



Application and comparison of different ensemble learning machines combining with a novel sampling strategy for shallow landslide susceptibility mapping

Zhu Liang¹ · Changming Wang¹ · Kaleem Ullah Jan Khan¹

Accepted: 30 September 2020 / Published online: 9 October 2020
© Springer-Verlag GmbH Germany, part of Springer Nature 2020

Abstract

The existence of shallow landslide brings huge threats to the human lives and economic development, as the Lang County, Southeastern Tibet prone to landslide. Landslide susceptibility mapping (LSM) is considered as the key for the prevention of hazard. The primary goal of the present study is to assess and compare four models: classification and regression tree, gradient boosting decision tree (GBDT), adaptive boosting-decision tree and random forest for the performance of landslide susceptibility modeling. Firstly, a landslide inventory map consisting of 229 historical shallow landslide locations was prepared and the same number of non-landslide points was determined by k-means clustering. Secondly, 12 conditioning factors were considered in the landslide susceptibility modeling. The prediction performance of the four models were estimated by fivefold cross validation and relative operating characteristic curve (ROC), area under the ROC curve (AUC) and statistical measures. The results showed that the GBDT performed best in the training and validation dataset, with the highest prediction capability (AUC = 0.986 and 0.940), highest accuracy value (95.3% and 88.1%) and highest kappa index (0.904 and 0.772), respectively. Therefore, the GBDT was considered to be the most suitable model and applied to the whole study area for LSM. The results of this study also demonstrate that the performance can be enhanced with the use of ensemble learning. The sampling strategy of non-landslide points can be improved by combining with clustering analysis which are more reasonable.

Keywords Shallow landslide · Susceptibility · Ensemble learning · K-means clustering

1 Introduction

In geomorphology, a “landslide” is the movement of a mass of rock, debris or earth down a slope, under the influence of gravity (Cruden and Varnes 1996). According to different variables, landslides can be divided into different types (Varnes 1978). Rainfall-induced shallow landslides is a natural phenomenon mainly occurring in mountainous areas which are widespread all over the world and caused damages on both human lives and economy directly or indirectly (Trigila and Iadanza 2012). Generally, damages can be decreased to a certain extent by predicting

the likely location of future disasters (Pradhan 2010). Therefore, landslide susceptibility prediction is the first step towards estimation or reduction of landslide hazard and risk.

The effectiveness of LSM depends greatly on the modeling methodology adopted. The approaches of LSM can be broadly classified as qualitative and quantitative (Liang et al. 2020a), which are also known as direct or indirect methods. Indirect approach like machine learning techniques such as neural networks, support vector machines, decision tree and logistic regression, have been used over large-scale study areas (Liang et al. 2020b; Merghadi et al. 2020; Lian et al. 2014; Colkesen et al. 2016). However, the performance of different learning machines fluctuates in different situations because of the inductive preference, which is an essential attribute of machine learning algorithms. Most single classifiers, of which generalization ability is slightly better than that of

✉ Changming Wang
wangcm@jlu.edu.cn

¹ College of Construction Engineering, Jilin University, Changchun 130000, People's Republic of China

random guessing, are less robust. There is a room for improvement and the emergence of ensemble learning made it possible (Woods and Kaatupitiya 2013).

The goal of ensemble learning methods is to construct a collection (an ensemble) of individual classifiers to reduce predictive error and improve the performance (Fan et al. 1999; Dietterich 2000). Due to the superior prediction capacity with high robustness, more and more ensemble learning approaches were applied and explored in LSM study (Bui et al. 2014; Youssef et al. 2015a, b; Pham and Prakash 2019). Boosting and Bagging were originally introduced by Schapire (1990) and Breiman (1994) respectively, which are two of the most popular and earliest techniques for machine learning and are usually used as the based idea of ensemble techniques (Dietterich 2000). However, these techniques rarely applied and discussed in LSM and need to be further exploration and comparison for a more improved model.

In general, the spatial prediction of geological disasters can be regarded as a binary classification process (Bennett et al., 2016). It is recommended to prepare equal amounts of disaster presence and absence observations (1:1 sampling) during the modeling process (Heckmann et al. 2014; Kornejady et al. 2017). In previous researches, negative samples were selected randomly or subjectively (Nefeslioglu et al. 2008; Hussin et al. 2015). However, there is significant uncertainty because the random points may be located in a highly prone area. Where there is a large amount of classification noise (i.e., training and test examples with incorrect class labels) it may cause overfitting or underfitting, resulting in low generalization ability of the model and affecting the accuracy of spatial prediction. High-purity samples are an important condition to ensure the performance of the model and clustering analysis is considered to be an effective method (Mingoti and Lima 2006).

The main objective of the present study is to explore and compare the potential application and performance of the decision tree and its three ensembles based on Boosting and Bagging for the selection of the best model for landslide susceptibility prediction. On the other hand, K-means clustering method is used to select negative samples that can better represent the characteristics of non-disasters areas and improve the purity of the samples. The Lang County in Southeastern Tibet chosen as the study area is prone to shallow landslides because of climatic and topographic conditions. Geomorphological, topographical and triggering factors were considered in modeling. The performance of the models were estimated by k-fold ($k = 5$ in this paper) cross validation using the receiver operating characteristic (ROC) curve, three statistical evaluation measures (Accuracy, Sensitivity, and Specificity) and kappa coefficient. Analysis and data processing has been

done using ArcMap 10.2, SPSS, SPSS Modeler 18.0 and Python 3.7 software.

2 Materials

2.1 Study area

The study area, Lang county was located in Nyingchi City, Southeastern Tibet, which is bounded by longitudes of $92^{\circ}25'$ E and $93^{\circ}31'$ E, latitudes of $28^{\circ}41'$ N and $29^{\circ}29'$ N (Fig. 1). It covers an area of about 4200 km^2 with a population of more than 15,000. The study area belongs to Qinghai-Tibet Plateau with the annual rainfall of $350 \sim 600 \text{ mm}$, mainly concentrated in June to August.

The study area belongs to the Yarlung Zangbo River deep fault zone in geological structure, with strong neotectonic activity and frequent earthquakes. The seismic intensity within the area has a degree of VIII on the modified Mercalli index. The strata is mainly composed of Jurassic, Triassic, Cretaceous and Quaternary.

The disasters in the study area mainly consist of rain-fed high frequency landslide, which pose a great threat to local villagers and engineering activities.

2.2 Data collection and preparation

2.2.1 Shallow landslide inventory

The statistically-based susceptibility models are based on an important assumption: future landslides will be more likely to occur under the conditions which led to the landslides past and present (Varnes 1984; Furlani and Ninfo 2015). Therefore, landslide inventory mapping as the initial step is essential. In this study, data comes from historical landslide records (from 1970 to 2010), field surveys (from 2000 to 2003) (Figs. 2 and 3) and Google Earth satellite images interpretation (20) (Fig. 4). All landslide inventories consist of polygons which bound the whole landslide perimeter and finally, a total of 229 landslide polygons were obtained. In this study, the identified landslides were shallow type according to Varnes classification system (1978), which were triggered by Rainfall.

