceptibility investigation：A literature review

MA Yanbin[a], LI Hongrui[a], WANG Lin[a,b], ZHANG Wengang[a,b], ZHU Zhengwei[a,b],
YANG Haiqing[a,b], WANG Luqi[a,b], YUAN Xingzhong[c]

（a. School of Civil Engineering；b. National Joint Engineering Research Center of Geohazards Prevention in the Reservoir Areas；c. School of Architecture & Urban Planning，Chongqing University，Chongqing 400045，P. R. China）

Abstract：There are many mountainous areas in China，with complex terrain，weak planes and geological structures and wide distribution of geohazards. Landslides are one of the most catastrophic natural hazards occurring in mountainous areas，leading to economic loss and casualties. Landslide susceptibility models are capable of quantifying the possibility of where landslides are prone to occur，which plays a significant role in formulating disaster prevention measures and mitigating future potential risk. Since expert-based models are difficult to quantify and generally depend on the subjective judgments，the accuracy and precision of landslide susceptibility models are now evolving from expert models and statistical learning toward the promising use of machine learning methods. This study presented critical reviews on current machine learning models for landslide susceptibility investigation，an extensive analysis and comparison between different machine learning techniques（MLTs）from case studies in the Three Gorges Reservoir area was presented. In combination with field survey information as well as historical data，machine learning models were used to map landslide susceptibility and help formulate landslide mitigation strategies. The advantages and limitations of several frequently employed algorithms were evaluated based on the accuracy and efficiency of landslide susceptibility forecasting models. As the result shows，the tree-based ensemble algorithms models achieved better compared with other commonly methods of papping landslide susceptibility. Furthermore，the effect of database quality and quantity is significant，and more applications of some advanced methods（i. e.，deep learning algorithms）are yet to be further explored in further researches.

Keywords：landslide；machine learning；landslide susceptibility；the Three Gorges Reservoir；deep learning

机器学习方法在滑坡易发性评价中的应用

马彦彬[a],李红蕊[a],王林[a,b],仇文岗[a,b],朱正伟[a,b],杨海清[a,b],王鲁琦[a,b],袁兴中[c]

（重庆大学 a. 土木工程学院;b. 库区环境地质灾害防治国家地方联合工程研究中心;c. 建筑城规学院,重庆 400045）

　摘　要：中国山区多、地形复杂,构造发育、地质灾害隐患分布广泛。滑坡作为山区最具灾难性的地质灾害之一,严重威胁着人民群众的生命及财产安全。构建滑坡易发性模型能够量化滑坡发生的

可能性,对制定防灾措施、减少潜在风险具有重要作用。由于经验驱动模型难以量化,且往往依赖主观判断,近年来,滑坡易发性模型的精度与准确度在从经验驱动和统计理论模型向新兴机器学习方向发展的过程中得到提升。对目前滑坡易发性评价常用的机器学习模型进行综合评述,并针对三峡库区的案例研究,对不同的机器学习技术进行广泛分析和比较。机器学习模型通过结合实地调查资料和历史数据,可绘制滑坡易发性地图,辅助制定滑坡减缓策略。根据滑坡易发性预测模型的准确性和效率,评价几种常用算法的优势和局限性。结果表明,与一些常用的滑坡易发性制图方法相比,基于树结构的集成算法模型性能更好。此外,高质量的数据库十分重要,深度学习算法的更多应用还有待进一步研究探索。

关键词:滑坡;机器学习;滑坡易发性;三峡库区;深度学习

1　Introduction

Landslide is a process in which the soil or rock on a slope falls，dumps，slides，spreads or flows due to the influence of various factors[1]. In recent years，landslide hazards have caused serious loss of human life and property，therefore，severely constrained economic and social development on a worldwide scale[2-5]. According to the statistics of China's Ministry of Natural Resources（Fig. 1），landslides are the most threatening geological hazards，causing incalculable damage[6-10]. To effectively reduce the damage caused by landslides，the status of landslide susceptibility identification using scientific and technical means based on existing landslide cases is the focus of current research.
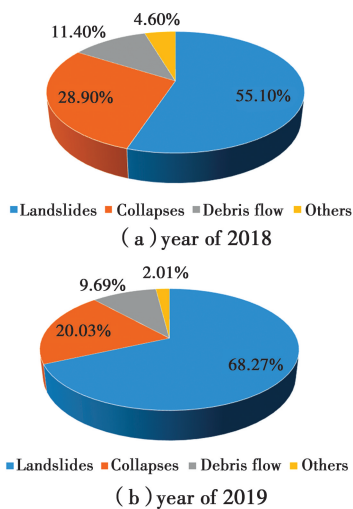

Fig. 1　Summary of the geological hazards in China

Landslide susceptibility is considered as the potential for slides to occur in an area under the influence of local topography and ambient factors[11] such as rainfall，earthquake，reservoir water level fluctuation，human engineering activity，etc. Landslide susceptibility assessments are generally divided into qualitative and quantitative methods. Qualitative approaches are based on the existing landslide inventories and the knowledge and experience of experts，defining landslide susceptibility through descriptive terms. Quantitative methods based on databases or physical models，predict landslide occurrences according to model calculation results. By comparison，quantitative methods reduce the subjectivity of qualitative methods[12]. With recent technical advances，using Geography Information Systems（GIS），Remote Sensing（RS），Global Positioning Systems（GPS）as sources of landslide database，combined with other data integration and analysis techniques is the most prevalent measure for landslide prediction and assessment. This requires considerations of a range of environmental factors，as well as the impact of the landslides that have occurred on the terrain，and is heavily dependent on the distribution and scale of the datasets[13]. However，the development of soft computing methods provides advanced alternative quantitative methods，which possesses the characteristics of low cost and high robustness by tolerating uncertainty，imprecision and incomplete true values.

As the most promising soft computing method，machine learning（ML）has been

developed since the 1950s，and has experienced two periods of declining popularity as artificial intelligence has affected the general public. In recent years，however，advances in computer technology and the rapid expansion of data and information have brought machine learning back into favor with scholars，mainly due to its ability to "learn" from large amounts of data by combining applied mathematics and computational intelligence to perform classification or regression tasks on unknown data. In the case of landslides in particular，machine learning methods are able to detect robust data structures that help model landslide susceptibility and predict landslide occurrences，despite missing values in monitoring information[14-16]. Generally，ML methods are more suitable extensive predictive modeling of landslide events，as well as classification tasks，considering that they are able to learn from complex and irregular data，establish relationships between data，and build algorithmic models without human intervention and prior assumptions. As an emerging technique in the fast development of the information age，ML is the product of statistical mathematics with artificial intelligence and other disciplines，and it is divided into supervised learning，semi-supervised learning and unsupervised learning. To the best of the authors' knowledge，the most common in current research contents is the use of supervised learning methods for landslide susceptibility modeling，solving classification or regression problems，and predicting the potential that landslides are likely to occur in the future[17-22].

