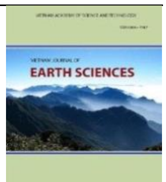
Vietnam Journal of Earth Sciences, 42(3), 208–227, Doi: 10.15625/0866-7187/42/3/14952

(VAST)

Vietnam Academy of Science and Technology

Vietnam Journal of Earth Sciences

http://www.vjs.ac.vn/index.php/jse



Shallow landslide susceptibility mapping: A comparison between classification and regression tree and reduced error pruning tree algorithms

Bahareh Ghasemian[1], Dawod Talebpoor Asl[2], Binh Thai Pham[3*], Mohammadtghi Avand[4], Huu Duy Nguyen[5], Saeid Janizadeh[4]

[1]*Department of Geomorphology, Faculty of Humanities, University of Mohaghegh Ardabili, Ardabil, Iran*
[2]*Department of Geomorphology, Faculty of Natural Resources, University of Kurdistan, Sanandaj, Iran*
[3]*Department of Geotechnical Engineering, University of Transport Technology, Hanoi, Vietnam*
[4]*Department of Watershed Management Engineering, College of Natural Resources, Tarbiat Modares University, Tehran, P.O. Box 14115-111, Iran*
[5]*Faculty of Geography, VNU University of Science, Hanoi, Vietnam*

Received 23 December 2019; Received in revised form 02 March 2020; Accepted 25 June 2020

ABSTRACT

Shallow landslides through land degrading not only lead to threat the properly and life of human but they also may produce huge ecosystem damages. The aim of this study was to compare the performance of two decision tree machine learning algorithms including classification and regression tree (CART) and reduced error pruning tree (REPTree) for shallow landslide susceptibility mapping in Bijar, Kurdistan province, Iran. We first used 20 conditioning factors and then they were tested by information gain ratio (IGR) technique to select the most important ones. We then constructed a geodatabase based on the selected factors along with a total of 111 landslide locations with a ratio of 80/20 (for calibration/validation). The performance of the models was checked by the true positive rate (TP Rate), false positive rate (FP Rate), precision, recall, F1-Measure, Kappa, mean absolute error, and area under the receiver operatic curve (AUC). Results of IGR specified that the slope angle and TWI had the most contribution to shallow landslide occurrence in the study area. Moreover, results concluded that although these models had a high goodness-of-fit and prediction accuracy, the CART model (AUC=0.856) outperformed the REPTree model (AUC=0.837). Therefore, the CART model can be used as a promising tool and also as a base classifier to hybrid with optimization algorithms and Meta classifiers for spatial prediction of shallow landslide-prone areas.

Keywords: Shallow landslide; machine learning; information gain ratio; classifier; GIS; Iran.

1. Introduction

Landslide is considered as one of the most dangerous disasters in the world because of its effects and damages on material and human. Steep topographies, heavy rainfall, weak lithology, poor human development on the land are some of the main factors that are responsible for a landslide occurrence (Chen et al., 2017a; Jaafari et al., 2018). Climate

[*]Corresponding author, Email: binhpt@utt.edu.vn

change in conjunction with rapid changes in land use increases the number and intensity of landslides around the world (Barik et al., 2017; Tien Bui et al., 2018a), causing damages and challenges to decision-makers (Hong et al., 2016; Jaafari et al., 2015). The landslide is a complicated phenomenon because of the time of rapid occurrence on large space and that is more difficult to fight against (Guzzetti et al., 2006). This disaster occurs in numerous regions of the world, caused great damage every year on material and population life, for example, 1.5 billion dollars in the United States, 2 billion in Japan (Bloechl and Braun, 2005). Iran is considered one of the countries most affected by this phenomenon because of climatic, topographical, human factors. This phenomenon causes about 500 billion rials of damages in Iran, these figures do not yet count the destruction of non-renewable goods (Pourghasemi et al., 2012). Landslide analysis is very necessary in order to reduce the damage. Because the landslide susceptibility analysis makes it possible to determine the spatial probability of landslide occurrences that can estimate and predict by analyzing the relationships between landslide events in the historical and geomorphologic, hydrologic, climatic and human activity factors (Ercanoglu and Gokceoglu, 2004). At the center of all measurements, landslide susceptibility maps must be available because of preliminary information on these maps for landslide response and management planning. Determining landslide susceptibility areas can contribute to reducing damage to agriculture, and livelihoods by avoiding new construction and developments in exposed areas. Technically, two principle method is used to build the landslide susceptibility map: the qualitative technique and the quantitative technique. The main qualitative method based on the knowledge and experiences of the experts (Pham et al., 2018b). The analytic hierarchy process (AHP) is the most known method of the qualitative method (Chen, 2016; Shirzadi et al., 2017a). With the development of computational power, the quantitative method is more used by researchers to analyze natural catastrophes like flash flood and landslide, for example, the statistical method includes frequency ratio (FR) (Guzzetti et al., 2006; Shirzadi et al., 2017a), information value (IV) (Chen et al., 2014), statistical index (SI) (Regmi et al., 2013), certainly factor (Chen, 2016; Hong et al., 2016), weights-of-evidence (WOE) (Chen et al., 2018a; Ozdemir and Altural, 2013), logistic regression (LR) (Chen et al., 2018a; Hemasinghe et al., 2018; Shirzadi et al., 2012; Tien Bui et al., 2019a), Fuzzy logic (Cloke and Pappenberger; Hemasinghe et al., 2018). Nowadays machine learning algorithms have been attended by researchers, include artificial neural network (ANN) (Hemasinghe et al., 2018; Le et al., 2020; Nguyen et al., 2019a; Shirzadi et al., 2017b; Tien Bui et al., 2019c), decision tree (DT) (Shirzadi et al., 2017b), Adaptive Neuro-Fuzz Inference System (ANFIS) (Shirzadi et al., 2017b), support vector machine (SVM) (Shirzadi et al., 2017b; Tien Bui et al., 2019e), k-nearest neighbour (KNN) (Shirzadi et al., 2017b), Bayesian logistic regression (BLR) (Taheri et al., 2019; Tien Bui et al., 2018b), kernel logistic regression (KLR) (Chen et al., 2020a; Chen et al., 2019a; Chen et al., 2018b), logistic model tree (LMT) (Abedini et al., 2019; Chen et al., 2018a; Chen et al., 2019b; Khosravi et al., 2019a), boosted regression tree (BRT) (Shafizadeh-Moghadam et al., 2018), naive bayes tree (NBT) (Chen et al., 2020b; Chen et al., 2017b; Khosravi et al., 2019b), J48 Decision Trees (J48) (Chen et al., 2020c) and alternate decision tree (ADTree) (Shirzadi et al., 2018; Tien Bui et al., 2018c; Tien Bui et al., 2019c).

There are many studies that used the CART model in other environmental fields such as air quality modeling (Choubin et al., 2020), earth fissure modeling (Choubin et al., 2019b), flood susceptibility assessment (Choubin et al., 2019a), groundwater modeling (Choubin et al., 2019c), river suspended sediment modeling (Choubin et al., 2018a), precipitation forecasting (Choubin et al., 2018b). However few studies have been focused on the REPTree on landslide susceptibility mapping. Researchers have a tendency to integrate individual models to become a hybrid artificial intelligence using optimization algorithms. For example, ANFIS coupled with the genetic algorithm (ANFIS-GA) (Bui et al., 2018; Hong et al., 2018); ANFIS integrated with particle swarm optimization (ANFIS-PSO) (Nguyen et al., 2019c), ANFIS coupled with the differential evolution algorithm (ANFIS- DE) (Hong et al., 2018); ANFIS coupled with the biogeography-based optimization (ANFIS-BBO) and BAT algorithms (ANFIS-BAT) (Ahmadlou et al., 2018), ANFIS and invasive weed optimization (ANFIS-IWO), ANFIS with Whale Optimization Algorithm (WOA) and Grey Wolf Optimizer (GWO) (Chen et al., 2019a), ANFIS built using the ICA and firefly algorithm (FA) (Bui et al., 2018), ANFIS coupled with the imperialistic competitive algorithm (ANFIS-ICA) (Wang et al., 2019), naïve bayes tree based on the random subspace (RS-NBT) (Shirzadi et al., 2017b), KNN classifiers based on bagging (KNNs-BA) ensemble (Shahabi et al., 2020).

To increase model accuracy, researchers combined machine learning meta classifiers with statistical base classifiers (Shafizadeh-Moghadam et al., 2018), For example, support vector regression (SVR) and ICA (SVR-ICA) (Nguyen et al., 2019b), ADTree coupled with the multiboost, bagging (BA), rotation forest (RF) and random subspace Meta classifiers (Shirzadi et al., 2018), rotation forest based ANN, SVM DT and naïve bayes (NB), (Pham et al., 2019a) and radial basis function neural network coupled with rotation forest algorithm (RBFRF) (Pham et al., 2018b), support vector machine and its ensembles (Tien Bui et al., 2019e), SVM, ABTree, ADTree and REPTree and different ensembles (Nguyen et al., 2019b; Pham et al., 2019b; Thai Pham et al., 2019; Tien Bui et al., 2019b; Tien Bui et al., 2019d), random forest integrated with random subspace Meta classifier (RS-RF) (Miraki et al., 2019).