2.2.2 Conditioning factors

Occurrence of landslides is controlled by multiple factors such as geomorphological, topographical and triggering factors (including natural and human factors). According to availability, reliability, and practicality of the data (van Westen et al. 2008), 12 landslide conditioning factors including annual rainfall (F1), maximum elevation difference (F2), altitude (F3), plan curvature (F4), profile

Fig. 1 Location map of the study area showing landslide inventory

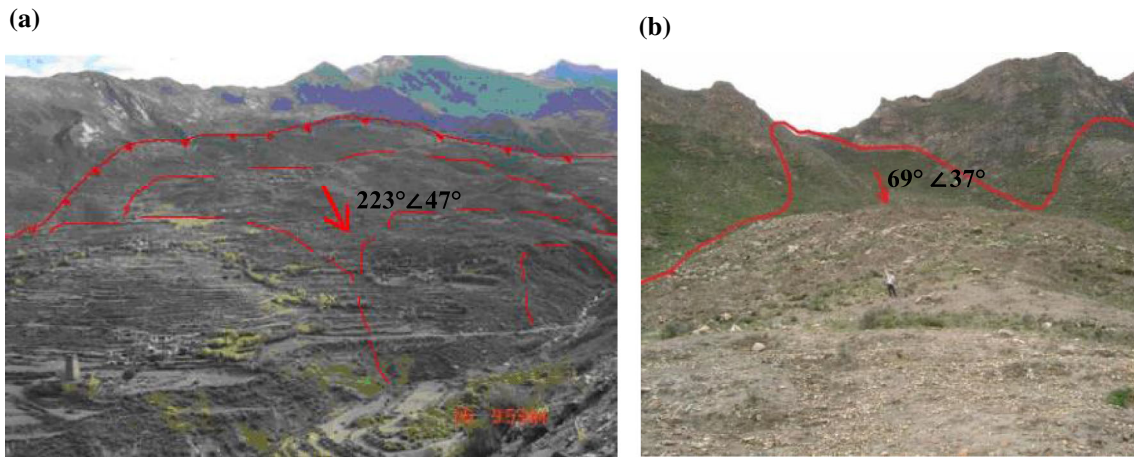
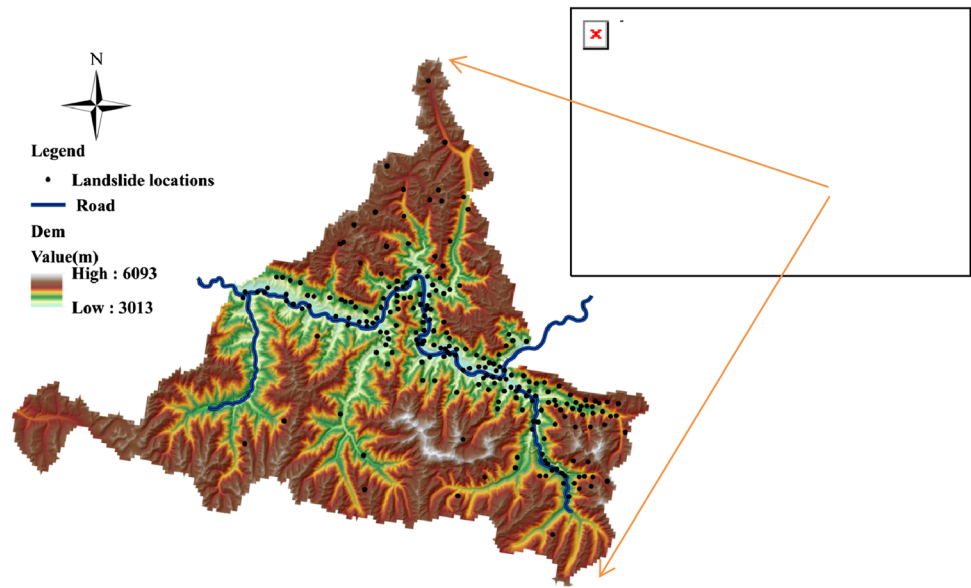


Fig. 2 Field investigation photos. **a** landslide in Jindong township; **b** traction landslide in Zhuogang Village

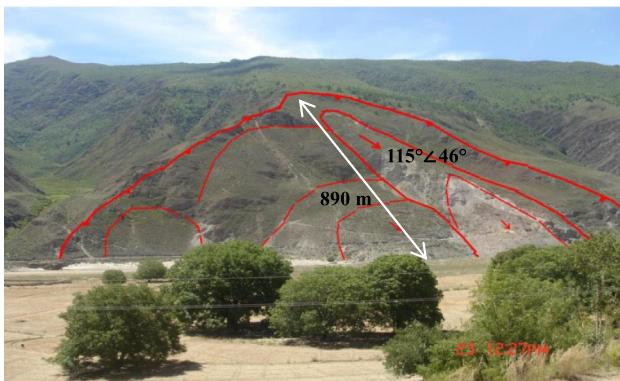


Fig. 3 The panorama of Jigoshan landslide

curvature (F5), slope angle (F6), topographic wetness index (F7), distance to roads (F8), distance to faults (F9), distance

to streams (F10), slope aspect (F11), and lithology (F12) were involved for landslide susceptibility modeling.

Topographic related factors (altitude, slope angle, slope aspect, maximum elevation difference, plan curvature, profile curvature and topographic wetness index) were derived from the (Digital Elevation Model) DEM with a resolution of 30×30 m from Shuttle Radar Topography Mission (SRTM) data. Altitude affect slope instability, precipitation properties and vegetation cover type, which was frequently used in LSM (Feizizadeh et al. 2014; Hong et al. 2015). Altitude was reclassified into 5 classes with an interval of 200 m (Fig. 5a). Slope angle was another major contributing factor, which controls the subsurface flow and the soil moisture (Magliulo et al. 2008). It was reclassified into 6 classed with an interval of 5° (Fig. 5b). Slope aspect map was reclassified into 8 classes according to the 8 cardinal directions (Fig. 5c). Maximum elevation

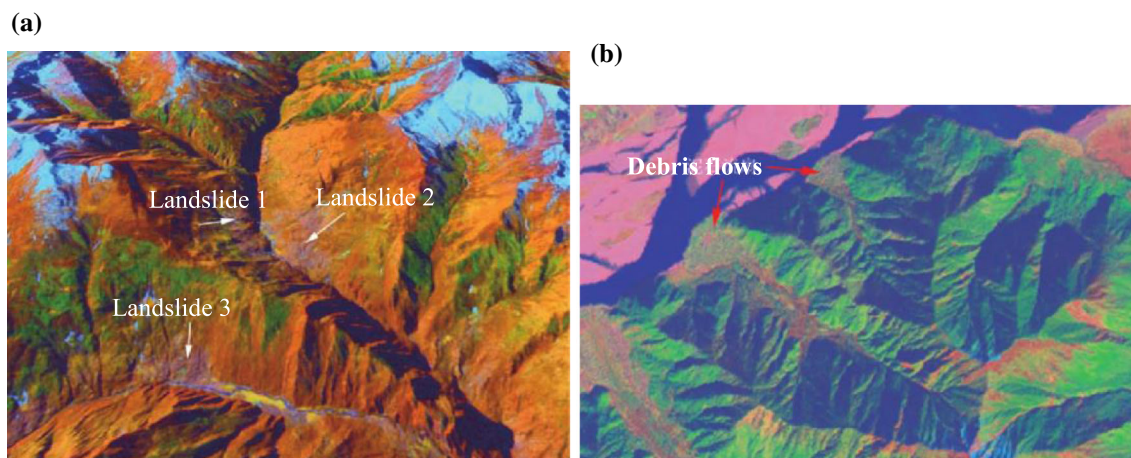


Fig. 4 Stereo remote sensing map of landslides in Lang country (Tong et al., 2019): **a** Landslides in Geda town; **b** Landslides in Zhongda town

difference was obtained by calculating the difference between the maximum and minimum values of elevation in the same slope unit. It was reclassified into 6 classes with an interval of 300 m (Fig. 5d). Curvature and topographic wetness index are morphometric parameters that represents basic terrain (Evans 1979). Topographic wetness index was reclassified into 6 classes (Fig. 5g). The plan curvature and profile curvature were reclassified into 7 classes (Figs. 5e, f).

Fault information were extracted from a geological map at a scale of 1:50,000. Faults decrease the rock strength, which act as potential weak planes in slopes (Bucci et al. 2016). The distance to faults maps with six classes (Fig. 5h) were constructed for < 3000, 3000–6000, 6000–9000, 9000–10,000, and > 10,000 m. Similarly, distance to roads (Fig. 5i) and distance to rivers (Fig. 5j) were constructed.

NDVI, which can be used to quantitatively estimate its impact on landslides and vegetation density (Chen et al. 2017), was prepared through information extracted from Landsat 8 LOI images (2016.5.11). It was reclassified into six classes (Fig. 5k).