In this paper，common ML classification and regression methods are reviewed，and the performance of artificial neural network（ANN），decision trees（DT），support vector machine（SVM），multivariate adaptive regression splines（MARS），random forest（RF）and extreme gradient boosting（XGBoost）models are compared based on the researches on landslide susceptibility

mapping（LSM）in the Three Gorges Reservoir area. A brief summary subsequently reviews the application of ML in landslide susceptibility prediction over the past two decades to illustrate the state of art and provide basic guidance for future research in this area.

2　Statistical analysis

2.1　Number of publications

In the period 2000-2020，publications on landslide susceptibility using "landslide"，"machine learning"，and "landslide susceptibility" as keywords from Web of Science were collected. A total of 614 papers in different publishing years were compiled，as shown in Fig. 2，it can be seen that there were less than ten studies per year about ML use in landslide susceptibility prediction during the first decade. After that，there was a significant increase in the number of published articles，most notably after 2016 when a sharp increase occurs，depending on the advancement of data acquisition technology and the continuous upgrading of computer hardware performance.
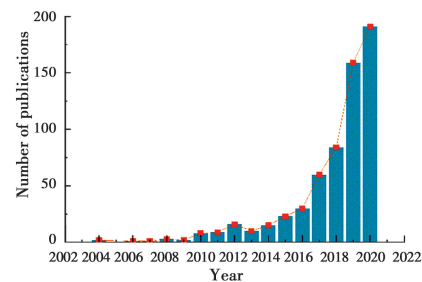
Fig. 2　Total number of publications in the last 20 years

2.2　Keywords

The keywords of 451 studies in the Science Citation Index Expanded（SCI-E）database out of 614 retrieved literature articles were analyzed. Statistically，a total of 521 of these keywords were used. The obvious is that the number of keywords has increased substantially in the last three years，consistent with the trend in the total number of publications，as shown in Fig. 3. Machine learning，GIS and landslides are the most commonly used keywords，and terms such as

landslide susceptibility and remote sensing have significantly received more attention in recent years due to the rapid development of data information and computer technology. In addition, machine learning methods such as random forest, support vector machine and deep learning have been used in recent years as data acquisition methods have diversified and data sizes have increased.
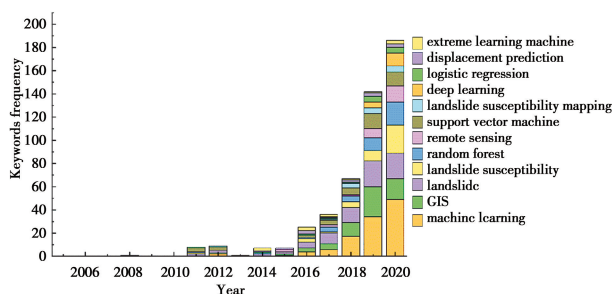

Fig. 3　Keyword trends in publications

2.3　Author

With 1 162 authors publishing in the area of machine learning methods for landslide susceptibility prediction, the author impact analysis was also established in the SCI-E database for the period 2000-2020. The top 15 authors with the most published articles are listed in Table 2, of which 10 authors have published more than 20 articles with a total of 1 861 citations, contributing significantly to the field of machine learning techniques (MLTs) for LSM researches. Among the publications by these authors, most focus on using logistic regression (LR)[23-24], decision trees (DT)[25-26], random forest (RF)[27-29], neural network (NNT)[30-31], and support vector machine (SVM)[32-33] methods. And from the methodological trends of these studies, it can be seen that scholars are more inclined to share research related to bagging and boosting algorithms which are representative machine learning methods.

To date, there is no consensus on which ML models are optimal for LSM due to the influence of landslide inducing factors, database quality, and the underlying assumptions of the ML algorithms. Thus, the next section focuses on discussing the

conditions of applicability, advantages and limitations of several representative ML models for critical comparison.

Table 1　The most productive authors in MLTs for LSM

Author	Total publications	Total citations	Average citations per pubs	Publications as first author
Bui D T	59	1 125	19.07	7
Pham B T	48	922	19.21	22
Shahabi H	34	736	21.65	2
Chen W	33	521	15.79	12
Shirzadi A	29	613	21.14	3
Prakash I	27	574	21.26	0
Pourghasemi H R	27	309	11.44	9
Pradhan B	24	455	18.96	0
Lee S	20	177	8.85	2
Bin Ahmad B	20	316	15.8	0
Hong H Y	19	246	12.95	1
Tang H M	19	164	8.63	0
Yin K L	15	168	11.2	0
Rahmati O	13	151	11.62	8

3　Comparison of MLTs for LSM

The Three Gorges Reservoir area refers to the area that was submerged or had migration tasks due to the Three Gorges Dam project on the Yangtze River, located in the Sichuan Basin and the middle and lower reaches of the Yangtze River plain. The excavation of the construction sites and the water storage process in the reservoir area has caused serious damage to the geological environment[34-38]. According to statistics, there were about 4 200 landslides in the whole Three Gorges Reservoir area, causing a large number of casualties and economic losses[39]. Four examples of landslides in the Three Gorges Reservoir area which are identified based on the interpretation of 1:10 000-scale color aerial photographs are shown in Fig. 4. Several published papers described landslides in the Three Gorges Reservoir area to study geological failure modes and increase slope stability by combining historical data, site

investigations，satellite images，and LSM has been commonly used to identify and map historical landslides and predict future landslide occurrences. Fig. 5 shows the methodology for LSM using MLTs. In this section，several methods such as artificial neural network（ANN），decision trees（DT），support vector machine（SVM），multivariate adaptive regression splines（MARS），random forest（RF）and extreme gradient boosting（XGBoost）and their applications in LSM are briefly described.
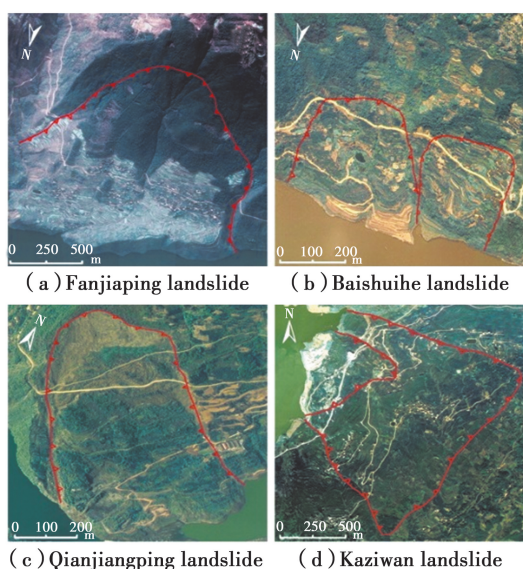

