Environmental Earth Sciences (2022) 81:51
https://doi.org/10.1007/s12665-021-10152-4

ORIGINAL ARTICLE

A comparison study on the quantitative statistical methods for spatial prediction of shallow landslides (case study: Yozidar-Degaga Route in Kurdistan Province, Iran)

Mitra Asadi[1] · Leila Goli Mokhtari[1] · Ataollah Shirzadi[2] · Himan Shahabi[3,4] · Shahram Bahrami[5]

Received: 17 July 2020 / Accepted: 27 December 2021 / Published online: 19 January 2022

Abstract

The main purpose of this study was to compare the performance of Support Vector Machines (SVM), Stochastic Gradient Descent (SGD), and Bayesian Logistic Regression (BLR) algorithms for landslide susceptibility modeling in the Yozidar-Degaga region, Iran. Initially, a distribution map with 175 landslides and 175 non-landslide locations was prepared and the data were classified into a ratio of 80% and 20% for training and model validation, respectively. Based on Information Gain Ratio (IGR) technique, 13 derived factors from topographic data, land cover and rainfall were selected for modeling. Then, the SVM, SGD, and BLR algorithms were selected based on size of the data and required accuracy of the output, to learn and prepare landslide susceptibility maps. Statistical criteria were employed to evaluate the models for both training and validation datasets. Finally, the performance of these models was evaluated by the area under the receiver operating curve (AUC). The results showed that SVM algorithm (AUC = 0.920) performed better than SGD (AUC = 0.918) and BLR (AUC = 0.918) algorithms. Therefore, the SVM model can be suggested as a useful tool for better management of landslide-affected areas in the study area. In this study, all three models (SVM, SGD and BLR) were implemented in WEKA 3.6.9 software environment to prepare landslide susceptibility maps.

Keywords Spatial prediction · Shallow landslide · Bayesian logistic regression · Stochastic gradient descend · Support vector machine · Prediction accuracy · Risk management

Introduction

Landslide, as one of the most important types of mass movements, is the down-slope movement of a mass of soil, rock and debris which can be affected by gravity (Varnes 1958). Numerous factors such as geological, morphological, hydrological conditions, topography of the region, climatic conditions, etc. cause slope instability and landslide occurrence, but only one factor causes the landslide to start. The factors that prepare the ground for a landslide and make the

✉ Leila Goli Mokhtari
L.mokhtari@hsu.ac.ir

Mitra Asadi
67avin@gmail.com

Ataollah Shirzadi
a.shirzadi@uok.ac.ir

Himan Shahabi
h.shahabi@uok.ac.ir

Shahram Bahrami
sh_Bahrami@sbu.ac.ir

[1] Department of Climatology and Geomorphology, Faculty of Geography and Environmental Science, Hakim Sabzevari University, 9617976487 Sabzevar, Iran

[2] Department of Rangeland and Watershed Management, Faculty of Natural Resources, University of Kurdistan, Sanandaj, Iran

[3] Department of Geomorphology, Faculty of Natural Resources, University of Kurdistan, Sanandaj, Iran

[4] Department of Zrebar Lake Environmental Research, Kurdistan Studies Institute, University of Kurdistan, Sanandaj, Iran

[5] Department of Physical Geography, Faculty of Earth Sciences, Shahid Beheshti University, Tehran, Iran

slope vulnerable are intrinsic factors, and those that initiate a landslide are called trigger or stimulus factors (Turner and Shuster 1996).

According to the Emergency Events Database at Université Catholique de Louvain (UCL)-CRED, from 2008 to 2017 (OFDA/CRED 2018), landslides have caused 10,338 deaths worldwide, with more than 3 million citizens affected and more than US$ 2.7 billion economic losses, also landslides caused by heavy rain killed 32 people in Ethiopia in May 2018 (Tiranti and Cremononi 2019). In addition, the Emergency Events Database reported that, landslides worldwide have caused 66,438 deaths and approximately 10.8 billion U.S. dollars economic loss from 1900 to 2020 (Guha-Sapir et al. 2020). In Iran, the Alpine–Himalayan seismic belt is mainly responsible for landslide occurrence so that it has been suffered about 12.7 billion $ economic loss from 4900 landslides only in 2007 (Farrokhnia et al. 2011). In land use planning, economic, social, and environmental parameters must be considered simultaneously (Bathrellos et al. 2012; Skilodimou et al. 2019). Therefore, it is necessary to use a proper strategy to reduce the damages caused by landslides, starting with the identification of areas prone to landslides such as landslide susceptibility mapping (LSM). LSM is the first step for assessing the risk and controlling landslides, which are useful in landslide hazard assessment (Anbalagan et al. 2015; Shadman Roodposhti et al. 2016). In this way, the land surface is divided into separate areas and ranked based on the actual degree or potential risk of landslides on slopes (Yalcin 2008). An accurate LSM can recognize the susceptible and high-risk areas, to be used by managers to reduce damages by providing solutions and controlling methods.

The methods that have been used and suggested for the LSM are generally divided into quantitative and qualitative methods. Qualitative models are thematic and typically use landslide inventory to detect areas prone to landslides with similar topographic, geological, and geomorphological features.

Some qualitative models such as the analytical hierarchical process (AHP) (Rozos et al. 2011; Zhang et al. 2016) are also considered as the expert-based approach, and might be called semi-quantitative models (Tamene et al. 2011). Quantitative statistical methods including probabilistic and definitive are based on mathematics and they are very useful in predicting landslide event. Due to the need for accurate geological data, the use of these models in large areas is difficult (Schilirò et al. 2016). Accordingly, data-driven machine learning and soft-computing methods are widely applied for landslide susceptibility assessment (Marjanović et al. 2011; Pham et al. 2019).

The most common quantitative statistical techniques and methods used in preparation of LSM are: statistical index (SI) (Pourghasemi et al. 2013a, b), logistic regression (LR) (Wang et al. 2015), certainty factor (CF) (Hong et al. 2017), bivariate statistical analysis (BSA) (Ayalew and Yamagishi 2005), frequency ratio (FR) (Pradhan and Lee 2010), multivariate adaptive regression spline (MARS) (Felicísimo et al. 2013), index of entropy (IOE), multivariate regression (MR) (Akgün and Türk 2011), discriminant analysis (DA) (Dong et al. 2009), spatial multi-criteria evaluation (SME) (Nsengiyumva et al. 2018), weight of evidence (WOE) (Kayastha et al. 2012), and evidential belief functions (EBFs) (Pourghasemi and Kerle 2016).

Among the machine learning techniques that are a subset of artificial intelligence, the following ones are worth to be mentioned: (1) artificial neural network (ANN) (Nhu et al. 2020b); (2) adaptive neuro-fuzzy inference (ANFIS) (Jaafari et al. 2019); (3) naive bayes (NB) (Tsangaratos and Ilia 2016); (4) random forest (RF) (Nhu et al. 2020c); (5) radial basis function (RBF) (Wang et al. 2020); (6) support vector machine (SVM) (Huang and Zhao 2018; Kavzoglu et al. 2014); (7) logistic model tree (LMT) (Chen et al. 2017); (8) random subspace (RS) (Pham et al. 2018; Tien Bui et al. 2019b); (9) alternating decision tree (ADT) (Nhu et al. 2020a); (10) reduced error pruning tree (REPT) (Tien Bui et al. 2019a); (11) bayesian logistic regression (BLR) (Das et al. 2012); (12) grey wolf optimizer (GWO) (Liao et al. 2020); and (13) random gradient descent (RGD) (Hong et al. 2020). There is no agreement among researchers to choose the best model (Pham et al. 2016) and each of the mentioned models has different disadvantages and advantages. Table 1 shows advantages and disadvantages of models previously used for LSM.