Rainfall as the most important external factors inducing landslides was selected. Considering that the rainfall in the study area is mainly concentrated in June–August, the annual average rainfall was applied. In addition, the elevation will be used as a cooperative influence on rainfall. Consequently, the rainfall map (Fig. 5l) was constructed from the annual average of rainfall data (1981 ~ 2000) using the collaborative kriging method in the spatial interpolation function of ArcGIS by collecting data of 9 precipitation stations near the area under study as a reference.

2.2.3 Mapping units

The selection of the mapping unit is an important prerequisite for susceptibility modelling (Guzzetti, 2006a). The mapping units are generally divided into four categories: grid cells, slope units, unique-condition units and watershed units (Zeze et al. 2017) among which the most popular one used is grid cells (Reichenbach et al. 2018). The comparison of the four mapping units may be referred to another literature (Guzzetti et al. 2006b). Slope units can be expressed as an individual slope or a small catchment based on the landslide type and suitable for landslide susceptibility mapping at principal. Accordingly, slope units, which allowed to better characterize the detachment area and represent the geomorphological and geological conditions of landslides, were applied in this study. ArcGIS is used to divide the study area into 1460 slope units and make artificial corrections according to remote sensing image. Landslide locations (229) were represented as a single point per landslide (Fig. 1). On the other hand, each factor was reclassified into 4 to 7 classes based on the equal spacing principle and the mean value in the unit was counted as the representative value of the unit (Fig. 2).

3 Methods

3.1 K-means clustering

K-means is popular for cluster analysis in data mining due to its efficiency and simple implementation (Anil 2010; Likas et al. 2003). The k-means clustering aims to partition n observations into k clusters, in which each data point is assigned to the cluster with the nearest mean, thus serving as the centroid of the cluster (Hartigan and Wong 1979).

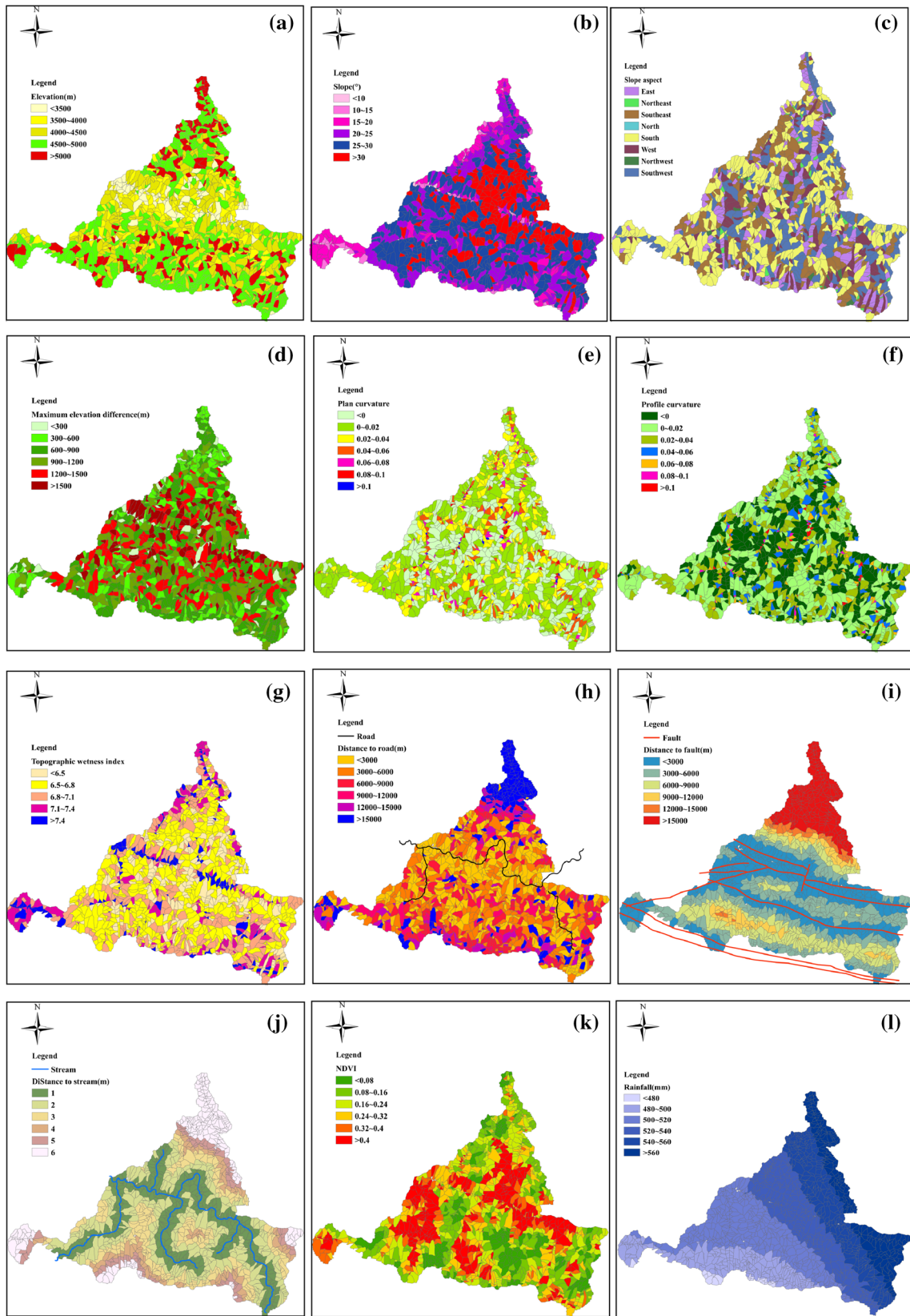


Fig. 5 Study area thematic maps: **a** Altitude; **b** Slope angle DTR; **c** Slope aspect; **d** MED; **e** Plan curvature; **f** Profile curvature; **g** TWI; **h** DTF; **i** DTR; **j** DTS; **k** NDVI; **l** Rainfall

The brief implementation process of K-means is as follows:

- (1) Predetermined initial clustering center;
- (2) Calculating the distance between each point and the initial clustering center (d_k) and determining a preliminary classification (MacQueen 1967);
- (3) Reacquiring the cluster centers of each new category and calculating the distance, and iterating repeatedly until the following equation is satisfied:

$$\frac{|u_{n+1} - u_n|}{u_{n+1}} \leq \varepsilon \quad (1)$$

where u_{n+1} represents the sum of squares of distances from each point to the cluster center after the n th clustering; ε represents the precision value.

In this paper, 229 positive samples (that is, disaster points) are used as the initial clustering center. The units farthest from the cluster center in each category were taken as a negative sample. The negative samples selected in this way will be representative and represent the characteristics of non-disaster areas. The final binary response variables consisted of an identical number of landslide presence and absence observations (1:1). Therefore, 229 negative samples were determined. Finally, the so-called cutoff value of the model result that discriminates between event and non-event is equal to 0.5.

3.2 Bagging and boosting

Bagging (also known as bootstrap aggregating), is one of the earliest ensemble methods introduced by Breiman (1994). The training set is constructed by multiple samples, which are generated from sampling randomly and replacement. Each subset is used to construct the predictive functions individually and then aggregated to make a final decision through voting procedure. It has been effectively used for generating classifier ensemble and accuracy is improved by controlling the variance of classification error. The out of bag (OOB) error is used to evaluate the generation ability of bagging and expressed as below:

$$E = \frac{1}{N} \sum_{C_i(x_j) \in N} \prod (L(X)^{OOB} \neq y) \quad (2)$$

where N represents the samples; X represents a vector and x represents variable; y represents the output; H represents the classifier and T is the number of classifiers; Π regards as 1 if it is true and 0 or else.

Boosting is another commonly used ensemble strategy and originally proposed by Schapire (1990). Boosting obtain sequentially training subsets by randomly sampling with replacement over weighted data. It provide controls

both bias and variance while Bagging performs better in terms of variance reducing (Jie et al. 2020). The errors of classifiers (E) is calculated as:

$$E = \sum_{x_j \in N, C_i(x_j) \neq y_j} Weight(x_j) / m \quad (3)$$

where m is the sequence of examples; N represents the samples; x_j and y_j represent the elements of samples.

In this paper, both Bagging and Boosting are adopted to compare and improve the performance of DT, which is a weak classifier, in landslide susceptibility modeling.