（a）Fanjiaping landslide   （b）Baishuihe landslide

（c）Qianjiangping landslide   （d）Kaziwan landslide

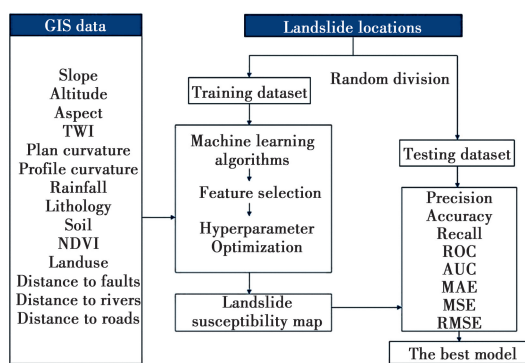**Fig. 4   Landslides aerial photographs（Peng et al. [40]）**

Fig. 5   Flow chart of LSM using MLTs

3.1   Artificial neural network

ANN is an algorithmic model of "artificial neurons" inspired by the generation mechanism of resting and action points of natural neurons in the human brain，and is one of the fastest growing research fields in recent years. This method has a strong nonlinear adaptive capability[41-45]. Not only can the information itself changes in many ways，but the nonlinear dynamic system is also constantly updating itself when processing information，which can simulate the intelligence of the human brain to actively adapt to the environment for learning. In general，it is suitable for building general mathematical models for datasets with no specific rules. However，since the dependence of the output variables on the input variables is nonlinear，it is not possible to visualize the effect of each input variable on the output value，so in fact，ANN is a black box model.

Depending on its advantage of handling variable relationships without relying on external rules，ANN algorithm becomes an effective tool for assessing landslide probability and predicting landslide occurrences[46-48]. In addition，Moayedi et al. [49] presented the optimization-artificial neural network（PSO-ANN）model to achieve model optimization to help establish the estimation of LSM. In general，the strong parallel processing and learning capabilities allow ANN to handle a large number of data samples，but precisely because of this，its model performance is very dependent on the quality of the landslide database and is relatively time-consuming.

3.2   Decision trees

DT is a typical supervised learning method and a predictive model that represents a mapping between object properties and object values. It uses a tree model to make decisions based on the properties of the data. The relationship between input variables and target variables can be linear or nonlinear，which allows for clear and concise explanations of variable associations or better visualization[50]，unlike ANN，it is not a black box model. DT can process data measured at different scales without making assumptions about frequencies or weights based on non-linear relationship between the data，this approach helps

explain complex relationship between variables and make predictions from the data[51-53].

The algorithm is actually a process of recursively selecting the optimal feature and partitioning the training data according to that feature, so that each sub-dataset has the best possible classification[54-56]. For the quantitative assessment of landslide susceptibility, there are numerous algorithms that can be used for DT method, such as iterative dichotomiser 3 (ID3), C4. 5[57-58], the classification and regression tree algorithm (CART)[59]. As a single decision tree model, the DT model is computationally simple and capable of handling both data-based and conventional attributes. And unlike black-box models, the corresponding logical expressions are easily introduced based on the resulting decision trees of a given model. Tree-based models such as RF, XGBoost are based on DT algorithm.

3. 3　Support vector machine

SVM is a supervised learning binomial classifier based on the risk minimization principle of structured architecture[60], originally developed by Vapnik[61]. Its basic model is a linear classifier with the largest spacing defined in the feature space, while the kernel approaches such as linear function, radial basis function and polynomial function, allow it to handle nonlinear problems[62]. Essentially, its approach seeks the best compromise between the learning accuracy of a particular training sample and the ability to correctly identify arbitrary samples based on limited sample information, in order to obtain the best generalization capability[63-66].

Among the many studies that predict landslides, SVM has been found to exhibit many unique advantages in handling small training samples, nonlinear and high-dimensional pattern recognition, and strong robustness can be obtained with very little model tuning compared with other multivariate statistical models[67-69]. The excellent generalization ability makes SVM one of the most commonly used and effective classifiers. And yet, SVM does not perform better than other algorithms in all landslide case studies. For large-scale training samples, it will waste a substantial amount of machine storage and operation time.

3. 4　Multivariate adaptive regression splines

MARS is a data analysis method developed by the American statistician Friedman[70]. It is a nonparametric regression technique that combines the advantages of spline regression and recursive partitioning, and can be seen as an extension of the linear model of interaction between variables[71-79]. In MARS, no need to make assumptions about the relationship between independent and dependent variables. It divides the training set of data into different segmented line segments, which are called basis functions, and the endpoints of each segment are called nodes. It has been proved by many researches on LSM that MARS has strong adaptive capability and generalization ability when dealing with high-dimensional data[80-81]. MARS can be seen as a refinement of the effect of CART in regression problems, compared to other MLTs, it can effortlessly procure the display expressions of variables whereas model accuracy is not satisfactory.

3. 5　Random forest

In 2001, Breiman combined decision trees into random forests with the idea of ensemble learning[82], so it essentially belongs to a large branch of machine learning: ensemble learning method. In RF, assuming that the problem to be solved is classification, so each decision tree is a classifier. And for an input sample, RF integrates the different voting results of n classifiers, and the category with the most votes is designated as the final output, which is one of the simplest bagging ideas[83-86].

The "forest" is composed of "trees", and when constructing a decision tree, it is necessary to

pay attention to the random sampling and complete splitting process，and the construction method has the following three steps：

（1）There is a put back to select n samples which means that each time a sample is randomly selected and then put back to continue the selection，with these n samples training set a decision tree. And these n sets are taken as the sample at the root nodes of the decision tree；

（2）Suppose that the feature dimension of each sample is M. When the nodes of the decision tree need to be divided，m attributes are randomly selected，which need to satisfy the condition $m \ll M$. Then optimal one is selected as the splitting attribute from the m attributes.

（3）Split each node of the decision tree to the maximum extent possible，without pruning.

Repeat steps（1）-（3）to construct a large number of decision trees，so that a random forest is formed.