To reach a satisfactory conclusion, Bui et al. (2012) emphasized that a good model depends not only on the quality of the dataset but also on the structure of the selected model. Das et al. (2012) used the BLR model to map landslide-susceptible areas along the Himalayan roads in India. BLR results compared to LR show that BLR is better in estimating parameters and especially in estimating uncertainty. Goetz et al. (2015) compared traditional statistical models with new statistical machine learning models for modeling landslide susceptibility in three areas of Austria. Lee et al. (2017) applied the SVM model for landslide susceptibility mapping in two study areas in Korea. The obtained results from SVM model showed that approximately 81.36% and 77.49% in the PyeongChang and Inje areas, respectively, was predicted correctly. These results indicate that SVMs can be useful and effective for landslide susceptibility analysis. Chen et al. (2018a) used support vector machine (SVM) with four kernel functions (linear-SVM, polynomial-SVM, radial basic function-SVM, and sigmoidal-SVM), and entropy models in landslide susceptibility mapping, in Shangzhou District, China. The results indicated that the entropy model had the highest success rate (0.7610), followed by polynomial-SVM (0.7526), the sigmoidal-SVM (0.7518), radial

Table 1 Some advantages and disadvantages of models previously used for LSM

Methods	Advantage	Disadvantage	Reference/s
Mora-Varson	Suitable for large areas	Inflexibility and inefficiency in low and non-uniform areas	Wu et al. (2014) Farhadi Nezhad et al. (2011)
Anbalagan	Suitable for zoning the risk of slope instabilities	Inflexible use only in mountainous conditions	Deljoee et al. (2016) Khezri et al. (2006)
AHP	Unity interdependency Hierarchy structure Consistency synthesis Balance (tradeoff) Judgment and consensus Repetition	The problem of standardizing units of measurement Possibility of expert error in determining weight Lack of standards for measuring quality criteria	Arjmandzadeh et al. (2019) Amir Ahmadi et al. (2010) Atash Afrooz and Safaeipour (2021)
LR	High flexibility with a variety of variables Simple implementation High accuracy and strong interpretation ability	Possibility of meaningless predicted values and lack of interpretability in practice Impossibility of comparing forecast values with each other Ignoring the relationship between the categories of each criterion	Pourtaghi and Pourghasemi (2014) Eker et al. (2015) Fang et al. (2020)
Weights-of-Evidence	Ability to use variable quantitative and qualitative data in complex systems Powerful visual display of variables and relationships between them in a system This model is suitable for LSM because its uncertainty is connected with landslide events and their associations with the complex landscape	Ignoring the impact of each criterion on modeling Ignoring feedback loops in system variables, for example, the effect of a variable returning to the predictor variable cannot be examined	Xu et al. (2012) Pourghasemi et al. (2013a, b) Razavizadeh et al. (2017) Mansoori and Shirani (2016)
Bayesian	Easy and fast Categorization of test data Ability to use previous information to achieve accurate results	Approximate impossibility of achieving the condition of independence in the real world Requires basic knowledge of many possibilities	Lee et al. (2002) Pourghasemi et al. (2012)
LMT	Having an efficient and effective algorithm for classifying large data sets	Sensitivity to noise and training set and unrelated traits	Chen et al. (2017) Qasemian et al. (2018)
Bivariate Statistical Method	High flexibility A simple and applicable method for different regions High computing speed	Separate investigation of factors affecting the occurrence of landslides	Gholami and Ajalloeean (2017) Deljoee et al. (2016) Chimidi et al. (2017) Lin and Tung (2003)
Multivariate Statistical Method MR	High flexibility Assessing the sum of the factors influencing the occurrence of landslides together	The process of collecting and analyzing information is long and time consuming	Chimidi et al. (2017) Gholami and Ajalloeean (2017) Deljoee et al. (2016)
SVM	Relatively simple training High generality Generating the optimal answer	Time consuming calculations Is difficult to solve the highly nonlinear characteristics of landslide disasters	Fang et al. (2020) Huang and Zhao (2018) Yamani et al. (2012) Naemitabar and Zanganeh Asadi (2021)

Table 1 (continued)

Methods	Advantage	Disadvantage	Reference/s
Artificial Neural Network (ANN)	High accuracy in predicting landslide risk Independent of the statistical distribution of data Ability to handle data without depending on the measurement scale	Time consuming calculations The difficulty of network training Stuck at local maximums	Fang et al. (2020) Eker et al. (2015) Bui et al. (2020) Narimani (2016)
Fuzzy logic	Stable, simple and understandable High accuracy ability to set and validate Flexibility of the spatial analysis process Capability in modeling complex nonlinear problems with extra accuracy By utilizing fuzzy sets, continuous equations are created and the unknowns are reduced which is an important factor in enhancing incomplete relations	Time consuming implementation of fuzzy logic on common hardware Difficult in choosing membership functions and basic rules	Pradhan (2010) Muthu et al. (2008) Hejazi and Najafvand. (2020) Abdollahizad et al. (2021) Azarafza et al. (2018)
Index of entropy	Identifying effective factors and their degree of effectiveness	Limited problem solving Low performance for independent variables	Pourghasemi et al. (2012) Devkota et al. (2013) Jaafari et al. (2014) Mou et al. (2020)
Decision tree	Working with large and complex data Easy to combine with other methods High accuracy where it has good results with small-sized data in high-dimensional feature space	They grow exponentially as the problem gets bigger Accumulation of layer errors on each other Building a decision consumes a lot of memory	Pang et al.(2012) Brodley and Friedl (1997) Tazeh et al. (2016) Jamali (2021) Arabameri et al. (2020)
GD	Successful in finding a lower accurate convergence rate	Choosing the right learning rate can be difficult More time consuming than other algorithms	Hong et al. (2020)

basic function-SVM (0.7446), and linear-SVM (0.7390) models.

So far, different modeling techniques including Generalized Logistic Regression (GLR), Generalized Increment Model (GAM), WOE, SVM, RF, and Analyzed Autonomous Decision Tree (BPLDA) have been used. Evaluations showed that the BPLDA, RF, and WOE techniques resulted in more heterogeneous mapping and classifications and GLR, GAM, and SVM models led to more homogenous mapping. Many models and methods have been used to prepare the landslide susceptibility mapping. However, there is an urgent need to use methods and techniques to increase the predictive accuracy of landslides in a regional scale. Accordingly, it is important to note that the results of machine learning algorithms vary from one region to another due to differences in conditioning factors. Therefore, it is necessary to test machine learning algorithms in different regions, as well as to select and generalize the best model for each region based on their characteristics. Hence, we seek to determine the highest performance algorithm for identifying landslide-prone areas in the study area.

This study was conducted to introduce and evaluate the spatial predictions of landslides based on a comparison between the two functional-based algorithms and a Bayes-based algorithm in the southwestern part of Kurdistan province (Yozidar-Dagaga linking route). Among landslide studies, some researchers have claimed that SVM is a powerful and robust benchmark machine learning algorithm that often has a high goodness-of-fit and prediction accuracy than other machine learning algorithms (Kavzoglu et al. 2014; Nhu et al. 2020d). We believe that not only SVM can predict the landslide-susceptible areas with high potential but SGD has a high performance that can also be used as a soft-computing benchmark algorithm in landslide susceptibility assessment worldwide.