3.3 CART

CART is a non-parametric and non-linear technique and first introduced by Breiman et al. (1984). It divides the data into subsets based on independent factors and meanwhile, it splits a node into yes/no answers as predicting values (Pourghasemi and Rahmati 2018), which is appropriate for generating both regression and classification trees (Samadi et al. 2014). CART is a well-known DT algorithm and runs in SPSS Modeler in this study.

3.4 Random Forests (RF)

RF is a popular integrated method and first proposed by Breiman (2001). It is constructed based on two powerful ideas: Bagging and random feature selection (Wu et al. 2014; Fernández-Delgado et al. 2014). Bagging technique is applied to select the trees at each node, random samples of variables and observations as the training data set for model calibration. More detailed statistical explanation on RF can be found in Breiman (2001) and Segal (2004). In this study, the scikit-learn package (Pedregosa et al. 2011) in Python version 3.7 was used for RF modeling. Two hyper-parameters, the number of trees (k) and the number of variables applied to divided the nodes (m), are required to determined in advance (Youssef et al. 2016). To ensure the algorithm convergence and favorable prediction results, the number of trees k has been fixed to 500 and the number of predictive variables m has been selected as 5 (Breiman et al. 2001). The error was assessed by mean decrease in node impurity (mean decrease Gini index) (Calle and Urrea 2010).

3.5 GBDT

GBDT is another ensemble learning machine with CART as the base classifier and Gradient Boosting as the ensemble method. It is an iterative decision tree algorithm, which transforms several weak learning classifiers into a strong learning classifier with high precision (Friedman

2001). The parameters of the weak classifier are set to approximate the gradient direction of the loss function established in the previous iteration. The GBDT was applied in Python 3.7 using the GBDT class library of scikit-learn.

3.6 AdaBoost-DT

AdaBoost (known as adaptive boosting) is of the boosting family and was introduced by Freund and Schapire (1997). An adaptive resampling technique is applied to choose training samples and then the classifiers are trained iteratively while the misclassified samples are given more weight in each iteration. Accordingly, the final classifier is a weighted sum of the ensemble predictions obtained by each training (Hong et al. 2018). DT as a common weak classifier expect considerate performance once combined with AdaBoost algorithm. AdaBoost-DT is applied in Python 3.7 using the AdaBoost class library of scikit-learn.

3.7 Performance assessment

Models would have a poor scientific value without a proper evaluation and/or validation. (Chung and Fabbri 2003). In earlier studies, single hold-out model performance measures were derived using one training and independent test set (Reichenbach et al. 2018). However, there is a need for a more reliable estimation of the model performance. The ability of the models to classify independent test data was elaborated using a k-fold cross validation procedure ($k = 5$ in this paper), where the data is randomly partitioned into k disjoint sets, and one set at a time is used for model testing while the combined remaining $k-1$ sets are used for model training (James et al. 2013).

Accuracy, Sensitivity, and Specificity are the three statistical evaluation measures generally applied to assess the performance of the landslide susceptibility models (Bui et al. 2016a, b).

$$\begin{aligned} \text{Accuracy} &= \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \\ \text{Sensitivity} &= \frac{\text{TP}}{\text{TP} + \text{FN}} \\ \text{Specificity} &= \frac{\text{TN}}{\text{FP} + \text{TN}} \end{aligned} \quad (4)$$

where True Positives (TP), i.e., cells predicted unstable and observed unstable, True Negatives (TN), i.e., cells predicted stable and observed stable, False Positives (FP), i.e., cells predicted unstable but observed stable and False Negatives (FN), i.e., cells predicted stable but observed unstable.

The kappa index is the consistency test index of binary classification results and its value varies from -1 to 1 . It is

equal to 1 if the two results are in perfect agreement, and is equal to -1 if they are completely different. It can be determined based on the equation below:

$$\text{Kappa} = \frac{P_p - P_{\text{exp}}}{1 - P_{\text{exp}}} \quad (5)$$

where P_p represents the observed agreements and P_{exp} is the expected agreements.

Finally, the overall performance of the models is assessed by ROC curve, which illustrates the sensitivity (also recall) and 1-specificity for model validation. The Area Under Curve (AUC) is the most popular metrics to estimate the quality of model, varying from 0.5 (diagonal line) to 1 , with higher values indicating a better predictive capability of the model. (Green and Swets 1966; Swets 1988).

In present study, both ROC curve, the contingency tables and Kappa index were used to evaluate the susceptibility models established for landslide.

3.8 Mapping landslide susceptibility

The ultimate purpose of this paper is to select the best method for landslide susceptibility mapping. The simplest approach to select an optimal model for prediction is to compare multiple evaluation indicators from cross-validation, where the modeling method with the lowest error estimate is determined as the best one to use (Goetz et al. 2015). According to the landslide susceptibility index, the study area was reclassified into five classes such as very low ($0 \sim 0.2$), low ($0.2 \sim 0.4$), moderate ($0.4 \sim 0.6$), high ($0.6 \sim 0.8$) and very high ($0.8 \sim 1$) based on the equal spacing principle.

4 Results and verification

4.1 Evaluation and comparison of different models

The performance of the four models was evaluated and compared using the chosen indexes including ROC curve, the contingency tables and Kappa index. Analyses of the contingency tables and Kappa index using the training set are shown in Table 1. The GBDT model indicates the best performance for classifying landslides (sensitivity = 97.0%), followed by the RF model (sensitivity = 92.8%), the Ada-DT model (sensitivity = 89.7%) and CART model (sensitivity = 88.3%). In terms of the classification of non-landslides zones, the best performance is GBDT model (specificity = 93.8%), followed by the RF model (specificity = 93.1%), the Ada-DT model (specificity = 92.6%) and the CART model

Table 1 Models' performance using training dataset

Indexes	CART	GBDT	Ada-DT	RF
TP (%)	92.8	93.6	92.9	93.1
TN (%)	87.7	97.1	89.3	92.8
FP (%)	7.2	6.4	7.1	6.9
FN (%)	12.3	2.9	10.7	7.2
Sensitivity (%)	88.3	97.0	89.7	92.8
Specificity (%)	92.4	93.8	92.6	93.1
Accuracy (%)	90.2	95.3	91.1	93.0
Kappa	0.802	0.904	0.820	0.895

(specificity = 92.4%). In addition, the GBDT model also has the highest accuracy (95.3%) and Kappa coefficient (0.904), indicating good agreement between the models and the training data. The GBDT has higher values in these three parameters compared to those of the other three models. The performance of CART is the worst among the four models, but it presented satisfactory performance. Regarding the values of AUC, all the models achieve a great performance. Among the, the GBDT model has the greatest value of ACU (0.986), which is the same as the RF model. Similarly, the CART model performed the worst with the value of 0.944 (Table 2). Ensemble algorithms can effectively enhance performance accuracy of a single model. The standard error values of AUC were all less than 0.05 and the errors associated with probability estimation were not obvious (Fig. 6).

To confirm the practicability of the proposed models, the performance of models in the validation set was more valuable and the results were showed in Tables 3 and 4. It can be noticed that the GBDT model perform the best with the highest values of sensitivity, specificity, accuracy, Kappa and AUC, which was 86.9%, 89.4%, 88.1%, 0.772 and 0.940, respectively. By comparing with the training data, the performance of the 4 models has declined, especially the RF model. The RF model performed the worst with the value of sensitivity, specificity, accuracy, Kappa and AUC, which was 74.7%, 67.3%, 70.4% 0.433 and 0.791, respectively. The performance of CART and Ada-

Table 2 ROC analysis of the three models using training data

Models	AUC	Standard Error	95% confidence interval
CART	0.944	0.005	0.934–0.955
GBDT	0.986	0.002	0.981–0.991
Ada-DT	0.961	0.004	0.953–0.970
RF	0.986	0.001	0.982–0.992

DT was relatively close and satisfactory. The standard error values of AUC were increased compared to the training set. The RF model indicated the highest standard error values of AUC, which was 0.021, followed by the CART model (0.017), Ada-DT model (0.014) and GBDT model (0.012).