RF is highly efficient and capable of processing large amounts of data with very high accuracy，and does not require dimensionality reduction in the face of input samples with high-dimensional features，so it is particularly applicable to explain the spatial relationship between landslide cases[87-89]. In addition，it can be used to evaluate the importance of individual features on the classification problems，which is very common in LSM[90-91]. In many current datasets，RF possesses a considerable advantage over other algorithms and can analyze the importance of the features. But since it presents a discontinuous output，the RF model performance in solving landslide susceptibility regression problems is not as impressive as on the classification problems.

3.6　Extreme gradient boosting

XGBoost is a tree boosting scalable machine learning system，proposed by Chen and Guestrin[92]. The learning method of this algorithm is mainly to sum the results of K（number of trees）

as the final predicted value，the forecast output of the ensemble model can be expressed as

$$\hat{y_i} = \phi(x_i) = \sum_{k=1}^{K} f_k(x_i)，\qquad f_k \in F \quad (1)$$

where f_k represents the regression tree，K is the number of regression trees，f_k is the k-th tree model，x_i is the i-th sample，y_i indicates the i-th category label，$\hat{y_i}$ is the model predicted value，F is the space of CART.

In fact，XGBoost is known for its regularized boosting technique，the objective function is defined in order to learn the set of functions used in the model，and prevent overfitting. Generally，there are three main parameters namely，colsample_bytree（subsample ratio of columns when constructing each tree），subsample（subsample ratio of the training instance）and nrounds（max number of boosting iterations）to be selected when using the XGBoost model for landslide susceptibility assessment[93-97]. When XGBoost takes CART as the base learner，the addition of the regular term compensates for the disadvantage that the DT algorithm is prone to overfitting. The linear model can also be used as the base learner，when XGBoost is equivalent to logistic regression or linear regression with a regular term. Excepting for the disadvantage of being time-consuming，the XGBoost model is well received in landslide susceptibility evaluation researches.

3.7　Comparative summary of MLTs

There are four steps in the process of constructing a landslide susceptibility model employing the machine learning algorithms：

（1）Obtain high quality spatial data such as topography and geological conditions from remote sensing images or landslide survey maps，with the aim of determining the triggering mechanism of landslides. It should be noted that different spatial resolutions have difference influence on various models adopted.

（2）The data analysis is carried out to optimize

the selection of suitable input factors. To date, there is still no universal guideline for determination of factors affecting landslides. Optimally selecting the input factors, ranking by importance, and removing factors with low contribution is a requisite. Landslide conditioning factors are required to be easily measured and estimated, and have the ability to perform functional calculations on them.

(3) The resampling method is implemented to supply the training set for training the different ML models. For some models, the performance relies not only on the quality of dataset and the assumptions of algorithm, but also on the optimization of the tuning of details.

(4) Evaluate the performance of different landslide susceptibility models, the fitting performance, prediction accuracy, and uncertainty of the results are all criteria for evaluating the model. In addition to the main measure accuracy (ACC), Receiver Operating Characteristic (ROC) curve and Cohen's kappa Index (kappa) are adopted as metrics to further assess the overall quality of the model.

For different study regions, different MLTs exhibit different model performance as well. Zhou et al.[98] classified the landslides in the Longju area as colluvial landslides and rockfalls, used SVM, ANN and LR for mapping landslide susceptibility. They found that the machine learning models outperformed the multivariate statistical models, and the SVM approach was more applicable to this case study. For the Zigui area, Zhang et al.[99] calculated the quantitative relationship between the landslide-conditioning factors and historical landslides using a RF model and found that the RF model achieved more accurately in LSM compared with the DT model; Wu et al.[100] used object-based data mining methods to study landslide susceptibility assessment, and the results showed that the object-based SVM model has a higher

accuracy than the others. Song et al.[101], Chen et al.[102] regarded the LSM problem as an imbalanced learning problem, built a more efficient prediction model using the weighted gradient boosting decision tree (weighted GBDT) method which is rarely used in previous researches. Besides, hybrid models were applied for LSM to build a more effective model and improve the adaptive capability of the model, such as the rough sets-SVM (RS-SVM) model[40], the model combined rough sets with back-propagation neural networks (BPNNs)[103], and the two steps self-organizing mapping-random forest (two steps SOM-RF) model[104].

Overall, the methods described above applied efficiently for solving many LSM problems in many individual case studies. Table 2 summarizes the applicable conditions, merits and disadvantages of the MLTs introduced in this study. Compared with the traditional ML method, the tree-based ensemble algorithms input most of the data without tedious data preprocessing process. Considering the sufficiency and accuracy of the landslide susceptibility models, the linear models classify the data in a sparse manner. By contrast, the tree-based model can better summarize the overall data structure and map the relationship between variables, and its results are less prone to overfitting problems and better handling of outliers. According to the study about the Three Gorges Reservoir area, the overall accuracy of the model was compared as a comparative index, and it was found that the accuracy of traditional linear models such as LR was in the range of 80.5%-88.3%, the accuracy of SVM model was 86.4%-93.6%, while the accuracy of RF algorithm model basically remained around 90.65%, and the accuracy of GBDT model was as high as 97.7%-99.7%. Therefore, a preliminary conclusion informed that for the Three Gorges Reservoir area, the tree-based algorithms achieve superiorly.

Table 2   Summary of the applicable conditions, merits and disadvantages for the MLTs

MLT	Applicable condition	Merit	Disadvantage
ANN	Systems with large data size, complex structure and unclear information	Strong ability of distributed storage and learning, high accuracy of classification	The learning process cannot be observed and the output is difficult to interpret as a black box model
DT	Explain the relationship between complex variables, make predictions on small data sets	Simple calculation, insensitive to intermediate missing values	Prone to overfitting
SVM	Small and medium datasamples, nonlinearity and high dimensionality classification problems	Improve generalization ability ofmodels	Sensitivity to missing data
MARS	Data system with large volume and high dimensionality	Ability to explain nonlinearity and interaction between variables	Susceptible to overfitting, low accuracy for sparse data
RF	Data system with large volume and high dimensionality	Handles thousands of input variables without variable deletion, estimates the importance of variables in the classification	Less accurate for low-dimensional data
XGBoost	Data system with large volume and high dimensionality	Uses regularized boosting technique to prevent overfitting, supports for column sampling	Susceptible to adjusting parameters issues

4   Discussion and conclusion

In the IT era, computer technology is constantly evolving, artificial intelligence is greatly improved, and machine learning has gradually covered the whole social life, which includes medical, transportation machinery, public safety, social science, disaster management and so on. This study focused on the research field of solving landslide susceptibility problems with machine learning techniques, and a systematic review of the relevant research in the past 20 years was presented. From the results of statistical analysis, it can be seen that as an emerging research method, the relevant publications are increasing year by year and rising rapidly, so it is obvious that the application of MLTs in LSM is effective.