Although there are several methods for assessing the susceptibility of the landslide; however, there is still no conclusion on the best method for other regions. Each region has different geomorphological, hydrological, climatic and human activity conditions, therefore the methods with high accuracy are required to analyze the landslide. In this article, we compiled the CART and REPTree algorithms to select the best methods for the spatial prediction of shallow landslides to assist land use decision-makers. The CART algorithm is fast and do not need a long time for the training process and is known as a time-consuming algorithm when the training dataset is huge. Additionally, the resulting of CART classification model can be easily interpreted that is considered as one of the strong advantages of it over other classification techniques (Lee et al., 2006). It is inherently a non-parametric that does not assume the distribution of values of the conditioning factors. Moreover, it can process the numerical dataset that is highly skewed or multi-modal as well as categorical factors with either ordinal or non-ordinal structure (Breiman et al., 1984). REPTree as one of the fast learner decision tree algorithms can build a decision tree using an information gain ratio technique for splitting criterion and also can prune the tree by reduced-error pruning technique (Zhao and Zhang, 2008). It is a useful and applicable decision tree algorithm because it can better fit to deal with different scenarios (Zhao and Zhang, 2008). For example, Devasena (2014) was performed a comparison among different decision tree algorithms such as random forest, and they concluded that REPTree algorithm outperformed the other decision tree algorithms (Barbon Jr et al., 2016). These two mentioned algorithms belong to decision tree algorithms that few studies have been conducted to compare them for landslide modeling worldwide. On the other hand, the study area is a semi-arid region that has been comprised of loose and erodible formations especially on slopes that after a low intensify rainfall shallow landslide will occur.

2. Data acquisition and preparation

2.1. Description of study area

The study area is located in northeast of Kurdistan province in Iran. This zone lies between 35°48′25″N and 35°59′50″N in latitude and 47°28′50″E and 47°46′44″E in longitude, with altitudes ranging from 1573.69 to 2550 meters above sea level (Fig. 1). Its area is about 598 square kilometers. The surface of this area is mostly mountainous and its slope is between 0 and 67.7 degrees. However, the average slope of the region is 5.73˚. Analysis of precipitation data for the period 1987 to 2014 shows that the average annual rainfall is 338 mm. In terms of land use, about 78% of the land is covered by arable land. Other categories of land cover are classified as irrigation lands, woodlands, rangelands, residential areas and barren land (1.2%) that cover the least. According to Köppen classification climate is classified as D (cold air). The number of snow days and frozen days are 35 and 104, respectively. The maximum and minimum of average daily temperatures are 4.409 and 13.401°C, respectively (Shirzadi et al., 2017b). Geologically, 191.46 square kilometers (31.97%) of the area comprises Qt1 (high level nonconforming, texturally variable, recent elastic sediments), 248.75 square kilometers (41.83%) is covered by Plc (pelagic conglomerate with marl cement and sand is covered by sandstone displacement) and 138.77 square kilometers (23.17%) is covered by Ksl (shale and dark green sandstone, locally metamorphosed into low-grade facies). The study area is located in the Sanandaj-Sirjan tectonic zone of Iran where most of the mountains are composed of carbonaceous rocks (Miocene Formation), while the hilly areas are mainly covered by Pliocene structures including shale and marl and Quaternary sediments. Most of the faults are located in the south of the study area, with the northern half almost without fault. Most of these faults are northwest-southwest (Shirzadi et al., 2017b; Shirzadi et al., 2019; Shirzadi et al., 2018).

2.2. Landslide inventory map

The quality of a landslide inventory map has an effective contribution to the performance of the models (Shirzadi et al., 2019). In this study, we collected a total of 111 shallow landslide polygons from Forests, Rangeland and Watershed Management Organization of Iran (FRW). Then, these locations were confirmed by field surveys, and interpretations of aerial photographs (1:40,000 scale) and satellite images. We selected the center of the scarp of the shallow landslide as shallow landslide locations with a ratio of 80% (89 cases) for training and 20% (22 cases) for validation datasets. Field surveys revealed that most of the shallow landslides had a slip depth lower than 3 meters (Shirzadi et al., 2017b; Shirzadi et al., 2018). According to Varnes et al. (1989), shallow landslides in the study area were mainly classified as rotational sliding (70.6%), complex type (22.4%), and the rotational falling type (6.3%). Comprehensive information about shallow landslide frequency and its dimension can be found in (Shirzadi et al., 2017b; Shirzadi et al., 2019).

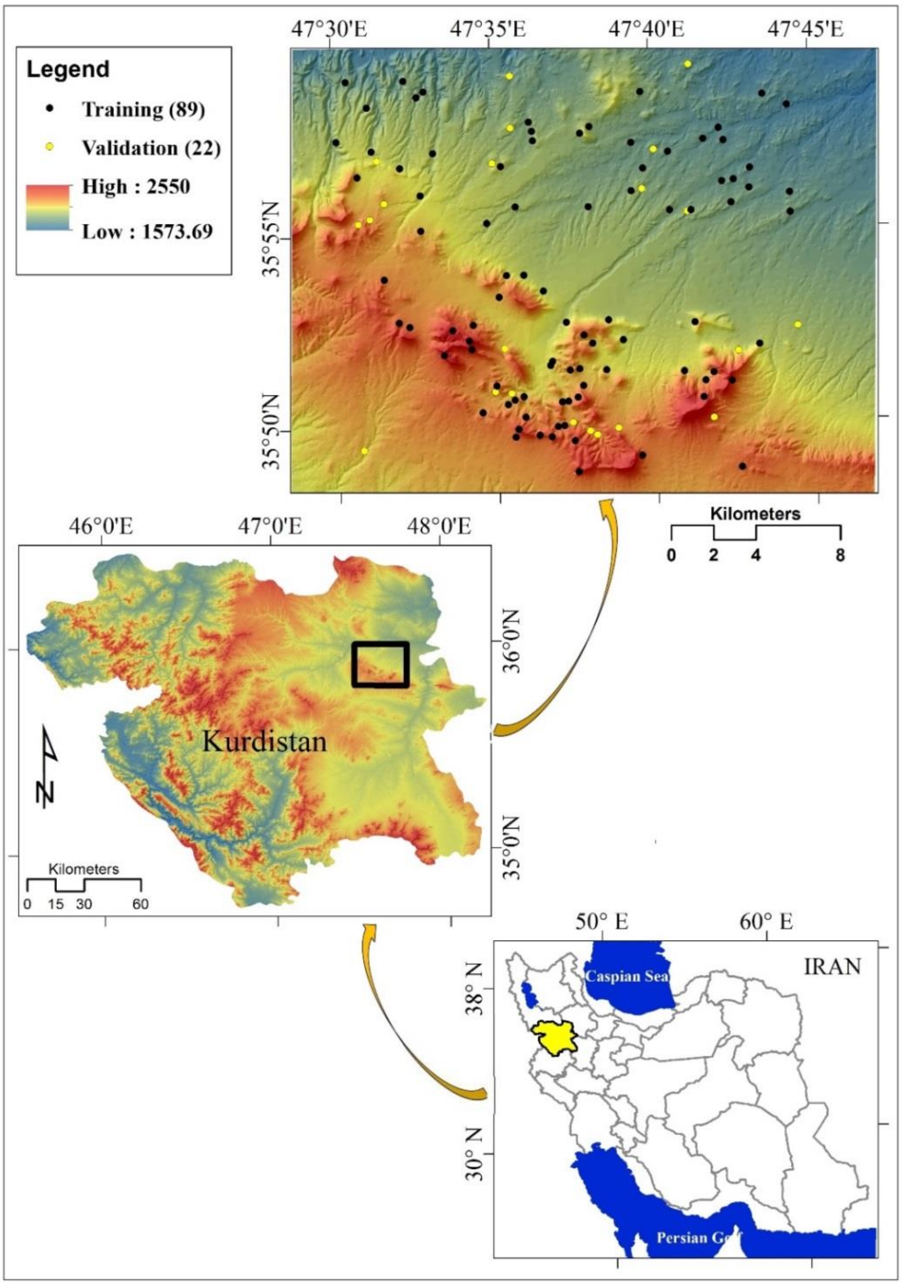


Figure 1. The location of shallow landslides of the study area in Iran and Kurdistan province

2.3. Landslide conditioning factors

In this study, we based on the literature review, data availability and our experiences of the study area selected 20 conditioning factors for developing landslide susceptibility prediction models (Fig. 2). According to this figure, the slope angle as a known important conditioning factor is widely used for developing landslide susceptibility prediction models (Chen et al., 2017c). The slope angle was extracted from the digital elevation model obtained from ASTER satellite image with a resolution of 30 m × 30 m and it was classified into 8 classes (Fig. 2b). The slope aspect was made and classified based on azimuth in 9 classes (Fig. 2c). The elevation factor has been used in all landslide susceptibility mapping studies as one of the affecting factors. In this study, the elevation map with 9 categories was prepared (Fig. 2a). The curvature map was also extracted from the DEM and classified into 6 classes ranges from -12.478 to 15.578 (Fig. 2m). The Profile curvature also controls the velocity changes of the mass of soil flowing down the slope. This factor varies in the study area from -6.738 to 10.386 which was then classified into 6 classes (Fig. 2o). Also, Plan curvature values that ranging from -10.722 to 7.455, were classified into 6 classes (Fig. 2n). Sediment transport index (STI) by surface flow determines the amount of transported sediments. The STI can be expressed as below: (Moore and Burch, 1986):

$$\mathrm{STI} = \left[\frac{\mathrm{A_s}}{22.13}\right]^{0.6}\left[\frac{\sin\beta}{0.0896}\right]^{1.3} \quad (1)$$

where A_s is specific catchment area (m^2) and β is slope gradient (radian). In this study, STI map was divided into six classes (Fig. 2p). Precipitation maps were prepared using the mean annual precipitation at 21 rainfall stations for the period 1996-2005. The map was categorized into 7 ranks (Fig. 2f). Also, the annual solar radiation (h) is calculated based on aspect and slope by ArcGIS 10.3 using the "Area solar radiation" tool. In this study, the solar radiation map was classified into 7 classes using 'natural break intervals' classification method (Fig. 2). Another factor influencing landslides is the stream power index (SPI), was calculated as (Moore and Wilson 1992):

$$SPI = A_s \times \tan\beta \quad (2)$$

Where A_s is the specific basin area (m^2) and β is represents the local degree of slope. The SPI show stream power to transport the sediment in the rivers that is affected by soil characteristic. Finally, the SPI map was divided into 6 categories (Fig. 2q). Another variable is Topographic wetness index (TWI), which is calculated as follows:

$$\mathrm{TWI} = \mathrm{Ln}\left(\frac{\mathrm{A_s}}{\tan\beta}\right) \quad (3)$$