4.2 Generation of landslide susceptibility map

The above analyses verified that the GBDT model showed a superior capability and robustness in predicting the landslide susceptibility over the other 3 models. Therefore, the GBDT was considered to be the most suitable model and applied to the whole study area for landslide susceptibility mapping. The landslide susceptibility map was reclassified into five classes: very low (0 ~ 0.2), low (0.2 ~ 0.4), moderate (0.4 ~ 0.6), high (0.6 ~ 0.8), very high (0.8 ~ 1) by using the equal spacing method (Fig. 7). The map should satisfy two spatial effective rules: (1) the existing disaster points should belong to the high-susceptibility class and (2) the high-susceptibility class should cover only small areas (Bui et al. 2012). The number of units belonging to high and very high class reached 46 and 340, accounting for 3.2% and 23.2%, respectively (Figs. 7 and 8). Landslide locations were mostly predicted in the dark (red or deep red) areas. The number of units belonging to very low and low class reached 770 and 260, accounting for 53.2% and 17.6%, respectively (Figs. 7 and 8). The units belonging to moderate class accounted for the smallest proportion, at 2.8% (Fig. 7). Most of the non-landslide points screened by K-means clustering were predicted to fall in low or very low susceptibility areas.

The very-high susceptibility areas of landslide are mainly distributed around the 306- provincial highway, which runs through the three townships in the study area, including Lang Town, Zhongda Town and Dongga Town. These areas are closed to faults and streams. Streams scour eroded slopes, and loose rock and soil tend to accumulate, which can easily cause landslides. More importantly, the areas are densely populated with highway projects and human activities which have been caused frequent disasters.

The very-low susceptibility region was almost entirely distributed on the south side of the 306-provincial highway with high elevation, sparse population and lush vegetation.

The low or moderate susceptibility region was mainly distributed on the north side of the 306-provincial highway. This area is far from streams and faults, but the terrain is undulating and the rainfall is sufficient.

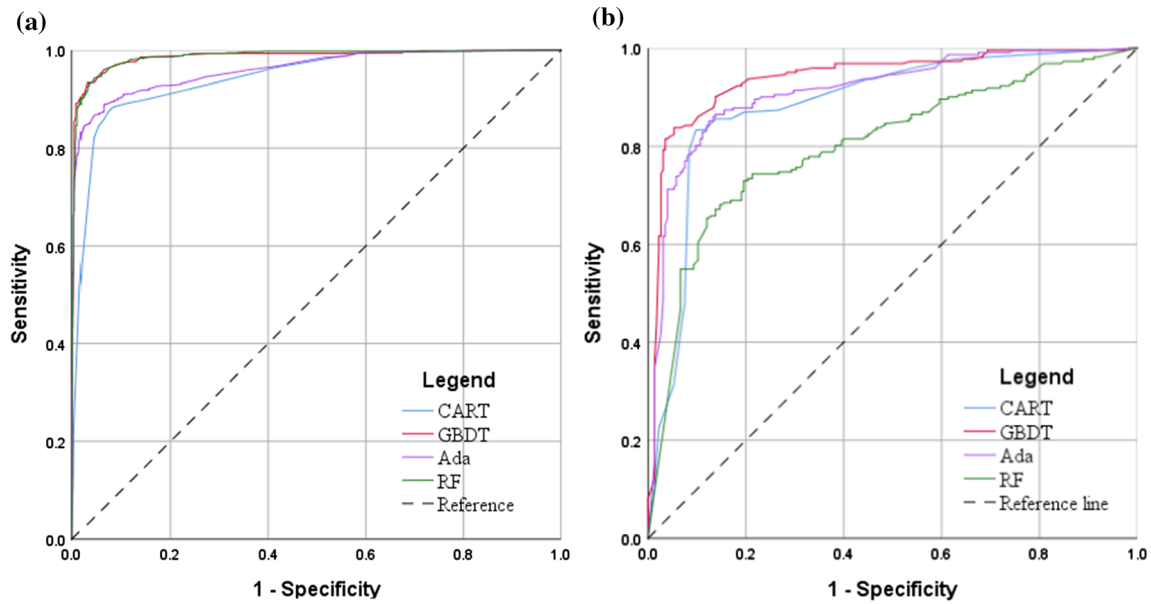


Fig. 6 Analysis of ROC curve for the landslide susceptibility map: **a** Success rate curve of landslide using the training dataset; **b** Prediction rate curve of landslide using the validation dataset

Table 3 Models’ performance using verification dataset

Indexes	CART	GBDT	Ada-DT	RF
TP (%)	87.1	89.8	88.0	61.6
TN (%)	85.6	86.5	84.7	79.1
FP (%)	12.9	10.2	12.0	38.4
FN (%)	14.4	13.5	15.3	20.9
Sensitivity (%)	85.8	86.9	85.2	74.7
Specificity (%)	86.9	89.4	87.6	67.3
Accuracy (%)	86.3	88.1	86.3	70.4
Kappa	0.731	0.772	0.731	0.433

Table 4 ROC analysis of the models using validating data

Models	AUC	Standard Error	95% confidence interval
CART	0.886	0.017	0.853–0.919
GBDT	0.940	0.012	0.916–0.963
Ada-DT	0.913	0.014	0.885–0.941
RF	0.791	0.021	0.761–0.844

4.3 Relative importance of conditioning factors for landslide

Determining the factors that have the obvious impact on the occurrence of landslide is an essential task for conditioning, stabilizing and reducing landslide risks. Based on

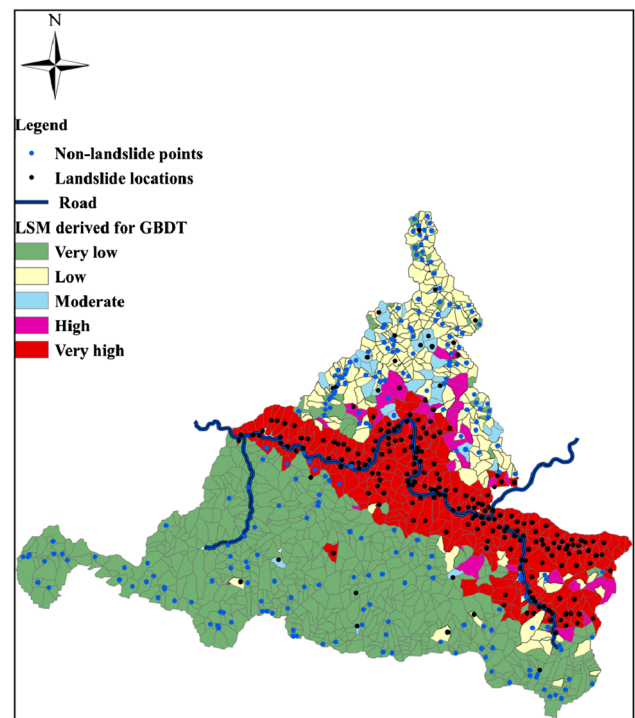


Fig. 7 Landslide susceptibility map using the GDBT model

the Gini index (the larger the value of the obtained result, the greater the contribution to the occurrence of landslide) (Breiman 2001), seven parameters, including elevation, MED, NDVI, rainfall, DTS, DTR and DTF had the obvious impact on landslide susceptibility. On the other hand, TWI, Slope angle and Slope aspect have insignificant impact on the occurrence of landslide (Fig. 9). The conditioning