In addition to the mentioned ANN, DT, SVM, MARS, RF, XGBoost methods, machine learning also encompasses other types of algorithms, and as research progresses, many hybrid methods have been proposed by worldwide researchers to improve the generalization ability of the models. These techniques have very powerful self-learning capability and can handle a large amount of data efficiently and accurately. They are mostly used for failure probability analysis of landslides and landslide displacement predictions, building landslide susceptibility maps to help make risk mitigation decisions. While reviewing the related research, it is important to note that this research area may face many challenges in the future as discussed below:

(1) It is difficult to conduct susceptibility analysis if the study area is large. On the one hand, data collection is a major problem, and obviously the lack of landslide information cannot build a good performance model. On the other hand, the predicted results of landslide high susceptibility areas may only account for a small portion of the total map, so that the accuracy of the susceptibility model cannot be guaranteed.

(2) Among the many cases used in literatures, there are different types of landslides with diverse geographical information and environmental conditions, so it is important to understand the types of landslides in the study area and then build an optimal model based on the assumptions and theoretical basis of different machine learning methods, which can enhance the credibility of the landslide area classification.

(3) The influence factors considered when

performing landslide susceptibility modeling using machine learning techniques are not comprehensive. In current studies，most of them focus on the geotechnical properties，slope，height，etc. of landslides. But in fact，uninvolved factors often have significant effects，such as climate and environmental changes，which are difficult to observe，and destructive to landslides. In addition，the role of groundwater and weathering erosion are often ignored.

（4）Few studies consider the occurrence and interactions of multiple geologic hazards in complex environments，such as rockfalls，debris flows，ground subsidence or landslide events that have occurred with unpredictable impacts. When analyzing landslide susceptibility in the study area，the performance of susceptibility maps can be improved if other hazards are modeled and predicted simultaneously in a cross-sectional study.

Previous researches have confirmed that MLTs can accurately predict landslides to reduce damage caused by disasters，and this technique has helped to make scientific decisions on disaster management. Moreover，authors hold the view that the application of MLTs in LSM is still evolving as an emerging research direction. Especially in recent years，deep learning，a derivative of machine learning，has received much attention，but the use of deep learning methods to develop superior landslide susceptibility models has not been widely investigated and reported，it is not difficult to imagine that such applications in the future are promising.

References：

［1］ VARNES D J，CRUDEN D M. Landslide types and processes ［M］// TURNER A K，SCHUSTER RL. Landslides： Investigation and mitigation： Transportation Reseach Boord Special Report 247，National Research Council，wasimgton，D. C. National Academy Press，1996：247.

［2］ LAN H X，ZHOU C H，WANG L J，et al. Landslide hazard spatial analysis and prediction using GIS in the Xiaojiang watershed，Yunnan，China ［J］. Engineering Geology，2004，76(1/2)：109-128.

［3］ GUZZETTI F，MONDINI A C，CARDINALI M，et al. Landslide inventory maps： New tools for an old problem ［J］. Earth-Science Reviews，2012，112(1/2)：42-66.

［4］ CHIGIRA M，WU X Y，INOKUCHI T，et al. Landslides induced by the 2008 Wenchuan earthquake，Sichuan，China ［J］. Geomorphology，2010，118(3/4)：225-238.

［5］ ZHANG W G，CHING J，GOH A T C，et al. Big data and machine learning in geoscience and geoengineering： Introduction ［J］. Geoscience Frontiers，2021，12(1)：327-329.

［6］ BAI S B，LÜ G，WANG J，et al. GIS-based rare events logistic regression for landslide-susceptibility mapping of Lianyungang，China ［J］. Environmental Earth Sciences，2011，62(1)：139-149.

［7］ XU C，DAI F C，XU X W，et al. GIS-based support vector machine modeling of earthquake-triggered landslide susceptibility in the Jianjiang River watershed，China ［J］. Geomorphology，2012，145/146：70-80.

［8］ CHEN W，PENG J B，HONG H Y，et al. Landslide susceptibility modelling using GIS-based machine learning techniques for Chongren County，Jiangxi Province，China ［J］. Science of the Total Environment，2018，626：1121-1135.

［9］ HONG H Y，LIU J Z，BUI D T，et al. Landslide susceptibility mapping using J48 Decision Tree with AdaBoost，Bagging and Rotation Forest ensembles in the Guangchang area（China）［J］. CATENA，2018，163：399-413.

［10］ WANG Y，FANG Z C，HONG H Y. Comparison of convolutional neural networks for landslide susceptibility mapping in Yanshan County，China ［J］. Science of the Total Environment，2019，666：975-993.

［11］ GUZZETTI F，GALLI M，REICHENBACH P，et al. Landslide hazard assessment in the Collazzone area，Umbria，Central Italy ［J］. Natural Hazards and Earth System Sciences，2006，6(1)：115-131.

［12］ COROMINAS J，WESTEN C，FRATTINI P，et al. Recommendations for the quantitative analysis of landslide risk ［J］. Bulletin of Engineering Geology and the Environment，2014，73(2)：209-263.

［13］ LUN N K，LIEW M S，MATORI A N，et al. Recent developments in machine learning applications in landslide susceptibility mapping［J］. AIP Conference Proceedings，2017，1905(1):040022.

［14］ LI C D，TANG H M，GE Y F，et al. Application of back-propagation neural network on bank destruction forecasting for accumulative landslides in the three Gorges Reservoir Region，China［J］. Stochastic Environmental Research and Risk Assessment，2014，28(6):1465-1477.

［15］ NGUYEN V，PHAM B，VU B，et al. Hybrid machine learning approaches for landslide susceptibility modeling［J］. Forests，2019，10(2):157.

［16］ COLKESEN I，SAHIN E K，KAVZOGLU T. Susceptibility mapping of shallow landslides using kernel-based Gaussian process，support vector machines and logistic regression［J］. Journal of African Earth Sciences，2016，118:53-64.

［17］ POUDYAL C P，CHANG C D，OH H J，et al. Landslide susceptibility maps comparing frequency ratio and artificial neural networks:A case study from the Nepal Himalaya［J］. Environmental Earth Sciences，2010，61(5):1049-1064.