Study area and landslide distribution map

The study area is the linking route of Kamyaran city to Marivan city which is located in northwest of Kamyaran city and southeast of Sarvabad city. It is located between 46° 21′ 12″ E to 46° 43′ 52″ E and 35° 06′ 03″ N to 35° 11′ 34″ N. The route, with a length of about 35 km, passes through the villages of Yozidar, Palangan, Tafin, Dahakan, Surah Tu, Kani Hosseinbag, Jrîje, Saroumal, and Dagaga. The maximum and minimum elevations of the study area are 2982 m (south of the area) and 758 m (northwest), respectively, indicating a height difference of 2224 m (Fig. 1).

Kurdistan province is divided into two parts including east, southeast and central areas (as eastern part), west and southwest areas (as the western part) based on topographic, geomorphological, geological, and climate characteristics. Shallow and deep-seated landslides of west Kurdistan are caused by the Arabian–Iranian plate convergence, and also by its location in the Zagros fold and thrust belt with intense faulting and fracturing. Hence, preparing accurate landslide susceptibility maps for this region is of high importance for the management of landslide-prone areas.

Based on climatic stations' data inside and near the study area, the mean annual rainfall is 513 mm. Rainfall in the study area is controlled by the Atlantic Ocean, the Mediterranean Sea, the Black Sea, and to a lesser extent the northern cold systems, of which the two Atlantic and Mediterranean precipitation systems have the greatest effect on rainfalls of the study area. Most of the rainfall is snow in winter and rain in spring. The minimum rainfall is in summer, which is due to the dominance of subtropical high-pressure system. The average annual temperature in the region is about 16.6 °C. The climate of the region is semi-arid based on the De Martonne index climatic classification. The study area, as a part of the Sirvan drainage basin, is geologically located in the Sanandaj-Sirjan zone, with the exception of its southwest, which is located in the High Zagros zone. The geological units of the area are: Paleozoic massive limestone; Cream-colored biomicrosparite units (thick to bulk and alternating thin-layer limestone); shale units, lithic sandstone, crystal tuff and chert (which is due to the marine conditions after the Pyrenean orogenic phase) corresponding to the Late Cretaceous and Paleocene; Marl, sandstone and limestone units from the Oligo-Miocene; Flysch, sandstone, and conglomerate sediments belonging to the Cenozoic and the alluvial terraces of the Quaternary period. About 23% of the study area is underlain by the Cream-colored biomicrosparite units. In the thrust zone of Palangan, the Nagel series has been thrust over Ophiolite Mélange series. The trend of most of the faults of the region (NW–SE) is parallel to the Zagros orogenic elements. In terms of land use, the study area is mainly covered by rangelands, scattered oak forests, dry lands and gardens.

Overall, due to the topographical and geological conditions (existence of landslide-susceptible formations such as marl and shale), active faults of the area (Zagros Main Recent Fault), abundant precipitation in the form of snow, unstable slopes and numerous geomorphological processes, the study area is among the landslide-prone areas of the Kurdistan province and the country. Furthermore, anthropogenic factors have also contributed to the intensification of instability and mass movements (particularly in the unstable zones of the slope, caused by road construction on Kashtar to Yozidar route and the removal of slope bases).

Accurate determination of the location of landslides and the establishment of a spatial database are essential for future risk studies and assessment. The locations of landslides were identified by field surveys and checked on aerial photographs and satellite imageries. Analysis of these

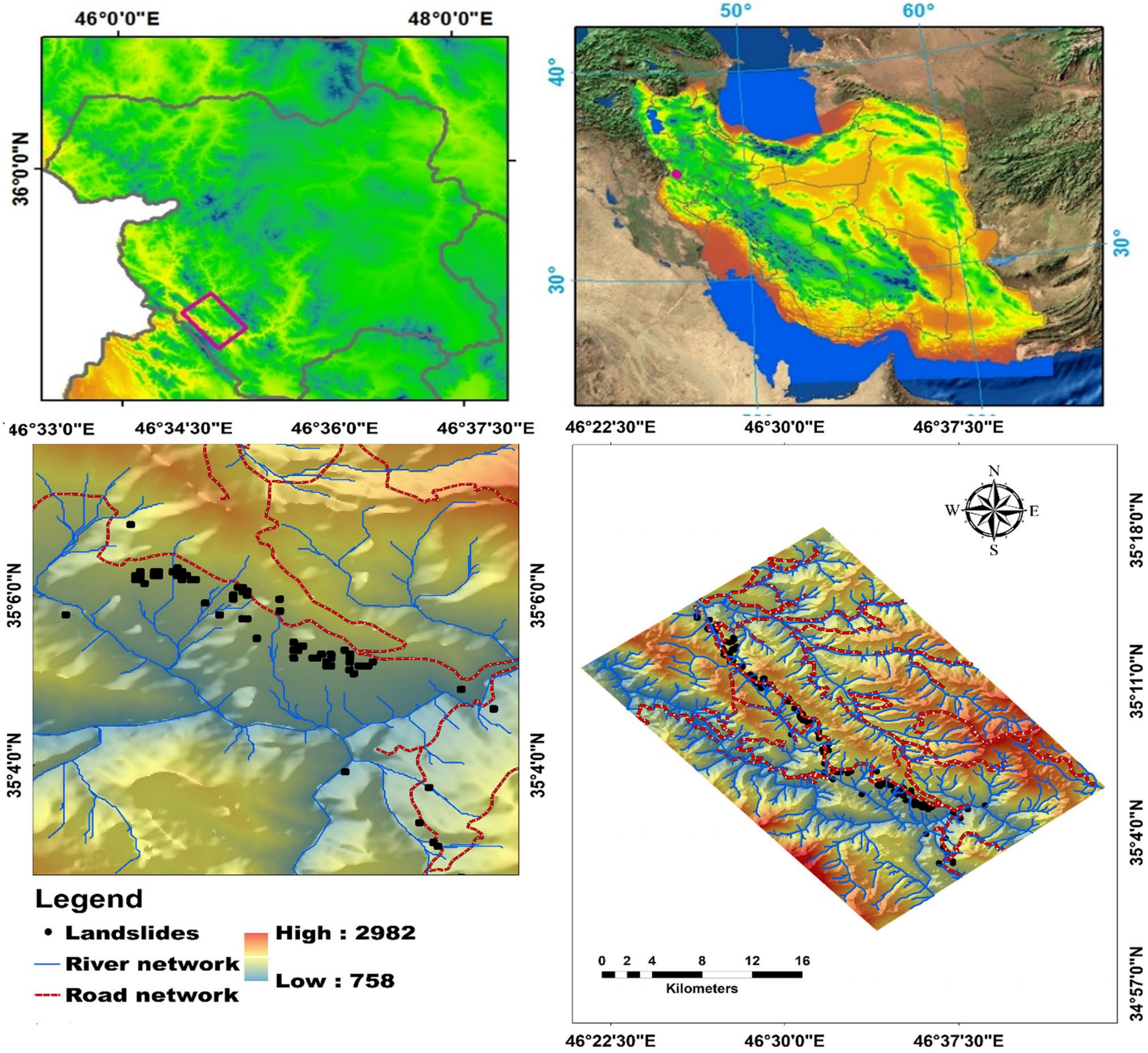


Fig. 1 Geographical position of the study area

sources showed that susceptible geological formations such as abundant calcareous and shale layers, incorrect construction of roads on slopes, changing land use especially in recent years as well as soil texture of the study area were the factors affecting the landslide occurrence. The field survey also indicated that most of landslides are rotational slides. A total of 175 landslides were identified, of which 123 were classified as training dataset and 52 as validation dataset. Along with the landslide dataset, 175 non-landslide locations were randomly selected in the places where landslides were not observed, especially on the flat areas as well as on slopes with hard lithology. Then, they were divided into the training and validation datasets similar to landslide datasets. Figure 2 shows some images of occurred landslides in the study area.