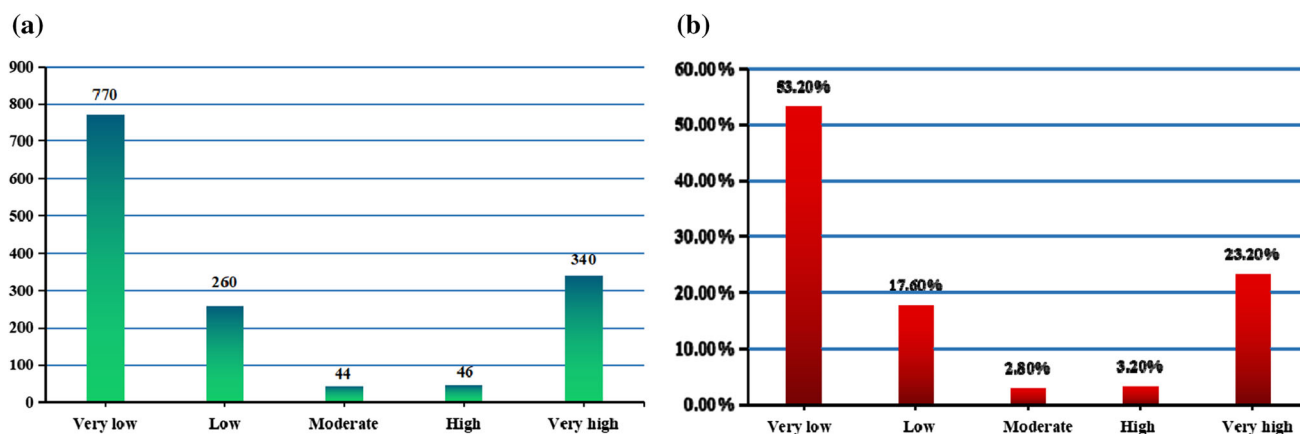


Fig. 8 Numbers and percentage of units in different susceptibility classes for landslide: **a** Numbers of units in different susceptibility classes for landslide; **b** Percentages of different susceptibility classes for landslide

Table 5 Conditioning factors assigned by the GBDT

Method	Elevation	MED	NDVI	Rainfall	DTS	DTR	DTF	Slope Angle	Slope aspect	TWI
GBDT	0.30	0.22	0.11	0.09	0.09	0.07	0.07	0.03	0.01	0.01

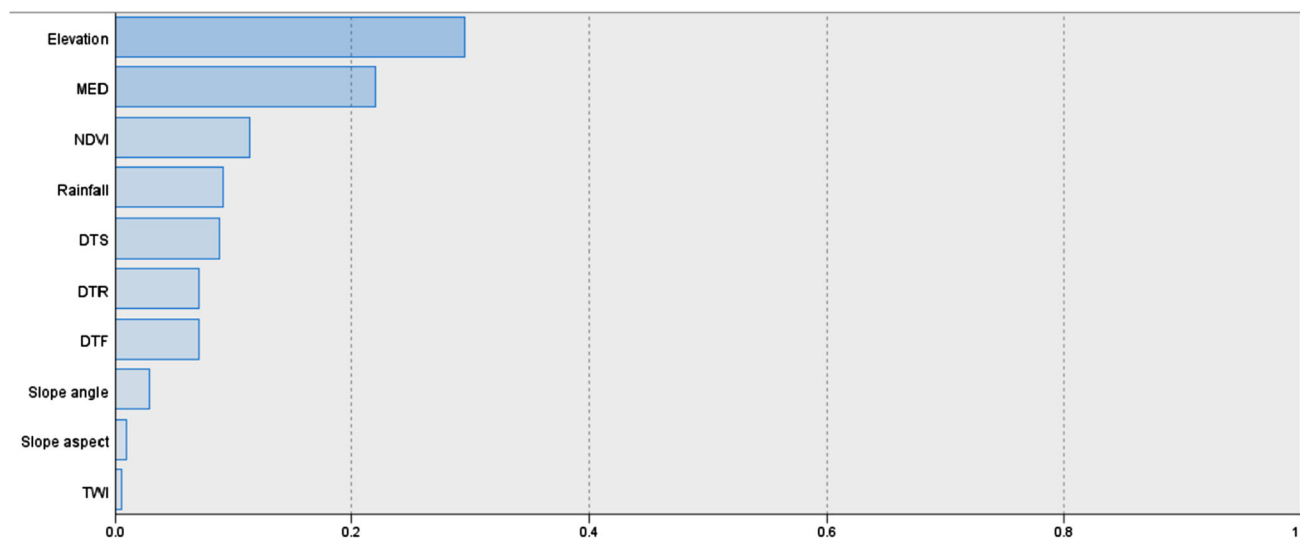


Fig. 9 Parametric importance graphics obtained from GBDT model

factors with significant effects were selected and normalized as shown in Table 5. The weight values of elevation, MED and NDVI were greater than 0.1, which were 0.30, 0.22, 0.11, respectively. The weight values of rainfall and DTS, DTR and DTF were close to 0.1, which were 0.09, 0.09, 0.07 and 0.07, respectively. The weight values of Slope angle, slope aspect and TWI, and Slope aspect were close to 0, which were 0.02, 0.02 and 0.01, respectively.

5 Discussion

5.1 Method used for modeling

A literature review declares that each model has its own strengths and weaknesses based on different assumptions and requirements (Reichenbach et al. 2018; Merghadi et al. 2020), and generally its performance varies with the characteristics of different study areas. The freedom of choice to decide which modeling method is most suitable for a particular application is challenging (Goetz et al.

2015). Although, many types of methods have been applied to and compared in landslide susceptibility mapping to obtain the best one for a given region (Goetz et al., 2015; Ciurleo et al., 2017; Binh et al. 2020; Liang et al. 2020b). However, most base learners are less robust due to their own limitations. Therefore, it is essential to investigate new methods or techniques for landslide susceptibility assessment. Our study implements a detailed comparison to evaluate the performances of some ensemble machine learning models based on DT (CART, GBDT, Ada-DT and RF) in predicting landslide susceptibility mapping. Bagging and boosting are two algorithms commonly used in ensemble learning. GBDT and Ada-dt are two representative algorithms in Boosting, while RF is based on bagging. CART is a tree-building algorithm that can be used for both classification and regression task.

The results of this study proved that the performance of landslide model can be enhanced with the use of machine learning ensembles. For training data sets, GDBT, Ada-DT and RF had better prediction performance than CART, which is in agreement with the findings of Bui et al. (2016a) which concluded that the prediction performance of landslide model is enhanced with the used of machine learning ensemble framework. This enhancement is related to the ability for reducing both bias and variance and avoid over-fitting problems (Dou et al. 2020). The GDBT model performed better than Ada-DT although both of them were based on boosting algorithm. For the verification group, the performance of GDBT and Ada-DT was still excellent, while RF has declined, even lower than CART. Therefore, the performance of GDBT was the best and stable among all models in our study. Based on different integration algorithms, the performance of the model varies and may fluctuate, but it is generally higher than that of a single primitive learner. In this paper, the GBDT model showed a superior capability and robustness in landslide susceptibility prediction over the other 3 models.

5.2 Determination of non-landslide points

Landslide susceptibility modeling is considered as a binary classification process, and construction of the binary classification needs both positive and negative database (Bennett et al. 2016). Positive database consists of landslide locations with the values of “1”, while negative (non-landslide) database with the values of “0”. Most previous studies emphasized a complete and accurate disaster inventory map, which consisted of a certain number of disaster points (Reichenbach et al. 2018). However, the determination of non-landslide points is rarely discussed. In most cases, negative samples were selected randomly or subjectively (Ciurleo et al. 2016; Cao et al. 2019), which is controversial. It is believed that non-landslide points

should only be selected from low-prone areas, which is difficult to achieve by selecting randomly or subjectively. Clustering is a effective method to solve the problem. The aim of the modelling of landslide susceptibility is to predicted the area prone to disaster or not prone to disaster. For a certain study area, the landslide area is very small compared to the non-landslide area. However, an equal number of non-landslide points (1:1) was need to the landslide database to avoid an over-estimation of non-landslide areas (Ayalew and Yamagishi 2005; Du et al. 2020), which means that the determination of non-landslide points is equally important. In this study, the existing landslide points were used as the initial clustering center, and the points with the farthest distance through iterative calculation were found as non-slide points. Therefore, 229 non-landslide points were determined through K-means clustering. The result showed that most of the non-landslide points fell in the areas with low or very low susceptibility which can be stated as the non-disaster points obtained by K-means clustering were reliable and representative.