［18］ PASCALE S，PARISI S，MANCINI A，et al. Landslide susceptibility mapping using artificial neural network in the urban area of Senise and San Costantino Albanese (Basilicata， southern Italy)［C］// Computational Science and Its Applications-ICCSA，2013:473-488.

［19］ JEBUR M N，PRADHAN B，TEHRANY M S. Manifestation of LiDAR-derived parameters in the spatial prediction of landslides using novel ensemble evidential belief functions and support vector machine models in GIS［J］. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing，2015，8(2):674-690.

［20］ CHEN W，XIE X S，WANG J L，et al. A comparative study of logistic model tree，random forest，and classification and regression tree models for spatial prediction of landslide susceptibility［J］. CATENA，2017，151:147-160.

［21］ VALENCIA ORTIZ J A，MARTÍNEZ-GRAÑA A M. A neural network model applied to landslide susceptibility analysis (Capitanejo，Colombia)［J］. Geomatics，Natural Hazards and Risk，2018，9(1):1106-1128.

［22］ DOU J，YUNUS A P，BUI D T，et al. Improved landslide assessment using support vector machine with bagging，boosting，and stacking ensemble machine learning framework in a mountainous watershed，Japan［J］. Landslides，2020，17(3):641-658.

［23］ TIEN BUI D，TUAN T A，KLEMPE H，et al. 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［J］. Landslides，2016，13(2):361-378.

［24］ KAVZOGLU T，SAHIN E K，COLKESEN I. Landslide susceptibility mapping using GIS-based multi-criteria decision analysis，support vector machines，and logistic regression［J］. Landslides，2014，11(3):425-439.

［25］ KHOSRAVI K，PHAM B T，CHAPI K，et al. A comparative assessment of decision trees algorithms for flash flood susceptibility modeling at Haraz watershed，northern Iran［J］. Science of the Total Environment，2018，627:744-755.

［26］ CHEN W，ZHANG S，LI R W，et al. Performance evaluation of the GIS-based data mining techniques of best-first decision tree，random forest，and naïve Bayes tree for landslide susceptibility modeling［J］. Science of the Total Environment，2018，644:1006-1018.

［27］ GOETZ J N，BRENNING A，PETSCHKO H，et al. Evaluating machine learning and statistical prediction techniques for landslide susceptibility modeling［J］. Computers & Geosciences，2015，81:1-11.

［28］ HONG H Y，POURGHASEMI H R，POURTAGHI Z S. Landslide susceptibility assessment in Lianhua County (China):A comparison between a random forest data mining technique and bivariate and multivariate statistical models［J］. Geomorphology，2016，259:105-118.

［29］ DOU J，YUNUS A P，TIEN BUI D，et al. Assessment of advanced random forest and decision tree algorithms for modeling rainfall-induced landslide susceptibility in the Izu-Oshima Volcanic Island，Japan［J］. Science of the Total Environment，2019，662:332-346.

［30］ CHOI J，OH H J，WON J S，et al. Validation of an artificial neural network model for landslide susceptibility mapping［J］. Environmental Earth Sciences，2010，60(3):473-483.

［31］BUI D T，TSANGARATOS P，NGUYEN V T，et al. Comparing the prediction performance of a Deep Learning Neural Network model with conventional machine learning models in landslide susceptibility assessment［J］. CATENA，2020，188：104426.

［32］PHAM B T，TIEN BUI D，PRAKASH I，et al. Hybrid integration of Multilayer Perceptron Neural Networks and machine learning ensembles for landslide susceptibility assessment at Himalayan area（India）using GIS［J］. CATENA，2017，149：52-63.

［33］RAHMATI O，FALAH F，DAYAL K S，et al. Machine learning approaches for spatial modeling of agricultural droughts in the south-east region of Queensland Australia ［J］. Science of the Total Environment，2020，699：134230.

［34］WANG F W，ZHANG Y M，HUO Z T，et al. The July 14，2003 Qianjiangping landslide，Three Gorges Reservoir，China ［J］. Landslides，2004，1（2）：157-162.

［35］JIAN W X，WANG Z J，YIN K L. Mechanism of the Anlesi landslide in the Three Gorges Reservoir，China ［J］. Engineering Geology，2009，108(1/2)：86-95.

［36］HUANG B L，YIN Y P，LIU G N，et al. Analysis of waves generated by Gongjiafang landslide in Wu Gorge，Three Gorges Reservoir，on November 23，2008［J］. Landslides，2012，9(3)：395-405.

［37］TANG H M，LI C D，HU X L，et al. Deformation response of the Huangtupo landslide to rainfall and the changing levels of the Three Gorges Reservoir ［J］. Bulletin of Engineering Geology and the Environment，2015，74(3)：933-942.

［38］HUANG D，GU D M，SONG Y X，et al. Towards a complete understanding of the triggering mechanism of a large reactivated landslide in the Three Gorges Reservoir ［J］. Engineering Geology，2018，238：36-51.

［39］YIN Y P，WANG H D，GAO Y L，et al. Real-time monitoring and early warning of landslides at relocated Wushan Town，the Three Gorges Reservoir，China ［J］. Landslides，2010，7(3)：339-349.

［40］PENG L，NIU R Q，HUANG B，et al. Landslide susceptibility mapping based on rough set theory and support vector machines：A case of the Three Gorges area，China ［J］. Geomorphology，2014，204：287-301.

［41］ KAWABATA D， BANDIBAS J. Landslide susceptibility mapping using geological data，a DEM from ASTER images and an Artificial Neural Network（ANN）［J］. Geomorphology，2009，113 (1/2)：97-109.

［42］TIAN Y Y，XU C，HONG H Y，et al. Mapping earthquake-triggered landslide susceptibility by use of artificial neural network（ANN）models：An example of the 2013 Minxian（China）Mw 5. 9 event ［J］. Geomatics，Natural Hazards and Risk，2019，10(1)：1-25.

［43］ZHANG W G，GOH A T C. Multivariate adaptive regression splines and neural network models for prediction of pile drivability［J］. Geoscience Frontiers，2016，7(1)：45-52.

［44］GOH A T C，ZHANG R H，WANG W，et al. Numerical study of the effects of groundwater drawdown on ground settlement for excavation in residual soils［J］. Acta Geotechnica，2020，15(5)：1259-1272.

［45］CHEN L L，ZHANG W G，GAO X C，et al. Design charts for reliability assessment of rock bedding slopes stability against bi-planar sliding：SRLEM and BPNN approaches［J］. Georisk：Assessment and Management of Risk for Engineered Systems and Geohazards，2020：1-16.