Where A_s is the cumulative upslope area draining through a point (M2) and β is the slope angle at that point. In this study, the TWI map was classified into 6 categories (Fig. 2r). The TWI, SPI and STI conditioning factors were prepared by SAGA software. In this study, the river map was extracted from 1:50000 topographic map. The distance to the river was produced in 5 categories (Fig. 2) and the river density map was classified into 7 classes using the 'natural break interval' method (Fig. 2k). Also, the lithology map of the studied basin was extracted from the geological map of Sanandaj with a scale of 1: 100,000. Then, the lithological map was prepared and classified into 3 categories (Fig. 2c) and the distance to fault map was made using lithological data fault lines by ArcGIS 10.3 and was classified into 6 classes (Fig. 2g). Finally, the fault density map was classified into 7 classes (Fig. 2j). The land use map was prepared using satellite imagery of Landsat 7 (ETM+) dated April 25, 2008, by the 'supervised classification' method in PCI Geomatica V9.1 into 6 classes (Figure 2d). Normal Difference Vegetation Index (NDVI) can evaluate the vegetation status of slope surfaces. The NDVI is calculated from the reflectance measurements in the red (band 3) and near-infrared spectrum (band 4) as:

$$\mathrm{NDVI} = \frac{(\mathrm{NIR\ (Band4)} - \mathrm{Red(Band3)})}{(\mathrm{NIR(Band4)} + \mathrm{Red(Band3)})} \quad (4)$$

NDVI varies from 1 to -1 and the map was classified into 7 classes (Fig. 2t). Road conditions are another important factor that is used in landslide risk assessment models. Thus, in the study area, distance to road map in 5 classes (Fig. 2i) and the road density map in 7 categories were prepared and used for shallow landslide modeling (Fig. 2l).

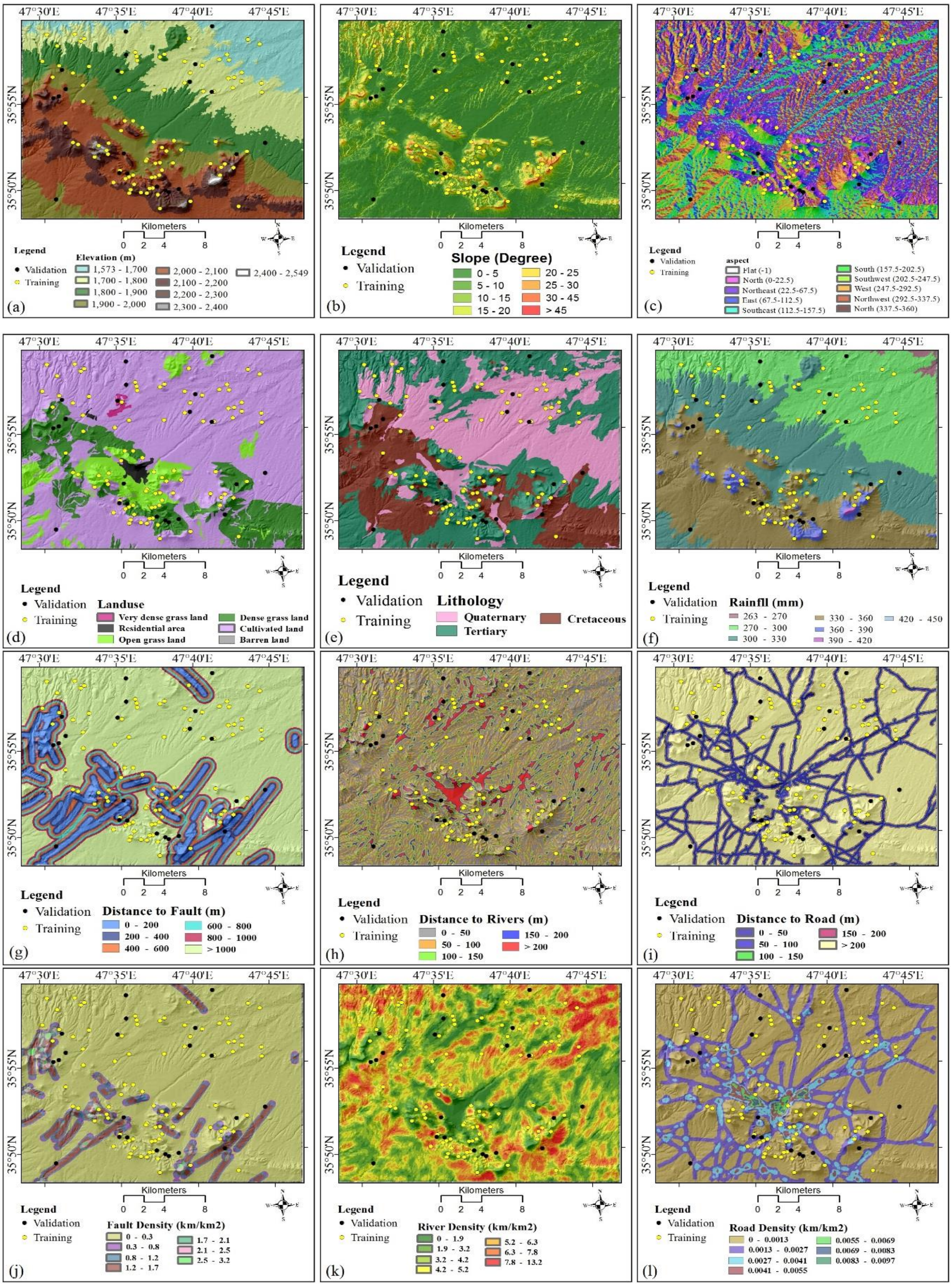


Figure 2. Landside conditioning factors map and their classes for landslide modeling in study area