6 Conclusion

In the present study, the CART and three ensemble learning machines, namely GBDT, Ada-DT and RF were explored and compared in the performance of landslide susceptibility prediction in Longzi County, Southeastern Tibet, China. An improved sampling method was applied to select non-landslide samples with the use of K-means clustering. The performance of the models was evaluated using ROC curves, three statistical measures and Kappa coefficient and The following conclusions can be drawn:

1. GBDT model showed a superior capability and robustness with the highest accuracy value (88.1%) and AUC value (0.940) compared with the other three models and was selected as the optimized model for the whole study area.
2. Although the performance of the model varies and may fluctuate based on different integration algorithms, the accuracy desire improvement with the integration of Boosting and Bagging algorithm.
3. According to comprehensive analysis of landslide conditioning factors based on GBDT model, Elevation, maximum elevation difference, NDVI, rainfall, DTS, DTR and DTF have relatively obvious importance compared to the rest factors. The zoning map obtained by GBDT model was reasonable in predicting the distribution of historical disaster points and the range of highly susceptibility regions.

4. The non-landslide samples selected by k-means clustering are more reasonable and representative, which avoids errors caused by random or subjective selection.
5. The combination of unsupervised learning and supervised learning further improves the reliability and accuracy of the landslide susceptibility modeling.

This study discussed the application of new learning machines that use decision trees as primitive learners and Boosting and Bagging as integrated ideas in landslide susceptibility mapping. Overall, the optimized model is effective for the improved management of land use and reducing the burden of landslides. In the future, more forms of comparison or combination for LSM still need to be explored. However, there are some limitations in our study:

1. More case studies, such as Stacking as an integration idea, support vector machine as a primitive learner, etc., should be conducted to explore and evaluate the the potential overall performance;
2. Other clustering methods, such as fuzzy C-means clustering and system clustering, can be used for comparison and verification;
3. Several hyperparameters applied in machine learning should be turned repeatedly.

Acknowledgements This work was supported by the National Natural Science Foundation of China (Grant Nos. 41972267 and 41572257).

References

- Anil K (2010) Data clustering: 50 years beyond K-Means. *Pattern Recogn Lett* 31:651–666
- Ayalew L, Yamagishi H (2005) The application of GIS-based logistic regression for landslide susceptibility mapping in the Kakuda-Yahiko Mountains, Central Japan. *Geomorphology* 65:12–31
- 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
- Bregoli F, Medina V, Chevalier G, Hürlimann M, Bateman A (2015) Debris-flow susceptibility assessment at regional scale: validation on an alpine environment. *Landslides* 12(3):437–454
- Breiman L (1994) Bagging predictors. *Machine Learn* 24:23–140
- Breiman L (2001) Random forests. *Mach Learn* 45(1):5–32
- Breiman L, Friedman JH, Olshen RA, Stone CJ (1984) Classification and regression trees. Chapman & Hall, New York
- Bucci F, Santangelo M, Cardinali M et al (2016) Landslide distribution and size in response to Quaternary fault activity: the Peloritani Range, NE Sicily. *Italy. Earth Surf Process Land* 41(5):711–720
- Bui DT, Pradhan B, Lofman O, Revhaug I, Dick OB (2012) Landslide susceptibility assessment in the Hoa Binh Province of Vietnam: a comparison of the Levenberg-Marquardt and Bayesian regularized neural networks. *Geomorphology*. <https://doi.org/10.1016/j.geomorph.2012.04.023>
- Calle ML, Urrea V (2010) Letter to the Editor: stability of random forest importance measures. *Brief Bioinform* 12(1):86–89. <https://doi.org/10.1093/bib/bbq011>
- Cao J, Zhang Z, Wang C, Liu J, Zhang L (2019) Susceptibility assessment of landslides triggered by earthquakes in the Western Sichuan Plateau. *CATENA* 175:63–76
- Chen W, Xie X, Wang J, Pradhan B, Hong H, Bui DT, Ma J (2017) A comparative study of logistic model tree, random forest, and classification and regression tree models for spatial prediction of landslide susceptibility. *CATENA* 151:147–160
- Chung CF, Fabbri AG (2003) Validation of spatial prediction models for landslide hazard mapping. *Nat Hazards* 30:451–472
- Ciurleo M, Calvello M, Cascini L (2016) Susceptibility zoning of shallow landslides in fine grained soils by statistical methods. *CATENA* 139:250–264
- Ciurleo M, Cascini L, Calvello M (2017) A comparison of statistical and deterministic methods for shallow landslide susceptibility zoning in clayey soils. *Eng Geol* 223:71–81
- Colkesen I, Sahin EK, Kavzoglu T (2016) Susceptibility mapping of shallow landslides using kernel-based Gaussian process, support vector machines and logistic regression. *J Afr Earth Sci* 118:53–64
- Cruden DM, Varnes DJ (1996) Landslide types and processes. In: Turner AK, Schuster RL (eds) Landslides, investigation and mitigation, Special Report 247. Transportation Research Board, Washington D.C., pp 36–75
- Dietterich TG (2000) An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting, and randomization. *Mach Learn* 40(2):139–157
- Dou J, Yunus AP, Bui DT, Merghadi A, Sahana M, Zhu Z, Pham BT (2020) Improved landslide assessment using support vector machine with bagging, boosting, and stacking ensemble machine learning framework in a mountainous watershed Japan. *Landslides* 17(3):641–658
- Du J, Glade T, Woldai T, Chai B, Zeng B (2020) Landslide susceptibility assessment based on an incomplete landslide inventory in the Jilong Valley, Tibet. *Chin Himal Eng Geol*. <https://doi.org/10.1016/j.enggeo.2020.105572>
- Evans IS (1979) An integrated system of terrain analysis and slope mapping. FinalReport on Grant DA-ERO-591-73-G0040. University of Durham, England
- Fan W, Stolfo SJ, Zhang J (1999). The application of AdaBoost for distributed, scalable and on-line learning. In: Proceedings of the fifth SIGKDD international conference on knowledge discovery and data mining (pp.362–366).
- Feizizadeh B, Blaschke T, Nazmfar H (2014) GIS-based ordered weighted averaging and Dempster-Shafer methods for landslide susceptibility mapping in the Urmia Lake Basin Iran. *Int J Digital Earth* 7(8):688–708
- Fernández-Delgado M, Cernadas E, Barro S et al (2014) Do we need hundreds of classifiers to solve real world classification problems? *J Mach Learn Res* 15(1):3133–3181
- 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
- Friedman JH (2001) Greedy function approximation: a gradient boosting machine. *Ann Stat* 29(5):1189–1232
- Furlani S, Ninfo A (2015) Is the present the key to the future? *Earth-Sci Rev* 142(C):38–46
- 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>
- Green DM, Swets JM (1966) Signal detection theory and psychophysics. Wiley, New York
- Guzzetti F, Reichenbach P, Ardizzone F, Cardinali M, Galli M (2006a) Estimating the quality of landslide susceptibility models. *Geomorphology* 81:166–184. <https://doi.org/10.1016/j.geomorph.2006.04.007>