［46］LEE S，RYU J H，WON J S，et al. Determination and application of the weights for landslide susceptibility mapping using an artificial neural network ［J］. Engineering Geology，2004，71(3/4)：289-302.

［47］NGUYEN H，MEHRABI M，KALANTAR B，et al. Potential of hybrid evolutionary approaches for assessment of geo-hazard landslide susceptibility mapping［J］. Geomatics，Natural Hazards and Risk，2019，10(1)：1667-1693.

［48］HARMOUZI H，NEFESLIOGLU H A，ROUAI M，et al. Landslide susceptibility mapping of the Mediterranean coastal zone of Morocco between Oued Laou and El Jebha using artificial neural networks（ANN）［J］. Arabian Journal of Geosciences，2019，12 (22)：1-18.

［49］MOAYEDI H，MEHRABI M，MOSALLANEZHAD M，et al. Modification of landslide susceptibility mapping using optimized PSO-ANN technique［J］. Engineering with Computers，2019，35(3)：967-984.

［50］SAITO H，NAKAYAMA D，MATSUYAMA H. Comparison of landslide susceptibility based on a

decision-tree model and actual landslide occurrence: The Akaishi Mountains, Japan [J]. Geomorphology, 2009, 109(3/4): 108-121.

[51] WITTEN I H, FRANK E. Data mining: Practical machine learning tools and techniques [J]. ACM Sigmod Record, 2011, 31(1):76-77.

[52] ZHANG R H, WU C Z, GOH A T C, et al. Estimation of diaphragm wall deflections for deep braced excavation in anisotropic clays using ensemble learning [J]. Geoscience Frontiers, 2021, 12(1): 365-373.

[53] ZHANG W G, LI H R, LI Y Q, et al. Application of deep learning algorithms in geotechnical engineering: A short critical review [J]. Artificial Intelligence Review, 2021: 1-41.

[54] BOU KHEIR R, CHOROWICZ J, ABDALLAH C, et al. Soil and bedrock distribution estimated from gully form and frequency: A GIS-based decision-tree model for Lebanon [J]. Geomorphology, 2008, 93 (3/4): 482-492.

[55] LOH W Y, SHIN Y S. Split selection methods for classification trees [J]. Statistica Sinica, 1997, 7(4): 815-840.

[56] SONG Y S, CHAE B G. Development to prediction technique of slope hazards in gneiss area using decision tree model [J]. The Journal of Engineering Geology, 2008, 18(1): 45-54.

[57] QUINLAN J R. C4. 5: Programs for machine learning [M]. Elsevier, 2014.

[58] YEON Y K, HAN J G, RYU K H. Landslide susceptibility mapping in Injae, Korea, using a decision tree [J]. Engineering Geology, 2010, 116(3/4): 274-283.

[59] BREIMAN L, FRIEDMAN J, STONE C J, et al. Classification and regression trees [M]. CRC Press, 1984.

[60] YAO X, THAM L G, DAI F C. Landslide susceptibility mapping based on support vector machine: A case study on natural slopes of Hong Kong, China [J]. Geomorphology, 2008, 101 (4): 572-582.

[61] VAPNIK V N. The nature of statistical learning theory [M]. Springer Science & Business Media, 2013.

[62] BUI D T, PRADHAN B, LOFMAN O, et al. Landslide susceptibility assessment in Vietnam using support vector machines, decision tree, and naive bayes models [J]. Mathematical Problems in Engineering, 2012, 2012:1-26.

[63] FROHLICH H, CHAPELLE O, SCHOLKOPF B. Feature selection for support vector machines by means of genetic algorithm [C]//Proceedings of 15th IEEE International Conference on Tools with Artificial Intelligence, November 5, 2003, Sacramento, CA, USA. IEEE, 2003: 142-148.

[64] TEHRANY M S, PRADHAN B, JEBUR M N. Flood susceptibility mapping using a novel ensemble weights-of-evidence and support vector machine models in GIS [J]. Journal of Hydrology, 2014, 512: 332-343.

[65] ZHANG W G, LI H R, WU C Z, et al. Soft computing approach for prediction of surface settlement induced by earth pressure balance shield tunneling [J]. Underground Space, 2021, 6(4): 353-363.

[66] ZHANG W G, ZHANG R H, WU C Z, et al. State-of-the-art review of soft computing applications in underground excavations [J]. Geoscience Frontiers, 2020, 11(4): 1095-1106.

[67] HUANG C, DAVIS L S, TOWNSHEND J R G. An assessment of support vector machines for land cover classification [J]. International Journal of Remote Sensing, 2002, 23(4): 725-749.

[68] BRENNING A. Spatial prediction models for landslide hazards: Review, comparison and evaluation [J]. Natural Hazards and Earth System Sciences, 2005, 5 (6): 853-862.

[69] KALANTAR B, PRADHAN B, NAGHIBI S A, et al. Assessment of the effects of training data selection on the landslide susceptibility mapping: A comparison between support vector machine (SVM), logistic regression (LR) and artificial neural networks (ANN) [J]. Geomatics, Natural Hazards and Risk, 2018, 9 (1): 49-69.

[70] FRIEDMAN J H. Multivariate adaptive regression splines [J]. The Annals of Statistics, 1991, 19(1): 1-67.

[71] ZHANG W G, GOH A T C, ZHANG Y M. Multivariate adaptive regression splines application for multivariate geotechnical problems with big data [J]. Geotechnical and Geological Engineering, 2016, 34 (1): 193-204.

[72] ZHANG W G, ZHANG Y M, GOH A T C. Multivariate adaptive regression splines for inverse analysis of soil and wall properties in braced excavation

［J］. Tunnelling and Underground Space Technology，2017，64：24-33.

［73］ ZHANG W G，ZHANG R H，GOH A T C. Multivariate adaptive regression splines approach to estimate lateral wall deflection profiles caused by braced excavations in clays ［J］. Geotechnical and Geological Engineering，2018，36(2)：1349-1363.

［74］ ZHANG W G，ZHANG R H，GOH A T C. MARS inverse analysis of soil and wall properties for braced excavations in clays ［J］. Geomechanics and Engineering，2018，16(6)：577-588.

［75］ ZHANG W G，GOH A T C，ZHANG Y M，et al. Assessment of soil liquefaction based on capacity energy concept and multivariate adaptive regression splines ［J］. Engineering Geology，2015，188：29-37.

［76］ ZHANG W G，GOH A T C. Multivariate adaptive regression splines for analysis of geotechnical engineering systems ［J］. Computers and Geotechnics，2013，48：82-95.