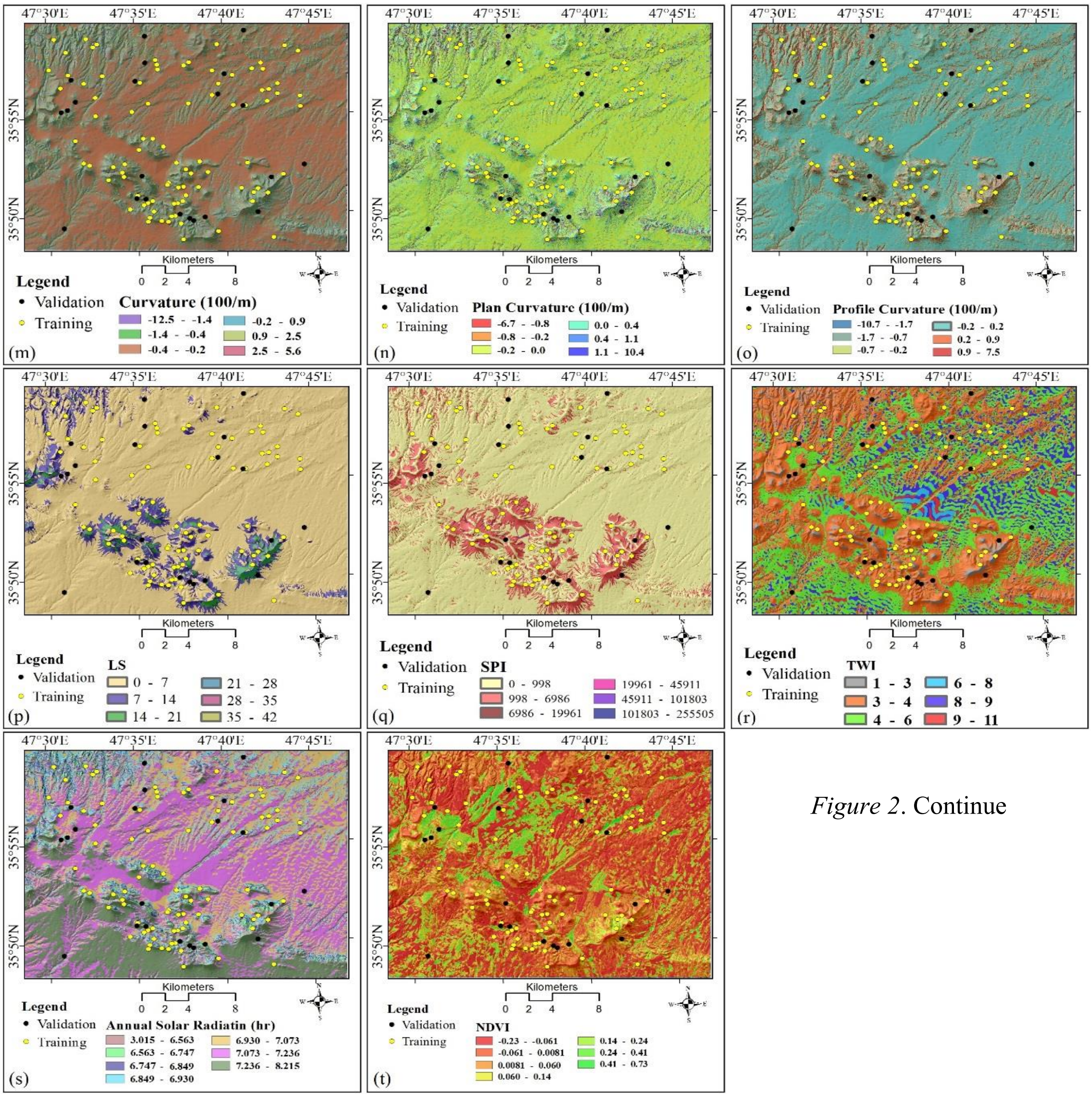


Figure 2. Continue

3. Methodology

3.1. Factor selection using information gain ratio technique

The role of conditioning factors on landslide occurrence is not the same because of their predictive powers. In the modeling process, the modeler should select appropriate factors with the highest capability. Although there are some methods to select these factors, we used of information gain ratio (IGR). This technique measure prefers to select attributes having a large number of values (Quinlan, 1986). The IGR technique is assigned the weights for each conditioning factor by entropy method and the importance of the factors are ranked based on the obtained weighs named average merit (AM). The IGR is considered the noise and over-fitting among the training distaste and it removes the factors that have no high predictive power (AM=0). Accordingly, the higher the value of IGR, the more importance the conditioning factor will be.

Given L as a training detest with m input sample, (Si, L) is the number of samples in training dataset L belonging to the class label (shallow landslide-non-shallow landslide). The average merit (AM) is obtained based on the entropy or expected information for each sample of the training dataset can be computed such as for slope angle using the following equation (Han and Kamber, 2001):

$$\mathrm{IGR}\left(\mathrm{L, Slope}\right) = \frac{\mathrm{E_n(L)\text{-}E_n(L, Slope)}}{\mathrm{SplitE_n(L, Slope)}} \quad (5)$$

$$\mathrm{E_n(L)} = -\sum_{i=1}^{2} \frac{\mathrm{n(S_i, Slope)}}{|\mathrm{L}|} \log_2 \frac{\mathrm{n(S_i, Slope)}}{|\mathrm{L}|} \quad (6)$$

$$\mathrm{E_n\ (L, Slope)} = \sum_{j=1}^{m} \frac{\mathrm{L_j}}{|\mathrm{L}|} \mathrm{E_n(L)} \quad (7)$$

$$\mathrm{SplitE_n}\ (L, Slope) = -\sum_{j=1}^{m} \frac{|L_j|}{|L} \log_2 \frac{|L_j|}{|L|} \quad (8)$$

3.2. Classification and regression tree algorithm

This method, which forms a decision tree with binary divisions, was introduced by Briman et al. This method is designed for quantitative variables but is applicable to any type of variable. According to this algorithm, statistical software under the name of CART is also developed, which is one of the most popular applications (Chen et al., 2017c). In this method and for the qualitative response variable, the Gini Index is introduced as a criterion for selecting appropriate variables. In the introduction of the tree model with binary divisions, other parameters such as entropy can be used. The advantage of the Gini index over entropy and other indices is its higher speed in computation. The CART model can be introduced as one of the best-known classification models for diagnosis and prediction (Hess et al., 1999; Loh, 2014). In the CART model pruning of the classification, tree is done on the basis of Cost-Complexity and the accuracy of the introduced tree is introduced using the test sample. One disadvantage of the CART model is its bias in the selection of variables (Timofeev, 2004). In addition, in qualitative variables with more than two levels, the results can be confusing. Since several levels of a variable may be assigned to a node, this does not make it easy to interpret the results (Youssef et al., 2016).

3.3. Reduced error pruning tree algorithm

This method is proposed by Quinlan. First, the tree is allowed to grow well enough. Nodes are then pruned that do not increase the accuracy of the clustering. The data are divided into two test and training sets. The tree is trained with training data. It is then deleted for an internal node (non-branch n) of sub-node n deleted (Pham et al., 2019b). This sub-branch is replaced by a leaf. This leaflet is referred to as the majority examples category, meaning the category of most examples placed under this category. The tree performance is tested on test examples if the pruned tree performs better than or equal to the current tree, the pruned tree is used. Pruning continues until pruning is no longer beneficial (Elomaa and Kaariainen, 2001; Fürnkranz and Widmer, 1994). Figure 3 shows the sections of the current research in the study area.

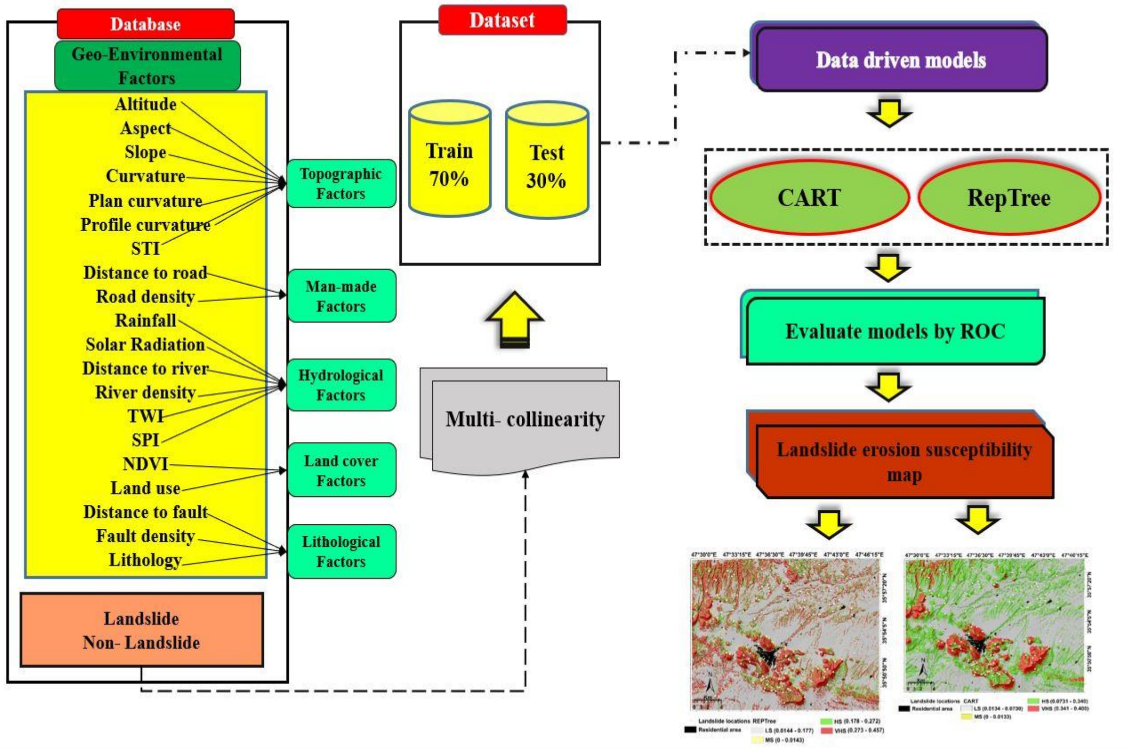


Figure 3. A flowchart of the methodology

3.4. Model validation and comparison

The knowledge generated during the model learning phase should be analyzed in the evaluation phase to determine its value and then determine the performance of the model learning algorithm. These criteria can be calculated for both the training data set at the learning stage and the test record set at the evaluation stage (Wu and Yen, 2009). We in this study used of the following measures to check the performance of the models:

3.4.1. Accuracy

The most important criterion for determining the efficiency of an accuracy classification algorithm is the classification rate, which calculates the accuracy of a whole cluster. In fact, this criterion is the most popular and most general criterion for calculating the efficiency of classification algorithms, which shows that the designed cluster correctly classifies a few percent of the entire set of experimental records. Classification accuracy is obtained using Equation 1, which states that the two values of TP and TN are the most important values that should be maximized in a two-dimensional problem (Wang et al., 2003).

$$CA = \frac{TN+Tp}{TN+FN+TP+FP} \tag{9}$$

where TP, FN, FP, and TN are true positives, false negatives, false positives and true negatives.

3.4.2. Precision

The Precision criterion measures the ratio of the number of "correct predictions" made for samples of a particular class to the number of "total predictions" for samples of the same class (this number includes all the correct predictions and false predictions). The high value for the accuracy criterion indicates the low number of data that is incorrectly categorized in a particular class (Pham et al., 2019).

$$Precision = \frac{TP}{TP+FP} \quad (10)$$

3.4.3. Recall

The Recall criterion represents the ratio of "the number of correctly categorized textual data" in a particular class to the total number of data that must be categorized in the same class. The high value for the Recall criterion indicates the low number of data that is not incorrectly categorized in that particular class. Using this criterion alone is not correct for evaluating system performance and should be used alongside the Precision criterion because it is easy to design textual classification models that are highly accurate and do not necessarily mean high precision (Stojanova et al., 2006).

$$Recall = \frac{TP}{TP+FN} \quad (11)$$

3.4.4. Kappa

The interrater reliability/precision of the data is estimated using kappa index. If the same data gives the same score the interrater will be happened among data. The kappa close to 1 indicates the degree of agreement between the data (Shirzadi et al., 2017b). The kappa can be computed according to the following equation (Shirzadi et al., 2018):

$$\text{kappa} = \frac{P_a - P_{est}}{1 - P_{est}} \quad (12)$$

$$Pa = (TP+TN) \Big/ (TP+TN+FN+FP)$$

$$P_{est} = \left((TP+FN)(TP+FP)+(FP+TN)(FN+TN) \Big/ \sqrt{(TP+TN+FN+FP)}\right) \quad (13)$$

3.4.5. Mean Absolute Error

Mean absolute error is an error-based measure to check the performance of the models. The lower the MAE will be, the better the performance of the model is. The MAE can be formulated as bellow (Bui et al., 2018):

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|x_i - x| \quad (14)$$

Where n is the number of errors, xi is the prediction and x is the true value.