Methodology

Data preparation

Factors influencing mass movement process can be classified as: geological factors including weathered, sensitive or discontinuous material, contrast in permeability, composition of rock, soil forming slope material; geomorphological factors such as tectonic and volcanic uplift, fluvial or glacial

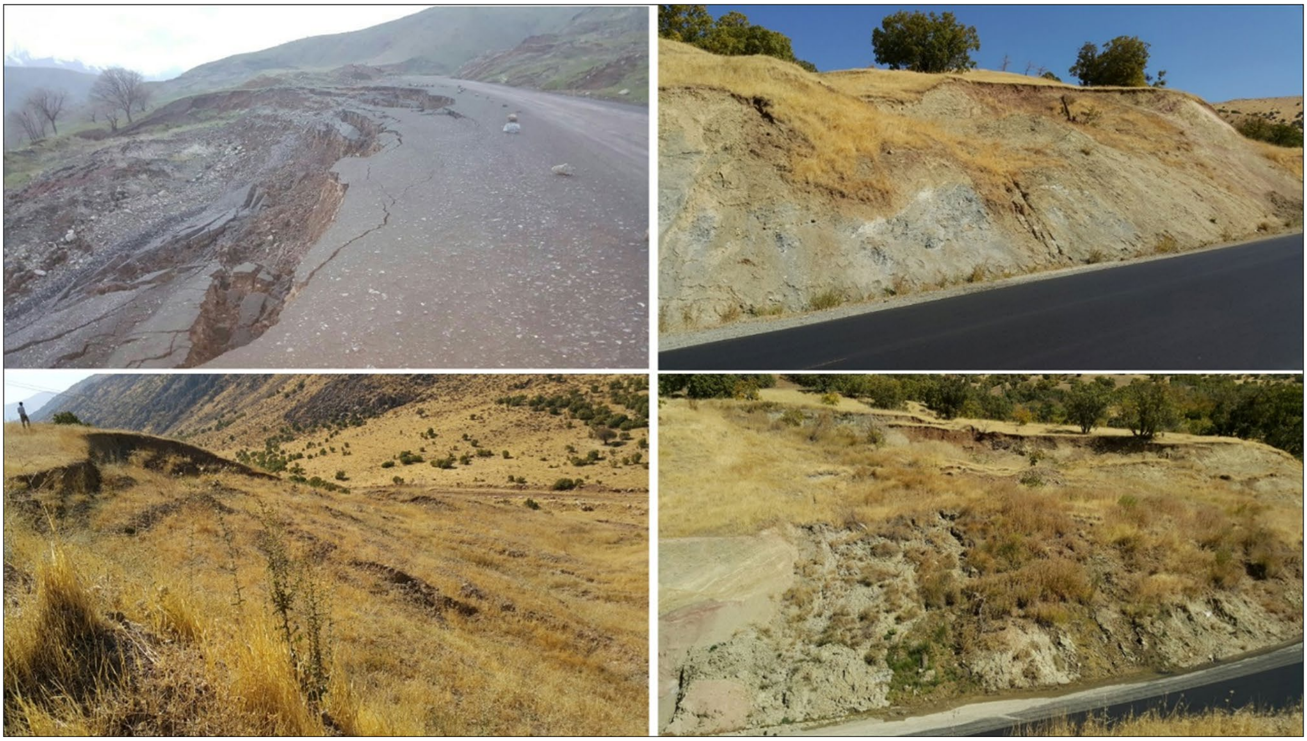


Fig. 2 The images show a number of landslides in the study area

erosion at toe of the slope or at lateral margins, subterranean erosion, topography or geometry of the slope, gradient of slope; climatic factors such as heavy rainfall, freeze and thaw cycles; and anthropogenic factors such as, excavation at toe of the slope, deforestation, irrigation, mining and land use (Pradhan et al. 2019).

At first, landslides were recorded based on field observations as well as checking them through aerial photographs at 1: 40,000 scale. Subsequently, their coordinates were identified by the global positioning system (GPS) and confirmed by the aerial photographs and satellite imageries. The first step in shallow landslide mapping in the study area is to convert vector layers into raster ones. Therefore, vector format layers were converted to the raster format ones with a resolution of 10 m by "resample" tool in ArcGIS 10.2 software. Then, all landslide locations were overlapped on the converted layers and the geodatabase was finally prepared for modeling by the WEKA 3.6.9 software. The 1:25,000 topographic maps, the 1:100,000 geological maps and the Landsat ETM + satellite imagery of 2018 were the main tools used in this study.

Twenty conditioning factors based on the literature review and data availability were identified (Fig. 3). Accordingly, the maps of slope angle (Fig. 3a), slope aspect (Fig. 3b), elevation (Fig. 3c), curvature (Fig. 3d), plan curvature (Fig. 3e), profile curvature (Fig. 3f), stream power index (SPI) (Fig. 3g), topographic wetness index (TWI) (Fig. 3h), slope length and steepness factor (LS) (Fig. 3i) were extracted from the digital elevation model (DEM) in ArcGIS 10.2. The land use map (Fig. 3j), and NDVI (Fig. 3k) were obtained from the Landsat ETM + satellite imagery. Lithology map (Fig. 3l), distance to faults (Fig. 3m), and fault density (Fig. 3n) were extracted from Kamyaran and Marivan geological maps at the scale of 1: 100,000. Rainfall map (Fig. 3o) was prepared based on the regression relationship between elevation and mean long-term annual rainfall of inside and outside rain-gauges in the study area. Maps of distance to rivers (Fig. 3p), river density (Fig. 3q), distance to road network (Fig. 3r), and road density (Fig. 3s) were provided, respectively, based on distance from the river and distance from the road network of the study area. To determine soil texture, 30 soil samples were taken from different lithological units of the study area. Then, the percentages of clay, silt and sand were obtained by hydrometer method. Subsequently, the texture of the soil samples was determined by the soil textural triangle (Fig. 3t).