［77］ GOH A T C，ZHANG W G，ZHANG Y M，et al. Determination of earth pressure balance tunnel-related maximum surface settlement：A multivariate adaptive regression splines approach ［J］. Bulletin of Engineering Geology and the Environment，2018，77(2)：489-500.

［78］ GOH A T C，ZHANG Y M，ZHANG R H，et al. Evaluating stability of underground entry-type excavations using multivariate adaptive regression splines and logistic regression ［J］. Tunnelling and Underground Space Technology，2017，70：148-154.

［79］ ZHANG W G，ZHANG R H，WANG W，et al. A Multivariate Adaptive Regression Splines model for determining horizontal wall deflection envelope for braced excavations in clays ［J］. Tunnelling and Underground Space Technology，2019，84：461-471.

［80］ BRIAND L C，FREIMUT B，VOLLEI F. Using multiple adaptive regression splines to support decision making in code inspections ［J］. Journal of Systems and Software，2004，73(2)：205-217.

［81］ VORPAHL P，ELSENBEER H，MÄRKER M，et al. How can statistical models help to determine driving factors of landslides? ［J］. Ecological Modelling，2012，239：27-39.

［82］ BREIMAN L. Random forests ［J］. Machine Learning，2001，45(1)：5-32.

［83］ CATANI F，LAGOMARSINO D，SEGONI S，et al. Landslide susceptibility estimation by random forests technique：Sensitivity and scaling issues ［J］. Natural Hazards and Earth System Sciences，2013，13(11)：2815-2831.

［84］ ZHANG W G，WU C Z，LI Y Q，et al. Assessment of pile drivability using random forest regression and multivariate adaptive regression splines ［J］. Georisk：Assessment and Management of Risk for Engineered Systems and Geohazards，2021，15(1)：27-40.

［85］ ZHANG W G，WU C Z，ZHONG H Y，et al. Prediction of undrained shear strength using extreme gradient boosting and random forest based on Bayesian optimization ［J］. Geoscience Frontiers，2021，12(1)：469-477.

［86］ ZHANG W G，LI Y Q，WU C Z，et al. Prediction of lining response for twin tunnels constructed in anisotropic clay using machine learning techniques ［J/OL］. Underground Space，https：//doi. org/10. 1016/j. undsp. 2020. 02. 007

［87］ KIM J C，LEE S M，JUNG H S，et al. Landslide susceptibility mapping using random forest and boosted tree models in Pyeong-Chang，Korea ［J］. Geocarto International，2018，33(9)：1000-1015.

［88］ TAALAB K，CHENG T，ZHANG Y. Mapping landslide susceptibility and types using Random Forest ［J］. Big Earth Data，2018，2(2)：159-178.

［89］ HONG H Y，MIAO Y M，LIU J Z，et al. Exploring the effects of the design and quantity of absence data on the performance of random forest-based landslide susceptibility mapping ［J］. CATENA，2019，176：45-64.

［90］ SHAFIZADEH-MOGHADAM H，MINAEI M，SHAHABI H，et al. Big data in Geohazard：pattern mining and large scale analysis of landslides in Iran ［J］. Earth Science Informatics，2019，12(1)：1-17.

［91］ YOUSSEF A M，POURGHASEMI H R. Landslide susceptibility mapping using machine learning algorithms and comparison of their performance at Abha Basin，Asir Region，Saudi Arabia ［J］. Geoscience Frontiers，2021，12(2)：639-655.

［92］ CHEN T Q，GUESTRIN C. XGBoost：A scalable tree boosting system ［C］//Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco，California，USA. New York，NY，USA：ACM，2016：785-794.

［93］ ZHANG W G，ZHANG R H，WU C Z，et al.

Assessment of basal heave stability for braced excavations in anisotropic clay using extreme gradient boosting and random forest regression ［J］. Underground Space，https：//doi. org/10. 1016/j. undsp. 2020. 03. 001

［94］ WANG L，WU C Z，TANG L B，et al. Efficient reliability analysis of earth dam slope stability using extreme gradient boosting method ［J］. Acta Geotechnica，2020，15(11)：3135-3150.

［95］ SAHIN E K. Comparative analysis of gradient boosting algorithms for landslide susceptibility mapping ［J］. Geocarto International，2020：1-25.

［96］ SAHIN E K. Assessing the predictive capability of ensemble tree methods for landslide susceptibility mapping using XGBoost，gradient boosting machine，and random forest ［J］. SN Applied Sciences，2020，2 (7)：1-17.

［97］ STANLEY T A，KIRSCHBAUM D B，SOBIESZCZYK S，et al. Building a landslide hazard indicator with machine learning and land surface models ［J］. Environmental Modelling & Software，2020，129：104692.

［98］ ZHOU C，YIN K L，CAO Y，et al. Landslide susceptibility modeling applying machine learning methods：A case study from Longju in the Three Gorges Reservoir area，China ［J］. Computers & Geosciences，2018，112：23-37.

［99］ ZHANG K X，WU X L，NIU R Q，et al. The assessment of landslide susceptibility mapping using random forest and decision tree methods in the Three Gorges Reservoir area，China ［J］. Environmental Earth Sciences，2017，76(11)：1-20.

［100］ WU X L，REN F，NIU R Q. Landslide susceptibility assessment using object mapping units，decision tree，and support vector machine models in the Three Gorges of China ［J］. Environmental Earth Sciences，2014，71(11)：4725-4738.

［101］ SONG Y X，NIU R Q，XU S L，et al. Landslide susceptibility mapping based on weighted gradient boosting decision tree in Wanzhou section of the Three Gorges reservoir area （China） ［J］. ISPRS International Journal of Geo-Information，2018，8(1)：4.

［102］ CHEN T，ZHU L，NIU R Q，et al. Mapping landslide susceptibility at the Three Gorges Reservoir，China，using gradient boosting decision tree，random forest and information value models ［J］. Journal of Mountain Science，2020，17(3)：670-685.

［103］ WU X L，NIU R Q，REN F，et al. Landslide susceptibility mapping using rough sets and back-propagation neural networks in the Three Gorges，China ［J］. Environmental Earth Sciences，2013，70 (3)：1307-1318.

［104］ LONG J J，LIU Y，LI C D，et al. A novel model for regional susceptibility mapping of rainfall-reservoir induced landslides in Jurassic slide-prone strata of western Hubei Province，Three Gorges Reservoir area ［J］. Stochastic Environmental Research and Risk Assessment，2020：1-24.

（编辑　黄廷）