3.4.6. Receiver Operating Characteristic (ROC)

Another important criterion used to determine the efficiency of a cluster is the Area under the Curve (AUC). The AUC represents the area under the Receiver Operating Characteristic (ROC) chart. The larger the value of a subclass, the more efficient the final clustering performance is evaluated. In fact, ROC curves are two-dimensional curves where DR is the True Positive Rate (TPR) on the Y axis and similarly False Positive Rate (FPR) which drawn on the X-axis (Janizadeh et al., 2019; Shirzadi et al., 2017b). The AUC value for a cluster that randomly determines the sample cluster under study is 0.5. The maximum value of this criterion is also equal to 1 and occurs for a condition that is ideal clustering and can detect all positive samples without any false warnings (Avand et al., 2019; Shirzadi et al., 2019).

4. Results and analysis

4.1. The most important factors using IGR technique for shallow landslide modeling

Conditioning factors play an important role in constructing the landslide susceptibility model. However, from the literature review, there is no guide to selecting these factors. The selections of these factors depend on several elements such as the availability of data, the condition of geomorphology, hydrology, climate, and human activities. 20 samples were generated from the four groups, so it's important to check these factors if some variables are useless and have similar values, which can reduce the performance and model and focus on the important factors. In this article, we used the IGR technique which is considered one of the most popular techniques for evaluating conditioning factors. It assigns a weighting to each factor to rank the prediction ability. Factors have higher values that are more important in the model. In this study, Slope angle, TWI, Curvature plane, LS, Curvature, Elevation, Solar radiation, Curvature profile, SPI, Aspect, Land use, rainfall are more important. In the study area, a landslide occurs in the high elevation zone with the loss of vegetation and the construction of humans (Fig. 4).

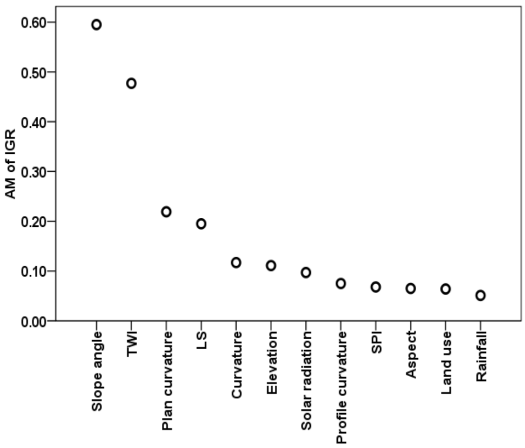


Figure 4. The most important conditioning factors using IGR technique for shallow landslide susceptibility mapping

4.2. Modelling process of shallow landslide

In this paper, the model is constructed by the training data and the test data is used to validate the models. Table 1 show some measures of statistics to evaluate how the model is performing. The CART model is better for training data (TP Rate = 0.932, FP Rate = 0.068, Precision = 0.933, Recall = 0.932, Fl-Mesaure = 0.932, AUC = 0.958, Kappa = 0.864, MAE = 0.1717) and validation dataset (TP Rate = 0.841, FP Rate = 0.158, Precision = 0.845, Recall = 0.841, Fl-Mesaure = 0.841, AUC = 0.839, Kappa = 0.673, MAE = 0.225), fallowed by REPTree with the traning data (TP Rate = 0.889, FP Rate = 0.111, Precision = 0.898, Recall = 0.889, Fl-Mesaure = 0.888, AUC = 0.889, Kappa = 0.778, MAE = 0.191) and the validation dataset ((TP Rate = 0.833, FP Rate = 0.167, Precision = 0.839, Recall = 0.833, Fl-Mesaure = 0.833, AUC = 0.833, Kappa = 0.667, MAE = 0.235) (Table 1).

Table 1 Modeling validation and comparison

	CART						REPTree					
	Training			Validation			Training			Validation		
	1*	0*	Average	1	0	Average	1	0	Average	1	0	Average
TP Rate	0.914	0.951	0.932	0.900	0.783	0.841	0.815	0.963	0.889	0.900	0.767	0.833
FP Rate	0.049	0.086	0.068	0.217	0.100	0.158	0.037	0.185	0.111	0.233	0.100	0.167
Precision	0.949	0.917	0.933	0.811	0.880	0.845	0.957	0.839	0.898	0.794	0.885	0.839
Recall	0.914	0.951	0.932	0.900	0.783	0.841	0.815	0.963	0.889	0.900	0.767	0.833
F1-Mesaure	0.931	0.933	0.932	0.831	0.850	0.841	0.880	0.897	0.888	0.844	0.821	0.833
AUC	0.958	0.958	0.958	0.839	0.839	0.839	0.889	0.889	0.889	0.833	0.833	0.833
Kappa	0.864			0.673			0.778			0.667		
MAE	0.117			0.225			0.191			0.235		

*1 and 0 are shallow landslide and non-shallow landslide

4.3. Development of Shallow landslide susceptibility mapping

After model validation, the CART and REPTree models are used to construct the landslide susceptibility map. This process is accomplished by feeding 12 selected conditioning factors out of 20 factors to the study area from two models. The values of susceptibility are variable from 0 to 1 which is reclassified in 4 groups using the natural break: low, moderate, high and very high. Figures 5 show the landslide maps using the CART and REPTree model. The landslide occurs mostly in the high elevation zone where vegetation loss and near the inhabitant area. These maps are useful for assisting land-use decision-makers and landslide risk management in reducing damage.

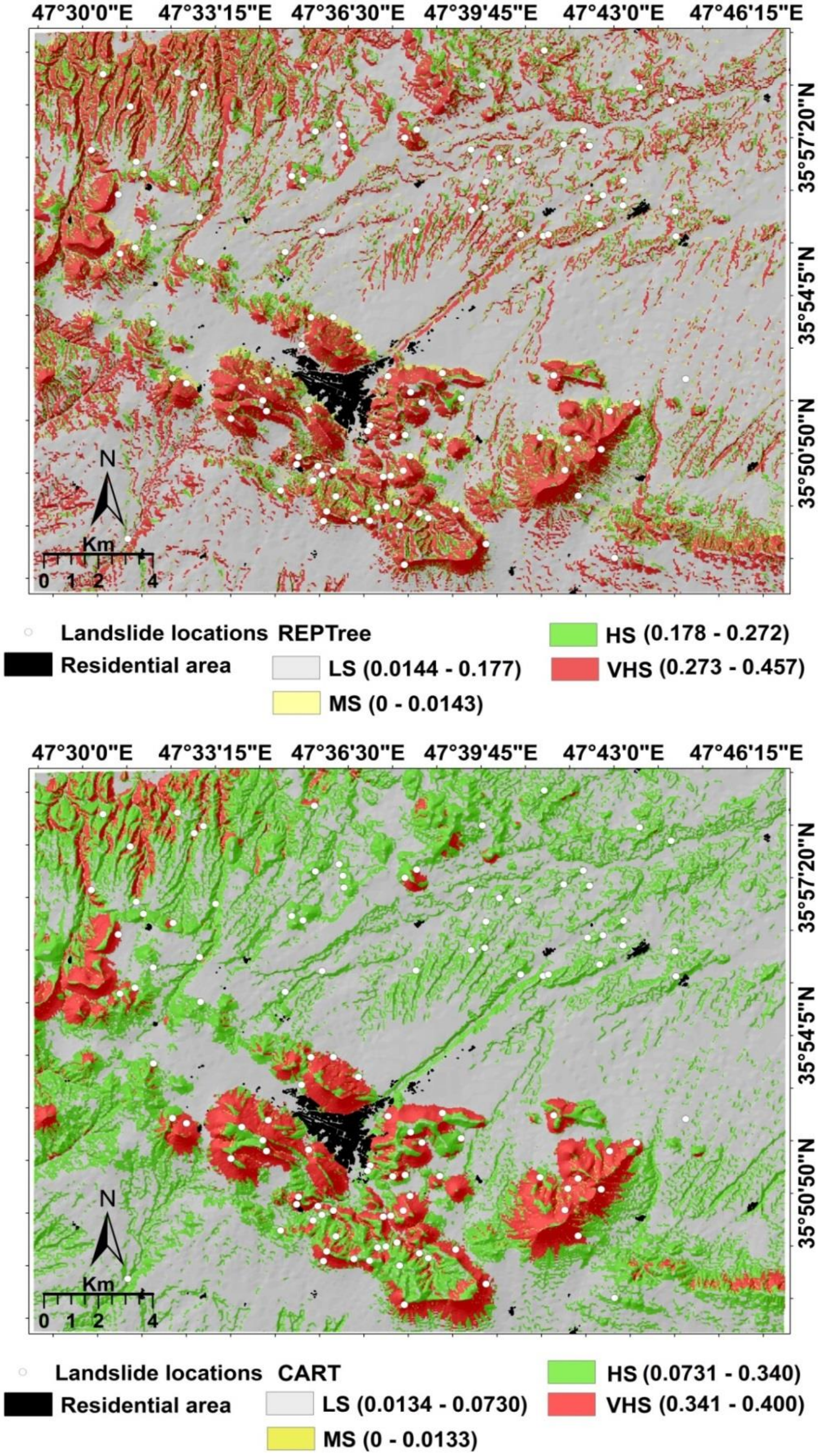


Figure 5. Shallow landslide susceptibility maps by REPTree and CART models

4.4. Evaluation of shallow landslide susceptibility maps

From the literature review, the predictive ability of the model is tested by the known technical: receiver operating characteristics (ROC). The performance of the model is presented by the sensitivity of the y-axes and the 1-specificity on the x-axes. The surfaces under the arch ROC is used to determine the model precisions with values from 0 to 1. If the values are equal to 1, the model is perfect. Figure 6 shows the AUC values of the two proposed models. The results indicate that the CART model is better than the REPTree model in terms of training data and validations data (AUC training data = 0.906 and AUC validation data = 0.856), followed by REPTree (AUC data from training = 0.857, AUC validation data = 0.837). Two models proposed are reasonable and reliable to build the map of landslide susceptibility in the mountain region in Iran where often touches by the landslide. However, risk management is still the limit. As a result, these methods can use as solutions for risk management Moreover, these models can apply to the other region in the world in general because the data is free and can collect from the global portal.