The curvature expresses the topographic shape such that the positive curvature represents the surface where the pixels are convex, and the negative ones denotes the surface at which the pixels are concave. Its zero value shows the surface with no slope and is straight/flat (Ohlmacher 2007). Profile curvature is a form of slope, defined as the curvature of a flow line formed by the intersection of the earth's surface with a vertical plane (Shirzadi et al. 2017). Plan curvature is

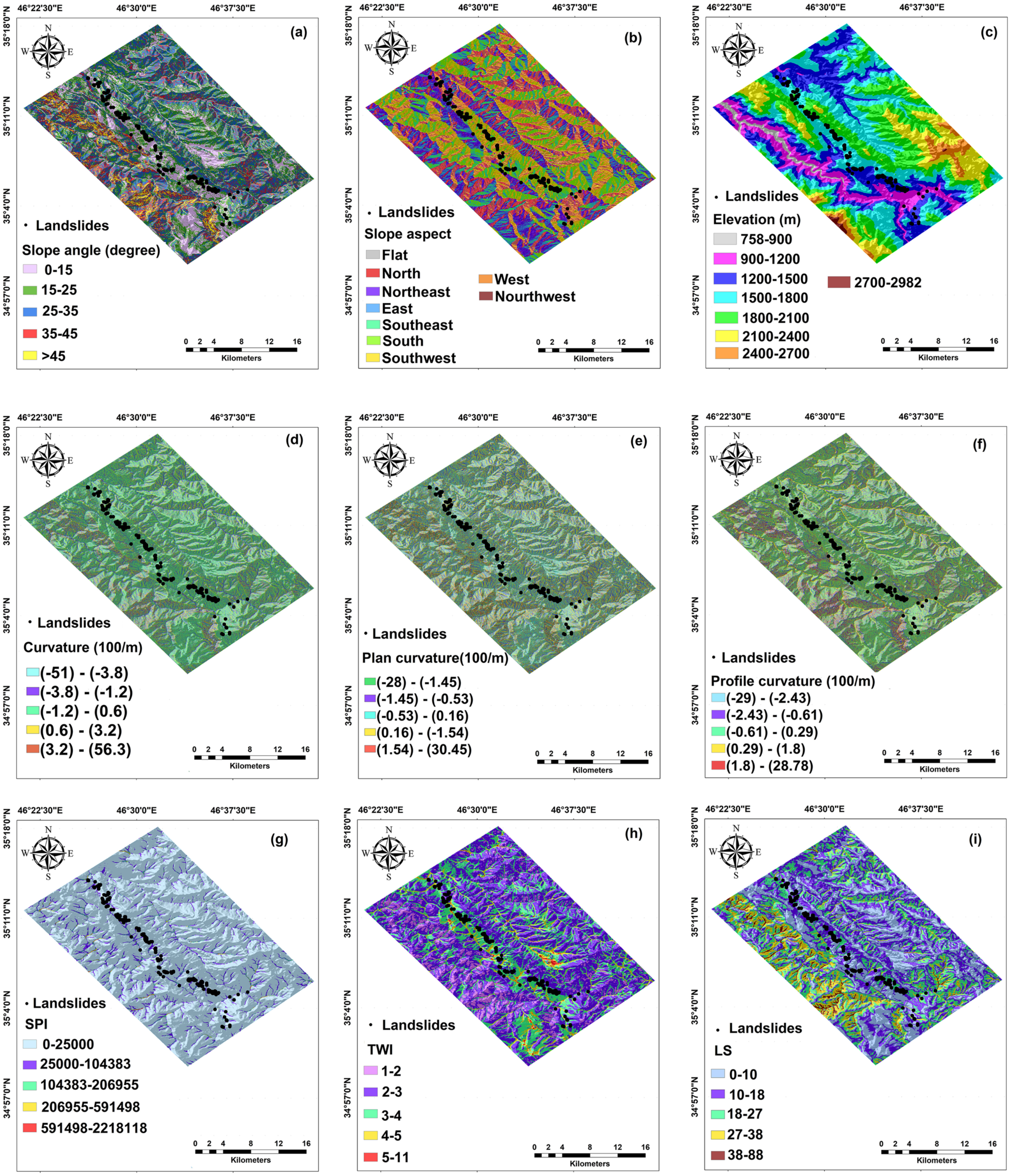


Fig. 3 Landslide conditioning factors used in this study: **a** slope angle, **b** slope aspect, **c** elevation, **d** curvature, **e** plan curvature, **f** profile curvature, **g** SPI, **h** TWI, **i** LS, **j** land use, **k** NDVI, **l** lithology, **m** distance to fault, **n** fault density, **o** rainfall, **p** distance to river, **q** river density, **r** distance to road, **s** road density, and **t** soil texture

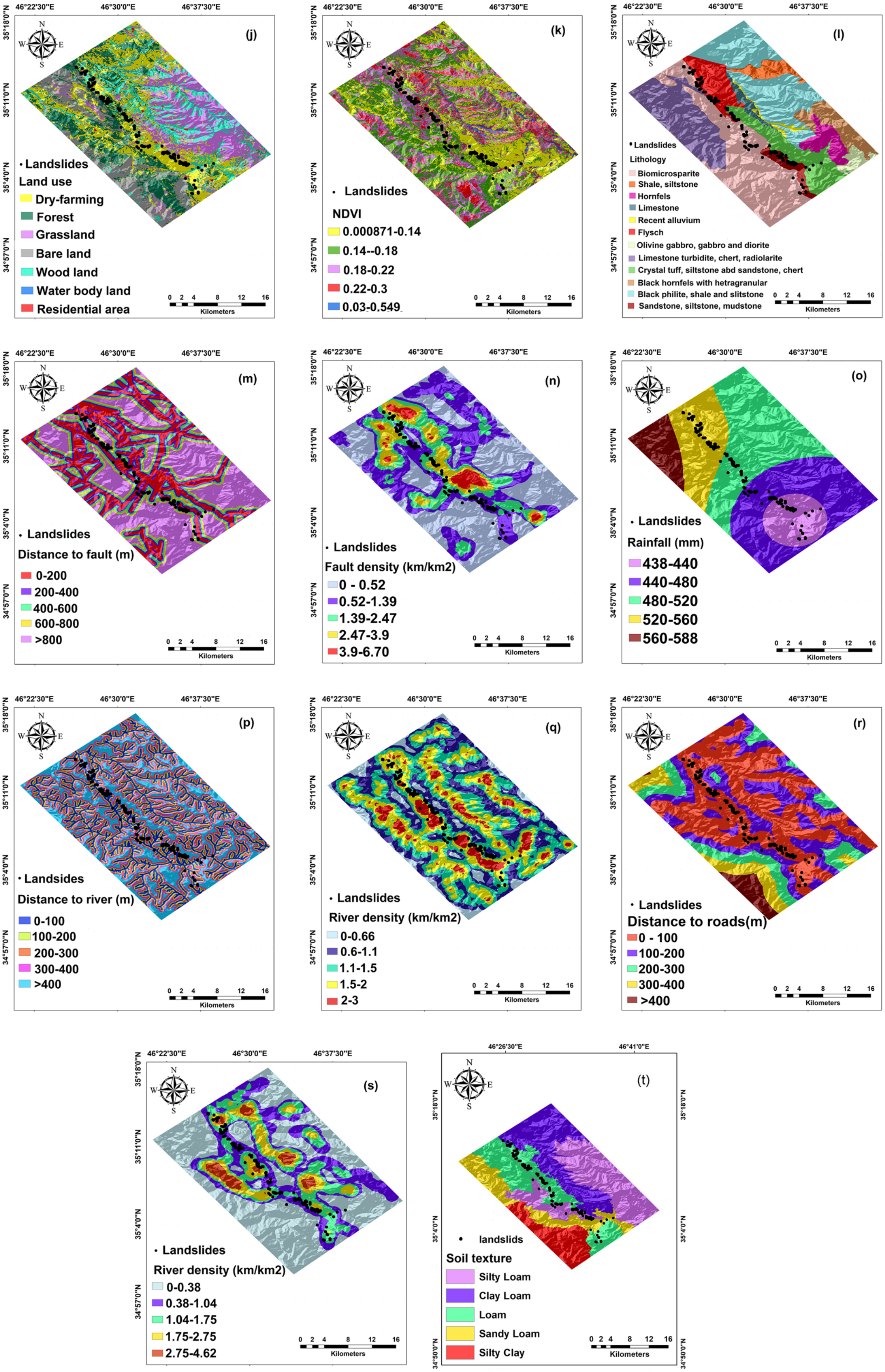


Fig. 3 (continued)