- Guzzetti F, Galli M, Reichenbach P, Ardizzone F, Cardinali M (2006b) Landslide hazard assessment in the Collazzone area, Umbria, central Italy. *Nat Hazard Earth Syst Sci* 6:115–131. <https://doi.org/10.5194/nhess-6-115-2006>
- Hartigan J, Wong M (1979) Algorithm AS 136: A K-means clustering algorithm. *J R Stat Soc C* 28:100–108
- Heckmann T, Gregg K, Gregg A, Becht M (2014) Sample size matters: investigating the effect of sample size on a logistic regression susceptibility model for debris flows. *Nat Hazards Earth Syst Sci* 14:259–278
- Hong H, Pradhan B, Xu C, Bui DT (2015) Spatial prediction of landslide hazard at the Yihuang area (China) using two-class kernel logistic regression, alternating decision tree and support vector machines. *CATENA* 133:266–281
- Hong H, Liu J, Bui DT, Pradhan B, Acharya TD, Pham BT, Zhu AX, Chen W, Ahmad BB (2018) Landslide susceptibility mapping using J48 decision tree with AdaBoost, bagging and rotation forest ensembles in the Guangchang area (China). *CATENA* 163:399
- Hussin HY, Zumpano V, Reichenbach P, Sterlacchini S, Micu M, van Westen C, Bălteanu D (2015) Different landslide sampling strategies in a grid-based bi-variate statistical susceptibility model. *Geomorphology* 253:508–523. <https://doi.org/10.1016/j.geomorph.2015.10.030>
- James G, Witten D, Hastie T, Tibshirani R (2013) An introduction to statistical learning. Springer, New York, p 441
- Kornejady A, Ownegh M, Bahremand A (2017) Landslide susceptibility assessment using maximum entropy model with two different data sampling methods. *CATENA* 152:144–162
- Lian C, Zeng Z, Yao W, Tang H (2014) Extreme learning machine for the displacement prediction of landslide under rainfall and reservoir level. *Stoch Environ Res Risk Assess* 28(8):1957–1972
- Liang Z, Wang C, Han S, Khan KUJ, Liu Y (2020a) Classification and susceptibility assessment of debris flow based on a semi-quantitative method combination of the fuzzy C-means algorithm, factor analysis and efficacy coefficient. *Nat Hazards Earth Syst Sci* 20:1287–1304. <https://doi.org/10.5194/nhess-20-1287-2020>
- Liang Z, Wang C, Zhang Z-M, Khan K-U-J (2020b) A comparison of statistical and machine learning methods for debris flow susceptibility mapping. *Stoch Environ Res Risk Assess*. <https://doi.org/10.1007/s00477-020-01851-8>
- Likas A, Vlassis N, Verbeek JJ (2003) The global K-means clustering algorithm. *Pattern Recogn* 36:451–461
- MacQueen J (1967) Some methods for classification and analysis of multivariate observations. *Proc 5th Berkeley Symp Math Stat Probab* 1(14):281–297
- Magliulo P, DiLisio A, Russo F, Zelano A (2008) Geomorphology and landslide susceptibility assessment using GIS and bivariate statistics: a case study in southern Italy. *Nat Hazards* 47:411–435
- Merghadi A, Yunus AP, Dou J, Whiteley J, Thai Pham Binh Bui DT, Ram A, Abderrahmane B (2020) Machine learning methods for landslide susceptibility studies: a comparative overview of algorithm performance. *Earth Sci Rev*. <https://doi.org/10.1016/j.earscirev.2020.103225>
- Mingoti SA, Lima JO (2006) Comparing SOM neural network with Fuzzy c-means, K-means and traditional hierarchical clustering algorithms. *Eur J Oper Res* 174(3):1742–1759
- Nefeslioglu HA, Gökçeoglu C, Sonmez H (2008) An assessment on the use of logistic regression and artificial neural networks with different sampling strategies for the preparation of landslide susceptibility maps. *Eng Geol* 97(3):171–191. <https://doi.org/10.1016/j.enggeo.2008.01.004>
- Pedregosa F, Varoquaux G, Gramfort A et al (2011) Scikit-learn: machine learning in python. *J Mach Learn Res* 12(10):2825–2830
- Pham BT, Prakash I (2019) A novel hybrid model of bagging-based naïve bayes trees for landslide susceptibility assessment. *Bull Eng Geol Env* 78(3):1911–1925
- Pham BT, Van Phong T, Nguyen-Thoi T, Trinh PT, Tran QC, Ho LS, Singh SK, Duyen TT, Nguyen LT, Le HQ, Van Le H, Hanh TB, Quoc NK, Prakash I (2020) GIS-based ensemble soft computing models for landslide susceptibility mapping. *Adv Space Res*. <https://doi.org/10.1016/j.asr.2020.05.016>
- Pourghasemi HR, Rahmati O (2018) Prediction of the landslide susceptibility: which algorithm, which precision? *Catena* 162:177–192
- Pradhan B (2010) Landslide susceptibility mapping of a catchment area using frequency ratio, fuzzy logic and multivariate logistic regression approaches. *J Indian Soc Remote Sens* 38(2):301–320
- Reichenbach P, Rossi M, Malamud BD et al (2018) A review of statistically-based landslide susceptibility models. *Earth Sci Rev* 180(5):60–91. <https://doi.org/10.1016/j.earscirev.2018.03.001>
- Samadi M, Jabbari E, Azamathulla HM (2014) Assessment of M5 model tree and classification and regression trees for prediction of scour depth below free overfall spillways. *Neural Comput Appl* 24:357–366
- Schapire RE (1990) The strength of weak learnability. *Mach Learn* 5(2):197–227
- Segal MR (2004) Machine Learning Benchmarks and Random Forest Regression. Center for Bioinformatics and Molecular Biostatistics UC, San Francisco. <https://eprints.cdlib.org/uc/item/35x3v9t4>.
- Swets JA (1988) Measuring the accuracy of diagnostic systems. *Science* 240:1285–1293
- Tien Bui D, Ho TC, Revhaug I, Pradhan B, Nguyen DB (2014) Landslide susceptibility mapping along the national road 32 of Vietnam using GIS-based J48 decision tree classifier and its ensembles Cartography from Pole to Pole. Springer, Berlin, pp 303–317
- Tien Bui D, Ho T-C, Pradhan B et al (2016a) GIS-based modeling of rainfall-induced landslides using data mining-based functional trees classifier with AdaBoost, Bagging, and MultiBoost ensemble frameworks. *Environ Earth Sci* 75:1101. <https://doi.org/10.1007/s12665-016-5919-4>
- Tien Bui D et al (2016b) GIS-based modeling of rainfall-induced landslides using data mining based functional trees classifier with AdaBoost, bagging, and MultiBoost ensemble frameworks. *Environ Earth Sci* 75:1101–1123
- Tong L, Qi W, An G, Liu C (2019) Remote sensing survey of major geological disasters in the Himalayas. *J Eng Geol* 27(03):496
- Trigila A, Catani F, Casagli N, Crosta G, Esposito C, Frattini P, Iadanza C, Lagomarsino D, Lari S, Scarascia-Mugnozza G, Segoni S, Spizzichino D, Tofani V (2012) The landslide susceptibility map of Italy at 1:1 million scale. *Geophysical Research Abstracts, European Geosciences Union — General Assembly 2012, Vienna 22–27 April 2012*
- van Westen CJ, Castellanos E, Kuriakose SL (2008) Spatial data for landslide susceptibility, hazard, and vulnerability assessment: an overview. *Eng Geol* 102(3–4):112–131
- Varnes DJ (1978) Slope movement types and processes. In: Schuster RL, Krizek RJ (eds), *Landslides: analysis and control*, National Research Council, Washington, D.C., Transportation Research Board, National Academy Press, Special Report 176, pp 11–33
- Varnes, D.J., 1984. Landslide hazard zonation: a review of principles and practice. Commission on Landslides of the IAEG, UNESCO Natural Hazards No. 3 (61 pp.).
- Woods M, Guivant J, Katupitiya J (2013) Terrain classification using depth texture features. In: *Proceeding Australian Conference of Robotics and Automation*, Sydney, NSW, Australia, 2013, pp 1–8

- Wu X, Ren F, Niu R (2014) Landslide susceptibility assessment using object mapping units, decision tree, and support vector machine models in the Three Gorges of China. *Environ Earth Sci* 71(11):4725–4738
- Youssef AM, Pradhan B, Jebur MN et al (2015a) Landslide susceptibility mapping using ensemble bivariate and multivariate statistical models in Fayfa area Saudi Arabia. *Environ Earth Sci* 73(7):3745–3761. <https://doi.org/10.1007/s12665-014-3661-3>
- Youssef AM, Pradhan B, Jebur MN, El-Harbi HM (2015b) Landslide susceptibility mapping using ensemble bivariate and multivariate statistical models in Fayfa area Saudi Arabia. *Environ Earth Sci* 73:3745–3761
- Youssef AM, Pourghasemi HR, Pourtaghi ZS, Al-Katheeri MM (2016) Landslide susceptibility mapping using random forest, boosted regression tree, classification and regression tree, and general linear models and comparison of their performance at Wadi Tayyah Basin, Asir Region Saudi Arabia. *Landslides* 13(5):839–856
- Zeze JL, Pereira S, Melo R et al (2017) Mapping landslide susceptibility using data-driven methods. *Sci Total Environ* 589:250–267

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.