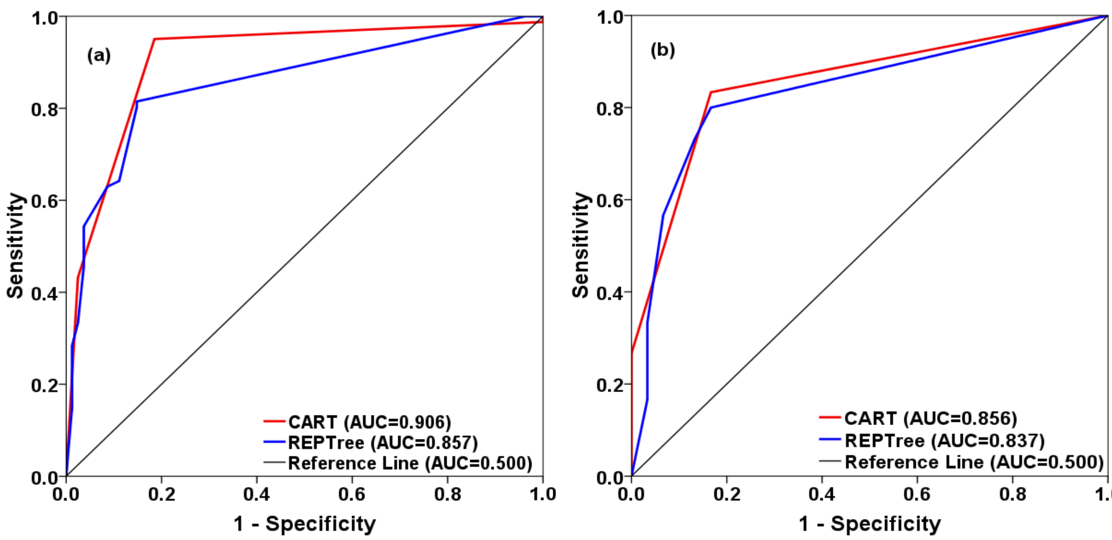


Figure 6. ROC curve and AUC: (a) training dataset, (b) validation dataset

5. Discussions

The results of the study of factors affecting landslide susceptibility showed that one of the factors affecting slope. The examination of the slope map in the study area shows that the higher slopes are more susceptible to landslide than the lower slopes. At low slopes, usually, the resistive forces such as soil friction are higher than the driving forces such as gravity. But on the higher slopes, the opposite was the case due to the climate and vegetation characteristics of the region, which provided high gravity forces for landslides. The results of this study are in agreement with the results of Schlögel et al. (2018); Tavoularis et al. (2018) and Kutlug Sahin and Colkesen, (2019). They also found that slope is one of the important factors affecting landslide sensitivity. The next factor that had the greatest impact on landslide sensitivity was TWI. The survey of landslide maps in the study area shows that the landslide susceptible areas in the study area are in accordance with the areas with high topographic moisture index. In areas with high TWI, due to increased soil moisture, soil erosion and soil stability increase. Areas with higher topographic wetness have a greater role in maintaining and concentrating subsurface waters, resulting in rapid water pressure within the pores. Surface landslides often occur in these areas. These findings are consistent with the results of Taalab et al. (2018) who showed that TWI is one of the important variables affecting landslide susceptibility in the study area.

Landslide susceptibility modeling was performed using two CART and REPTree data mining models. Statistical criteria were used to evaluate the efficiency of the models. The results of the modeling showed that both models had a good performance in determining landslide zones, however, the CART model had higher performance than the REPTree models. The obtained result is in agreement with Pham et al. (2018a). The main advantage of the CART model is inherently a non-parametric method that can control very small or multidimensional numerical data (Loh, 2014). In addition, no assumptions should be made during the learning process regarding the fundamental distribution of predictive variable values. Therefore, it can eliminate time analysis to determine whether variables are normally distributed for classification (Lewis, 2000). Similarly, CART is able to search for all possible variables to identify "split" variables. It is also effective in dealing with missing variables. In addition, CART is a fast and simple data mining method (Lewis, 2000). Various researchers, such as (Pham et al., 2018a; Youssef et al., 2016) in the study of natural phenomena, have shown that the CART model can predict these phenomena well.

6. Conclusions

Landslides are one of the major geographical hazards that cause the loss of lives and property around the world. Landslide susceptibility mapping is a useful tool in reducing landslide damage through land-use planning and decision making. In the present study, in order to map the susceptibility to landslide hazards in Bjar, Kurdistan province, Iran, 12 important conditioning factors were used. The CART and REPTree data mining models were used for data analysis and modeling. According to the results of the analysis, the slope and TWI variables are of great importance in determining landslide risk in the study area. It can also be concluded that the CART model is a promising method for predicting the location of landslides that can be used in other landslides. Its performance is even better than other machine learning techniques (REPTree). The landslide susceptibility map obtained from this study can be of great help to managers and planners in better decision making and management and prevention of potential hazards.

Acknowledgments

The authors would like to express their deepest gratitude to the Forests, Rangelands, and Watershed Management Organization of Iran for providing the reports of landslide locations in the study area.

References

Abedini M., Ghasemian B., Shirzadi A., Bui D.T., 2019. A comparative study of support vector machine and logistic model tree classifiers for shallow landslide susceptibility modeling. Environmental Earth Sciences, 78, 560.

Ahmadlou M., Karimi M., Alizadeh S., Shirzadi A., Parvinnezhad D., Shahabi H., Panahi M., 2018. Flood susceptibility assessment using integration of adaptive network-based fuzzy inference system (ANFIS) and biogeography-based optimization (BBO) and BAT algorithms (BA). Geocarto International, 1–38.

Barbon Jr S., Barbon A., Mantovani R.G., Barbin D.F., 2016. Comparison of SVM and REPTree for classification of poultry quality. Proceedings of the

Modelling, Simulation and Identification/Intelligent Systems and Control (MSI 2016. Doi: 10.2316/P.2016.840-039.

Barik M., Adam J., Barber M., Muhunthan B., 2017. Improved landslide susceptibility prediction for sustainable forest management in an altered climate. Engineering geology, 230, 104–117.

Bloechl A., Braun B., 2005. Economic assessment of landslide risks in the Swabian Alb, Germany - Research framework and first results of homeowners' and experts' surveys. Natural Hazards and Earth System Science, 5.

Breiman L., Friedman J., Olshen R., Stone C., 1984. Classification and regression trees. Wadsworth & Brooks. Cole Statistics/Probability Series.

Bui D.T., Panahi M., Shahabi H., Singh V.P., Shirzadi A., Chapi K., Khosravi K., Chen W., Panahi S., Li S., 2018. Novel hybrid evolutionary algorithms for spatial prediction of floods. Scientific reports, 8, 15364.

Chen W., 2016. GIS-based landslide susceptibility mapping using analytical hierarchy process (AHP) and certainty factor (CF) models for the Baozhong region of Baoji City, China. Environmental Earth Sciences, 75.

Chen W., Li W., Hou E., Zhao Z., Deng N., Bai H., Wang D., 2014. Landslide susceptibility mapping based on GIS and information value model for the Chencang District of Baoji, China. Arabian Journal of Geosciences, 7, 4499–4511.

Chen W., Li Y., Tsangaratos P., Shahabi H., Ilia I., Xue W., Bian H., 2020a. Groundwater Spring Potential Mapping Using Artificial Intelligence Approach Based on Kernel Logistic Regression, Random Forest, and Alternating Decision Tree Models. Applied Sciences, 10, 425.

Chen W., Li Y., Xue W., Shahabi H., Li S., Hong H., Wang X., Bian H., Zhang S., Pradhan B., 2020b. Modeling flood susceptibility using data-driven approaches of naïve bayes tree, alternating decision tree, and random forest methods. Science of The Total Environment, 701, 134979.

Chen W., Panahi M., Pourghasemi H.R., 2017a. Performance evaluation of GIS-based new ensemble data mining techniques of adaptive neuro-fuzzy inference system (ANFIS) with genetic algorithm (GA), differential evolution (DE), and particle swarm optimization (PSO) for landslide spatial modelling. CATENA, 157, 310–324.