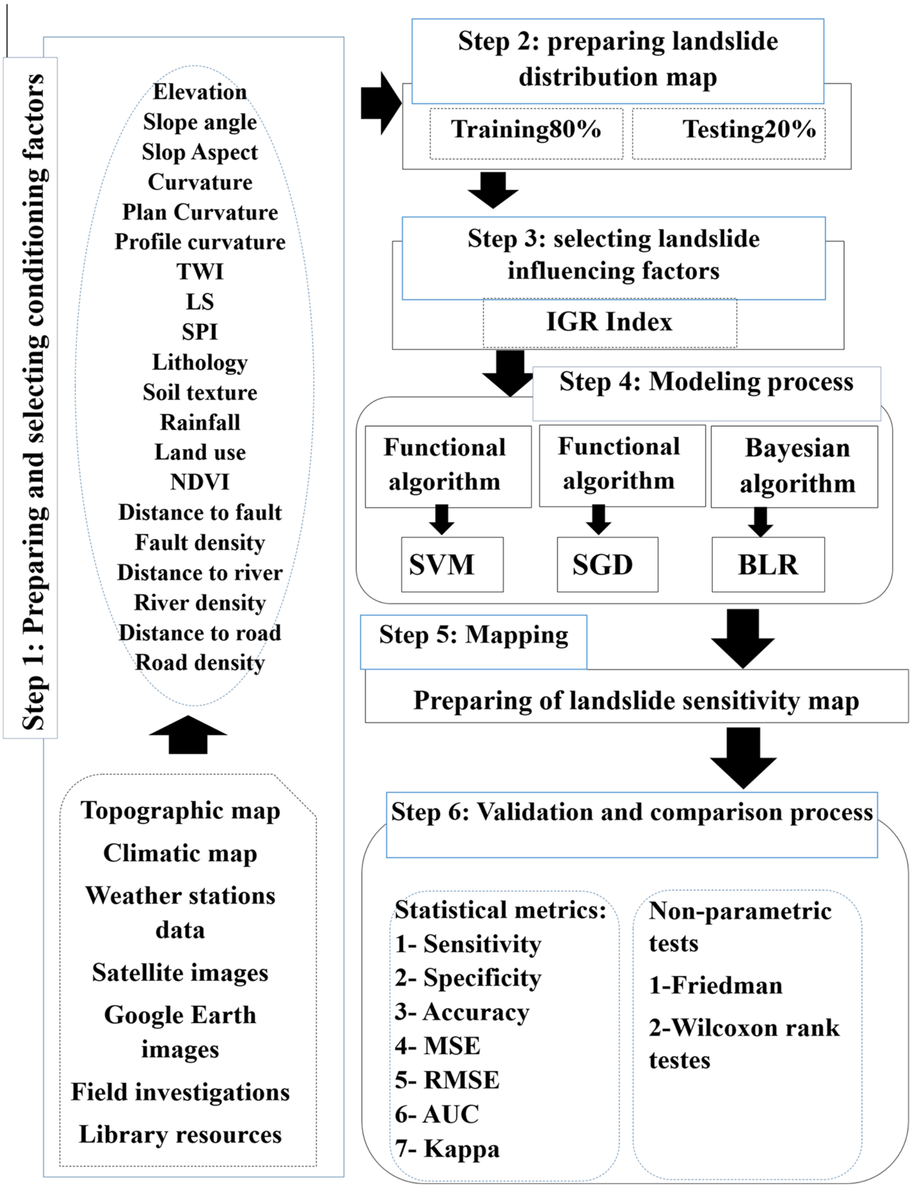


Fig. 4 Methodology flowchart of the research

defined as the line formed by the intersection of the earth's surface with a horizontal plane (Chen et al. 2018b). This factor influences the convergence and divergence of water and materials that form a landslide (Regmi et al. 2010). The combined slope length and steepness (LS-factor) is obtained from the mean LS value of the cells based on equation proposed by Moore and Wilson (1992) as follows:

$$LS = \left(\frac{A_s}{22.13}\right)^{0.4}\left(\frac{\sin b}{0.0896}\right)^{1.3}, \tag{1}$$

where A_s is the specific watershed area and b is the local slope angle in degrees. This index was developed based on the DEM in the SAGA software environment. The TWI is calculated as the ratio between the area of the specific watershed and the slope angle (Wilson and Gallant 2000). This index is calculated based on the following formula (Beven and Kirkby 1979):

$$TWI = \text{Ln}\left(\frac{\alpha}{\tan \beta}\right), \tag{2}$$

where α is cumulative upstream area of drainage at one point, and β is slope angle at a point of slope. This index indicates the spatial distribution of soil wetness or soil saturation.

The framework of methodology

The current study consists of the following steps (Fig. 4):

Step 1: *Preparing and selecting conditioning factors*. Based on data availability and literature review, the twenty conditioning factors were selected.

Step 2: *Preparing landslide distribution map*. Based on field survey, aerial photographs and satellite imageries, 175 landslide locations were recorded.

Step 3: *Selecting landslide conditioning factors*. Most effective factors for landslide modeling were selected based on information gain ratio technique.

Step 4: *Modeling the process*. In this step, machine learning models including SVM, SGD and BLR were used for landslide spatial prediction.

Step 5: *Preparing of landslide susceptibility map*. This step was conducted using output of machine learning models.

Step 6: *Validation and comparison process*. Susceptibility maps were validated based on statistical measures.

The most important factors: information gain ratio technique

There are several techniques for identifying the competence and predictability of variables affecting the occurrence of a phenomenon. One of the most important techniques is the Information Gain Ratio (IGR), which suggested by Quinlan (1993). The basis of IGR is information theory that, by reducing entropy, determines the importance of effective factors and is examined as a standard way to measure the ability of predicting factors affecting the occurrence of a data mining process (Bui et al. 2014). Higher values of IGR indicate its higher ability to be an effective measure in modeling. Therefore, the IGR was employed to identify the prominent factors affecting the occurrence of shallow landslides.

If S is training dataset with an input sample of $n\left(L_i.S\right)$ belonging to the L_i class (landslide, non-landslide). Then,

$$\text{info}(s) = -\sum_{i=1}^{2}\frac{n\left(L_i,S\right)}{|S|}\log_2\frac{n\left(L_i,S\right)}{|S|}. \tag{3}$$

The information required to divide S into series (S_1, S_2, $\cdots$, S_n) as estimated below:

$$\text{Info}(S,A) = \sum_{j=1}^{m}\frac{S_j}{|S|}\text{Info}(S). \tag{4}$$

The IGR index is calculated for a given factor such A *(e.g.,* slope angle) from the following equation:

$$\text{InformationGainRatio}(S,A) = \frac{\text{Info}(S) - \text{Info}(S,A)}{\text{SplitInfo}(S,A)}, \tag{5}$$

where Split Info denotes the information produced by the S split into the m subset calculated from the following equation:

$$\text{SplitInfo} = -\sum\nolimits_{i=1}^{m}\frac{S_j}{|S|}\log_2\frac{\left|S_j\right|}{S}. \tag{6}$$

Bayesian logistic regression

The Bayesian logistic regression (BLR) algorithm, first proposed by Friedman et al. (1997), is recognized as an effective way of presenting knowledge affected by uncertainty (Pearl 2014). Parameter estimates in BLR are probabilistic estimates rather than point estimates and therefore Bayesian algorithms provide alternatives to conventional methods that facilitate uncertainty estimation methods and show higher accuracy of parameter estimation (Mila et al. 2003; Das et al. 2012).