Chen W., Shahabi H., Shirzadi A., Hong H., Akgun A., Tian Y., Liu J., Zhu A.-X., Li S., 2019a. Novel hybrid artificial intelligence approach of bivariate statistical-methods-based kernel logistic regression classifier for landslide susceptibility modeling. Bulletin of Engineering Geology and the Environment, 78, 4397–4419.

Chen W., Shahabi H., Shirzadi A., Li T., Guo C., Hong H., Li W., Pan D., Hui J., Ma M., Xi M., Ahmad B.B., 2018a. A Novel Ensemble Approach of Bivariate Statistical Based Logistic Model Tree Classifier for Landslide Susceptibility Assessment. Geocarto International, 1–32.

Chen W., Shahabi H., Zhang S., Khosravi K., Shirzadi A., Chapi K., Pham B., Zhang T., Zhang L., Chai H., 2018b. Landslide susceptibility modeling based on gis and novel bagging-based kernel logistic regression. Applied Sciences, 8, 2540.

Chen W., Shirzadi A., Shahabi H., Ahmad B.B., Zhang S., Hong H., Zhang N., 2017b. A novel hybrid artificial intelligence approach based on the rotation forest ensemble and naïve Bayes tree classifiers for a landslide susceptibility assessment in Langao County, China. Geomatics, Natural Hazards and Risk, 8, 1955–1977.

Chen W., Xie X., Wang J., Pradhan B., Hong H., Bui D.T., Duan Z., Ma J., 2017c. A comparative study of logistic model tree, random forest, and classification and regression tree models for spatial prediction of landslide susceptibility. Catena, 151, 147–160.

Chen W., Zhao X., Shahabi H., Shirzadi A., Khosravi K., Chai H., Zhang S., Zhang L., Ma J., Chen Y., 2019b. Spatial prediction of landslide susceptibility by combining evidential belief function, logistic regression and logistic model tree. Geocarto International, 1–25.

Chen W., Zhao X., Tsangaratos P., Shahabi H., Ilia I., Xue W., Wang X., Ahmad B.B., 2020c. Evaluating the usage of tree-based ensemble methods in groundwater spring potential mapping. Journal of Hydrology, 124602.

Choubin B., Abdolshahnejad M., Moradi E., Querol X., Mosavi A., Shamshirband S., Ghamisi P., 2020. Spatial hazard assessment of the PM10 using machine learning models in Barcelona, Spain. Science of The Total Environment, 701, 134474.

Choubin B., Darabi H., Rahmati O., Sajedi-Hosseini F., Kløve B., 2018a. River suspended sediment modelling using the CART model: A comparative study of machine learning techniques. Science of the Total Environment, 615, 272–281.

Choubin B., Moradi E., Golshan M., Adamowski J., Sajedi-Hosseini F., Mosavi A., 2019a. An ensemble prediction of flood susceptibility using multivariate discriminant analysis, classification and regression trees, and support vector machines. Science of the Total Environment, 651, 2087–2096.

Choubin B., Mosavi A., Alamdarloo E.H., Hosseini F.S., Shamshirband S., Dashtekian K., Ghamisi P., 2019b. Earth fissure hazard prediction using machine learning models. Environmental research, 179, 108770.

Choubin B., Rahmati O., Soleimani F., Alilou H., Moradi E., Alamdari N., 2019c. Regional groundwater potential analysis using classification and regression trees, Spatial modeling in GIS and R for earth and environmental sciences. Elsevier, 485–498.

Choubin B., Zehtabian G., Azareh A., Rafiei-Sardooi E., Sajedi-Hosseini F., Kişi Ö., 2018b. Precipitation forecasting using classification and regression trees (CART) model: a comparative study of different approaches. Environmental Earth Sciences, 77, 314.

Cloke H., Pappenberger F., 2009. Ensemble flood forecasting: A review. Journal of hydrology, 375, 613–626.

Devasena C.L., 2014. Comparative analysis of random forest, REP tree and J48 classifiers for credit risk prediction. International Journal of Computer Applications, 0975–8887.

Elomaa, T., Kaariainen, M., 2001. An analysis of reduced error pruning. Journal of Artificial Intelligence Research, 15, 163–187.

Ercanoglu M., Gokceoglu C., 2004. Use of fuzzy relations to produce landslide susceptibility map of a landslide prone area (West Black Sea Region, Turkey). Engineering Geology, 75, 229–250.

Fürnkranz J., Widmer G., 1994. Incremental reduced error pruning, Machine Learning Proceedings. Elsevier, 70–77.

Guzzetti F., Reichenbach P., Ardizzone F., Cardinali M., Galli M., 2006. Estimating the quality of landslide susceptibility models. Geomorphology, 81, 166–184.

Han J., Kamber M., 2001. Data mining concepts and techniques San Francisco Moraga Kaufman.

Hemasinghe H., Rangali R.S.S., Deshapriya N.L., Samarakoon L., 2018. Landslide susceptibility mapping using logistic regression model (a case study in Badulla District, Sri Lanka). Procedia Engineering, 212, 1046–1053.

Hess K.R., Abbruzzese M.C., Lenzi R., Raber M.N., Abbruzzese J.L., 1999. Classification and regression tree analysis of 1000 consecutive patients with unknown primary carcinoma. Clinical Cancer Research, 5, 3403–3410.

Hong H., Chen W., Xu C., Youssef D.A., Pradhan B., Tien Bui D., 2016. Rainfall-induced landslide susceptibility assessment at the Chongren area (China) using frequencyratio, certainty factor, and index of entropy. Geocarto International, 32, 139–154.

Hong H., Panahi M., Shirzadi A., Ma T., Liu J., Zhu A.-X., Chen W., Kougias I., Kazakis N., 2018. Flood susceptibility assessment in Hengfeng area coupling adaptive neuro-fuzzy inference system with genetic algorithm and differential evolution. Science of The Total Environment, 621, 1124–1141.

Jaafari A., Najafi A., Rezaeian J., Sattarian A., Ghajar I., 2015. Planning road networks in landslide-prone areas: A case study from the northern forests of Iran. Land Use Policy, 47, 198–208.

Jaafari A., Zenner E.K., Pham B.T., 2018. Wildfire spatial pattern analysis in the Zagros Mountains, Iran: A comparative study of decision tree based classifiers. Ecological Informatics, 43, 200–211.

Khosravi K., Melesse A.M., Shahabi H., Shirzadi A., Chapi K., Hong H., 2019a. Flood susceptibility mapping at Ningdu catchment, China using bivariate and data mining techniques, Extreme Hydrology and Climate Variability. Elsevier, 419–434.

Khosravi K., Shahabi H., Pham B.T., Adamowski J., Shirzadi A., Pradhan B., Dou J., Ly H.-B., Gróf G., Ho H.L., 2019b. A comparative assessment of flood susceptibility modeling using Multi-Criteria Decision-Making Analysis and Machine Learning Methods. Journal of Hydrology, 573, 311–323.

Kutlug Sahin E., Colkesen I., 2019. Performance analysis of advanced decision tree-based ensemble learning algorithms for landslide susceptibility mapping. Geocarto International, 1–23.

Le T.-T., Pham B.T., Ly H.-B., Shirzadi A., Le L.M., 2020. Development of 48-hour Precipitation Forecasting Model using Nonlinear Autoregressive Neural Network, CIGOS 2019, Innovation for Sustainable Infrastructure. Springer, 1191–1196.

Lee T.-S., Chiu C.-C., Chou Y.-C., Lu C.-J., 2006. Mining the customer credit using classification and regression tree and multivariate adaptive regression splines. Computational Statistics & Data Analysis, 50, 1113–1130.

Lewis R.J., 2000. An introduction to classification and regression tree (CART) analysis, Annual meeting of the society for academic emergency medicine in San Francisco, California.

Loh W.Y., 2014. Classification and regression tree methods. Wiley StatsRef: Statistics Reference Online.

Miraki S., Zanganeh S.H., Chapi K., Singh V.P., Shirzadi A., Shahabi H., Pham B.T., 2019. Mapping groundwater potential using a novel hybrid intelligence approach. Water resources management, 33, 281–302.

Moore I., Burch G., 1986. Sediment transport capacity of sheet and rill flow: application of unit stream power theory. Water Resources Research, 22, 1350–1360.

Nguyen M.D., Pham B.T., Tuyen T.T., Yen H., Phan H., Prakash I., Vu T.T., Chapi K., Shirzadi A., Shahabi H., 2019a. Development of an Artificial Intelligence Approach for Prediction of Consolidation Coefficient of Soft Soil: A Sensitivity Analysis. The Open Construction and Building Technology Journal, 13.

Nguyen P.T., Tuyen T.T., Shirzadi A., Pham B.T., Shahabi H., Omidvar E., Amini A., Entezami H., Prakash I., Phong T.V., 2019b. Development of a novel hybrid intelligence approach for landslide spatial prediction. Applied Sciences, 9, 2824.

Nguyen V.V., Pham B.T., Vu B.T., Prakash I., Jha S., Shahabi H., Shirzadi A., Ba D.N., Kumar, R., Chatterjee J.M., 2019c. Hybrid machine learning approaches for landslide susceptibility modeling. Forests, 10, 157.

Ozdemir A., Altural T., 2013. A comparative study of frequency ratio, weights of evidence and logistic regression methods for landslide susceptibility mapping: Sultan Mountains, SW Turkey. Journal of Asian Earth Sciences, 64, 180–197.

Pham B.T., Bui D.T., Prakash I., 2018a. Application of Classification and Regression Trees for Spatial Prediction of Rainfall-Induced Shallow Landslides in the Uttarakhand Area (India) Using GIS, Climate Change, Extreme Events and Disaster Risk Reduction. Springer, 159–170.