This algorithm is based on the Bayesian theory for graphical and probabilistic expression of the correlation between variables (Marcot et al. 2006). The Bayes-based theory algorithm is used extensively for modeling complex systems (Song et al. 2012).

This algorithm is a combination of a Bayes-based theory and a logistic regression function to obtain the weight of each example of the training dataset based on the relations between dependent and independent variables (Nhu et al. 2020d). In landslide modeling, first, a Bayesian function is constructed using a prior probability function based on the behavior and response of the conditioning factors (Nhu et al. 2020d) in three phases including, (i) identifying the prior probability of parameters, (ii) identifying the likelihood function for data, and (iii) applying a posterior distribution function for parameters. Then, a logistic regression function is used to calculate the weights of posterior probability function for samples belonging to a specific class of landslides and non-landslides, as follows:

$$P\left(\text{Class}|x_1,x_2,\ldots,x_n\right) = {}^{1}\Big/\left(1+\exp^{\left(a+w_0*b+\sum_{i=1}^{n}w_i*f(x_i)\right)}\right), \tag{7}$$

where x_i are the landslide conditioning factors of training dataset, x, b is the prior log odds ratio ($b = \log P$ (landslide class label/non - landslide class label), and a is the bias of the model. The weights that are trained by the training dataset are w_0 and w_i and ith factors of landslide conditioning factors are used to calculate the $f\left(x_i\right)$ function using the prior log odds ratio (b).

Support vector machine

The Support Vector Machine (SVM) is based on the principle of structural risk minimization and can be used to work with small sample datasets (Zhao and Zhou 2021). The SVM algorithm is based on the theory of statistical learning that the rate of the learning machine error for unclassified data can be considered as the generalized error rate (Xuegong 2000). These boundaries are a function of the set of training error rates that show the degree of complexity of the classifiers. There are two main ideas in SVM modeling to determine the type of statistical problems. The first is distinct optimal linear meta-schemes, which are discrete data patterns. The second idea is to use the core functions to transform the original non-linear data patterns into a form that is a distinct line in a high-dimensional space (Vapnik 1999). The explanatory details for SVM modeling in this study are as follows:

$x_i = (i = 1, 2, \ldots, n)$ is a landslide training dataset that consists of two classes (landslide (+ 1), and non-landslide (− 1)), and it is characterized by their maximum slot. This equation is mathematically expressed as follows:

$$\frac{1}{2} = \|W^2\|, \tag{9}$$

which is subject to the limitation of Eq. (10):

$$Yi = \left(\left(w.x_i\right) + b\right) \geq 1, \tag{10}$$

where $\|W^2\|$ is a rule of the normal meta-scheme with “1” being a numerical basis and “0” denoting the operation of numerical production, and its value is calculated using the method of Lagrange multipliers to define its function as follows:

$$L = \frac{1}{2}(W(^{2} - \sum_{i=1}^{n} \lambda_i(Y_i(\left(W.X_i\right) + b) - 1), \tag{11}$$

where λ_i is a Lagrange multiplier and can be zero or non-zero. Only datasets whose coefficients are non-zero are entered in the final equation and these datasets are known as support vectors (Schölkopf et al. 2000), and the core functions are used in the SVM model.

Stochastic gradient descent (SGD)

The SGD is mainly applied for solving large-scale learning issues with a high excellence performance (Wang et al. 2015). An arbitrary input *x* (conditioning factor) and a scalable output *y* (landslide and non-landslide) constitute a sample of *z (x, y)*. There is an *h (y-y)* function that measures the prediction cost of *y* when the real answer is *y*, and a function *f (x)* is chosen by a weight vector. Then, we look for the function *f*, which can minimize the coefficient $D(z.\theta) = h(f_\theta.(x).y)$.

$$R(f) = \int h\left(f_\theta(x), y\right) dp(Z); R_n(f) = \frac{1}{n}\sum_{i=1}^{n} h\left(f_\theta(x), y\right), \tag{12}$$

where *R(f)* measures the generalization efficiency and *Rn (f)* measures the efficiency of the training dataset. The gradient descent is an optimization algorithm to find the minimum of a function. In this algorithm, work begins with a random point on the function and moves in the negative direction of the function gradient to reach the local/global minimum. The SGD algorithm is an extreme simplification without the *Rn (f)* slope (Wang et al. 2015). The advantage of this algorithm is that it does not need the storage of gradients and, therefore, in complex problems of machine learning such as neural network learning or structured prediction is more easily applicable (Johnson and Zhang 2013).

Validation of landslide modeling

Statistical criteria

In this study, for evaluation and comparison of modeling results, the Percentage of Correct Predictions test was made on data. A 2×2 matrix was used to derive the criteria. This matrix consists of four possibilities including; false negative (FN), true negative (TN), false positive (FP), and true positive (TP). TP is the factor of ratio of number of pixels correctly divided as landslides, FN is the number of pixels with landslides (1) that are classified mistakenly as pixels without landslides (0), TN is the number of pixels without landslides (0) that are classified correctly as pixels without landslides (0), FP is the number of pixels without landslides (0) that are classified mistakenly as pixels with landslide (1) (Tsangaratos and Benardos 2014). Finally, the best result of these four states is when the TP value is high and the FP value is low (Althuwaynee et al. 2014). Sensitivity, is the ratio of landslide pixels that are correctly classified as landslides (Bui et al. 2016). This criterion indicates how good the predictive power of the landslide model is to classify landslide pixels (Pham et al. 2016). Specificity is the ratio of non-landslide pixels that are correctly classified as non-landslide (Pham et al. 2016). This criterion indicates how good the predictive power of the landslide model is to classify pixels of non-landslide (Pham et al. 2016). Accuracy refers to the ratio of occurrence and non-occurrence of landslides pixels that are correctly classified (Bennett et al. 2013). This criterion indicates how good the model performance is (Pham et al. 2016). Root Mean Square Error (RMSE) shows how much error is in the data (Bennett et al. 2013). The lower the RMSE, the better the landslide model performance (Pham et al. 2016). The kappa coefficient assesses the pairwise agreement or reliability between two or more measures (Carletta 1996). Mean absolute error (MAE) is an error which shows the difference between the paired observations that have widely been used in evaluating the accuracy of an algorithm (Pham et al. 2016). All the mentioned evaluation measures used in this study have been formulated as follows:

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

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

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

$$\text{Kappa} = \frac{P_{\text{c}} - P_{\text{exp}}}{1 - P_{\text{exp}}} \tag{16}$$

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|x_{\text{pred.}} - x_{\text{act.}}\right| \tag{17}$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(X_{\text{pred.}} - X_{\text{act.}})^2}{n}}, \tag{18}$$

where P_{c} is the proportion of observations in agreement and P_{exp} is the proportion in agreement due to chance, $x_{\text{pred.}}$ and $x_{\text{act.}}$ are the predicted and actual (output) values and n is the total samples.

Receiver operating characteristic (ROC) curve

An important strategy to provide meaningful interpretation of the results of predictive models, is outcome validation (Pourghasemi et al. 2013a, b). The ROC curve is a graphical curve that the “1-specificiy” denotes the *X*-axis and the *Y*-axis is defined by the “sensitivity”. The percentage of the area under the ROC curve (AUC) is a quantitative indicator to determine the overall performance of the models (Shirzadi et al. 2017). The larger the AUC is, the better the model performance will be. The range of this index varies from 0.5 (model with poor performance) to 1 (accurate performance of the model) (Bui et al. 2016).