Pham B.T., Prakash I., Dou J., Singh S.K., Trinh P.T., Tran H.T., Le T.M., Van Phong T., Khoi D.K., Shirzadi A., 2019a. A novel hybrid approach of landslide susceptibility modelling using rotation forest ensemble and different base classifiers. Geocarto International, 1–25.

Pham B.T., Prakash I., Singh S.K., Shirzadi A., Shahabi H., Bui D.T., 2019b. Landslide susceptibility modeling using Reduced Error Pruning Trees and different ensemble techniques: Hybrid machine learning approaches. Catena, 175, 203–218.

Pham B.T., Shirzadi A., Bui D.T., Prakash I., Dholakia M., 2018b. A hybrid machine learning ensemble approach based on a radial basis function neural network and rotation forest for landslide susceptibility modeling: A case study in the Himalayan area, India. International Journal of Sediment Research, 33, 157–170.

Pourghasemi H.R., Mohammady M., Pradhan B., 2012. Landslide susceptibility mapping using index of entropy and conditional probability models in GIS: Safarood Basin, Iran. Catena, 97, 71–84.

Quinlan J.R., 1986. Induction of decision trees. Machine learning, 1, 81–106.

Regmi A., Devkota K., Yoshida K., Pradhan B., Pourghasemi H.R., Kumamoto T., Akgun A., 2013. Application of frequency ratio, statistical index, and weights-of-evidence models and their comparison in landslide susceptibility mapping in Central Nepal Himalaya. Arabian Journal of Geoscience, 7.

Schlögel R., Marchesini I., Alvioli M., Reichenbach P., Rossi M., Malet J.-P., 2018. Optimizing landslide susceptibility zonation: Effects of DEM spatial resolution and slope unit delineation on logistic regression models. Geomorphology, 301, 10–20.

Shafizadeh-Moghadam H., Valavi R., Shahabi H., Chapi K., Shirzadi A., 2018. Novel forecasting approaches using combination of machine learning and statistical models for flood susceptibility mapping. Journal of environmental management, 217, 1–11.

Shahabi H., Shirzadi A., Ghaderi K., Omidvar E., Al-Ansari N., Clague J.J., Geertsema M., Khosravi K., Amini A., Bahrami S., 2020. Flood Detection and Susceptibility Mapping Using Sentinel-1 Remote Sensing Data and a Machine Learning Approach: Hybrid Intelligence of Bagging Ensemble Based on K-Nearest Neighbor Classifier. Remote Sensing, 12, 266.

Shirzadi A., Chapi K., Shahabi H., Solaimani K., Kavian A., Ahmad B.B., 2017a. Rock fall susceptibility assessment along a mountainous road: an evaluation of bivariate statistic, analytical hierarchy process and frequency ratio. Environmental Earth Sciences, 76.

Shirzadi A., Saro L., Joo O.H., Chapi K., 2012. A GIS-based logistic regression model in rock-fall susceptibility mapping along a mountainous road: Salavat Abad case study, Kurdistan, Iran. Natural hazards, 64, 1639–1656.

Shirzadi A., Shahabi H., Chapi K., Bui D.T., Pham B.T., Shahedi K., Ahmad B.B., 2017b. A comparative study between popular statistical and machine learning methods for simulating volume of landslides. CATENA, 157, 213–226.

Shirzadi A., Solaimani K., Roshan M.H., Kavian A., Chapi K., Shahabi H., Keesstra S., Ahmad B.B., Bui D.T., 2019. Uncertainties of prediction accuracy in shallow landslide modeling: Sample size and raster resolution. Catena, 178, 172–188.

Shirzadi A., Soliamani K., Habibnejhad M., Kavian A., Chapi K., Shahabi H., Chen W., Khosravi K., Thai Pham B., Pradhan B., 2018. Novel GIS based machine learning algorithms for shallow landslide susceptibility mapping. Sensors, 18(11), 3777. https://doi.org/10.3390/s18113777.

Stojanova D., Panov P., Kobler A., Dzeroski S., Taskova K., 2006. Learning to predict forest fires with different data mining techniques, Conference on data mining and data warehouses (SiKDD 2006), Ljubljana, Slovenia, 255–258.

Taalab K., Cheng T., Zhang Y., 2018. Mapping landslide susceptibility and types using Random Forest. Big Earth Data, 2, 159–178.

Taheri K., Shahabi H., Chapi K., Shirzadi A., Gutiérrez F., Khosravi K., 2019. Sinkhole susceptibility mapping: A comparison between Bayes-based machine learning algorithms. Land Degradation & Development, 30, 730–745.

Tavoularis N., Koumantakis I., Rozos D., Koukis G., 2018. The Contribution of Landslide Susceptibility Factors Through the Use of Rock Engineering System (RES) to the Prognosis of Slope Failures: An Application in Panagopoula and Malakasa Landslide Areas in Greece. Geotechnical and Geological Engineering, 36, 1491–1508.

Thai Pham B., Shirzadi A., Shahabi H., Omidvar E., Singh S.K., Sahana M., Talebpour Asl D., Bin Ahmad B., Kim Quoc N., Lee S., 2019. Landslide susceptibility assessment by novel hybrid machine learning algorithms. Sustainability, 11(16), 4386. https://doi.org/10.3390/su11164386.

Tien Bui D., Khosravi K., Shahabi H., Daggupati P., Adamowski J.F., Melesse A.M., Thai Pham B., Pourghasemi H.R., Mahmoudi M., Bahrami S., 2019a. Flood spatial modeling in northern Iran using remote sensing and gis: A comparison between evidential belief functions and its ensemble with a multivariate logistic regression model. Remote Sensing, 11(13), 1589. https://doi.org/10.3390/rs11131589.

Tien Bui D., Shahabi H., Omidvar E., Shirzadi A., Geertsema M., Clague J.J., Khosravi K., Pradhan B., Pham B.T., Chapi K., 2019b. Shallow landslide prediction using a novel hybrid functional machine learning algorithm. Remote Sensing, 11(8), 931. https://doi.org/10.3390/rs11080931.

Tien Bui D., Shahabi H., Shirzadi A., Chapi K., Alizadeh M., Chen W., Mohammadi A., Ahmad B., Panahi M., Hong H., 2018a. Landslide detection and susceptibility mapping by airsar data using support vector machine and index of entropy models in cameron highlands, malaysia. Remote Sensing, 10(10), 1527. https://doi.org/10.3390/rs10101527.

Tien Bui D., Shahabi H., Shirzadi A., Chapi K., Pradhan B., Chen W., Khosravi K., Panahi M., Bin Ahmad B., Saro L., 2018b. Land subsidence susceptibility mapping in south korea using machine learning algorithms. Sensors, 18(8), 2464. https://doi.org/10.3390/s18082464.

Tien Bui D., Shahabi H., Shirzadi A., Chapi K., Pradhan B., Chen W., Khosravi K., Panahi M., Id B., Bin A., Lee C.-W., Lee S., 2018c. Land Subsidence Susceptibility Mapping in South Korea Using Machine Learning Algorithms. Sensors, 18(8), 2464.

Tien Bui D., Shirzadi A., Chapi K., Shahabi H., Pradhan B., Pham B.T., Singh V.P., Chen W., Khosravi K., Bin Ahmad B., 2019c. A Hybrid Computational Intelligence Approach to Groundwater Spring Potential Mapping. Water, 11(10), 2013. https://doi.org/10.3390/w11102013.

Tien Bui D., Shirzadi A., Shahabi H., Chapi K., Omidavr, E., Pham B.T., Talebpour Asl D., Khaledian H., Pradhan B., Panahi M., 2019d. A Novel Ensemble Artificial Intelligence Approach for Gully Erosion Mapping in a Semi-Arid Watershed (Iran). Sensors, 19(11), 2444. https://doi.org/10.3390/s19112444.

Tien Bui D., Shirzadi A., Shahabi H., Geertsema M., Omidvar E., Clague J.J., Thai Pham B., Dou J., Talebpour Asl D., Bin Ahmad B., 2019e. New Ensemble Models for Shallow Landslide Susceptibility Modeling in a Semi-Arid Watershed. Forests, 10(9), 743. https://doi.org/10.3390/f10090743.

Timofeev R., 2004. Classification and regression trees (CART) theory and applications. Humboldt University, Berlin.

Varnes D.J., Radbruch-Hall D.H., Savage W.Z., 1989. Topographic and structural conditions in areas of gravitational spreading of ridges in the western United States. United States Geological Survey, Professional Paper, (USA), 1496.

Wang J.-H., Deng P.S., Fan Y.-S., Jaw L.-J., Liu Y.-C., 2003. Virus detection using data mining techinques, IEEE 37th Annual 2003 International Carnahan Conference onSecurity Technology. Proceedings. IEEE, 71–76.

Wang Y., Hong H., Chen W., Li S., Panahi M., Khosravi K., Shirzadi A., Shahabi H., Panahi S., Costache R., 2019. Flood susceptibility mapping in dingnan county (China) using adaptive neuro-fuzzy inference system with biogeography based optimization and imperialistic competitive algorithm. Journal of environmental management, 247, 712–729.

Wu S.-Y., Yen E., 2009. Data mining-based intrusion detectors. Expert Systems with Applications, 36, 5605–5612.

Youssef A.M., Pourghasemi H.R., Pourtaghi Z.S., Al-Katheeri M.M., 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, 839–856.

Zhao Y., Zhang Y., 2008. Comparison of decision tree methods for finding active objects. Advances in Space Research, 41, 1955–1959.