The Friedman and Wilcoxon nonparametric tests

Friedman nonparametric test is also used to compare the performance of BLR, SVM and SGD methods. Nonparametric methods do not require any statistical assumptions (Derrac et al. 2011). The Friedman test can be applied as a nonparametric test even if the data are normally distributed (Martínez-Álvarez et al. 2013). In this test, it is first assumed that there is no difference between the performances of two models. After using the *p* value index (hypothesis probability), if the index is correct (< 5%), the hypothesis is rejected, and if the *p* value index is incorrect (> 5%), the hypothesis is confirmed. It should be noted that in comparing between the two or more models, if the *p* value in the Friedman test is true for all models (> 5%), the obtained results are not usable for comparing the models (Bui et al. 2015). To solve this problem, Wilcoxon nonparametric test is used to systematically investigate statistically significant differences between the two or more models.

Results and discussion

Determining the most important factors

The IGR technique was employed to identify the most important factors affecting the occurrence of shallow landslides in the study area. Figure 5 shows the ranking waterfall chart of the IGR index for 20 selected landslides conditioning factors in the study area. Accordingly, the highest values of the IGR index were allocated to distance to road, lithology, and road density, respectively. The factors of SPI, curvature, profile curvature, plan curvature, river density, distance to river, and LS, due to assigning zero value to this index, were excluded from the final modeling and modeling was performed with thirteen remaining factors.

Preparing shallow landslide susceptibility maps

According to the research methodology, shallow landslide susceptibility maps were prepared based on BLR, SVM, and SGD algorithms using quantile, natural breaks and geometrical interval methods in ArcGIS 10.2 environment. Finally, based on the landslide frequency histogram in each susceptibility class of these maps, the best method was selected. The results showed that natural breaks method was the best method and accordingly, all landslide susceptibility maps were classified into the five classes: very low susceptibility (VLS), low susceptibility (LS), moderate susceptibility (MS), high susceptibility (HS), and very high susceptibility (VHS). Figure 6 shows these maps for the BLR, SVM and SGD algorithms, respectively. The results of the SVM model showed that about 7.22% of the area was very susceptible to landslide; however, these rates in the BLR and SGD models were 20.61% and 18.77%, respectively.

Model validation and comparison

Table 2 shows the results of modeling evaluation using SVM, BLR, and SGD to check the goodness-of-fit/performance and prediction accuracy by the training and validation datasets, respectively. The results of goodness-of-fit or performance based on training dataset showed that the SGD and SVM (89%) algorithms had the highest sensitivity, followed by BLR (86%) algorithm. In terms of specificity, results indicated that the SVM with a value of 86% had

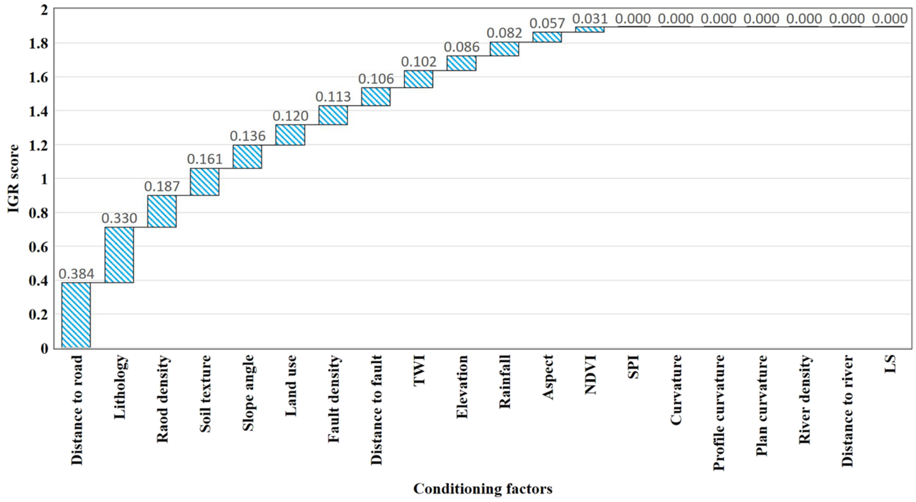


Fig. 5 The ranking waterfall chart for IGR values to show the importance of conditioning factors

higher performance than the SGD (84%) and BLR (81%) models. Moreover, the accuracy of the SVM was higher (87.8%) compared to the SGD (87%) and BLR (83.7%) algorithms. Moreover, the accuracy of the SVM was higher (90%) compared to the SGD (88%) and BLR (84%) algorithms. However, results of prediction accuracy of the algorithms by the validation dataset revealed that the SVM algorithm had higher prediction accuracy than the SGD and BLR algorithms. Overall, although the SGD and BLR models performed well, but the SVM model performed better.

Accuracy assessment of landslide susceptibility maps of the study area

Figure 7 shows the ROC curve based on the training (Fig. 7a) and validation (Fig. 7b) datasets. Results showed that the AUC value in the SVM method was 0.950, indicating that this method was capable of predicting landslide-susceptible areas, with a predictability of 95%, while the SGD and BLR methods had the predictability of 95.2% and 93.9%, respectively. However, for the validation dataset, the areas under the ROC curve in the SVM, SGD, and BLR algorithms were 0.920, 0.918 and 0.890, respectively. Although the results showed the excellent performance for all the three algorithms, the SVM algorithm had the highest ability in landslide classification and susceptibility mapping in the study area. In addition to the area under the ROC curve, landslide density index was also used to check the capability of landslide models in spatial predicting. Results pointed out that from VLS to HS classes, this index is added, indicated that the areas of high susceptibility had a higher incidence of landslides.

Figure 8 shows the landslide density in susceptibility classes of SVM, SGD, and BLR algorithms. Results illustrated that, in all the three algorithms, the landslide density was increased with increasing susceptibility to landslides, and hence obtained prediction accuracies by the algorithms were confirmed.

The Friedman and Wilcoxon nonparametric tests

The results obtained from the Friedman test are presented in Table 3. The average rankings of SVM, SGD, and BLR were 2.06, 3.19, and 3.39, respectively. Since the statistical significance (Sig.) was less than 5% (0.000) among all three models, the null hypothesis was rejected, which indicates that there is a significant difference between the algorithms. The results of the Wilcoxon signed-rank test are given in Table 4. These results showed that there was a significant difference between the algorithms at 5% level of statistical significance (p value < 0.05). The Wilcoxon signed-rank test was used to examine the statistical significance of the three landslide algorithms. In this test, there is a comparison between the two algorithms at 5% significance level. P value and z value are used to statistically evaluate the landslide susceptibility maps. The null hypothesis was rejected, as the P value was less than 0.05 and the z value exceeded the threshold value of z (-1.96 and $+1.96$), implied that the performance of the three models were significantly different.

The results of factor analysis using IGR showed that seven factors including SPI, curvature, plan curvature, profile curvature, river density, distance to river, and LS indices had zero values and no effect on landslides, and therefore, they were excluded from the final modeling. Moreover, distance to road, lithology and road density had the highest effects on the occurrence of landslides. This might due to the presence of landslide-susceptible formations such as marl and shale, along with improper human policies such as road construction. Improper cutting of the heel of the slopes during the construction of roads causes more opportunity for water to penetrate into the sensitive soil formations and by saturating these soils under the force of gravity on the slopes (slope

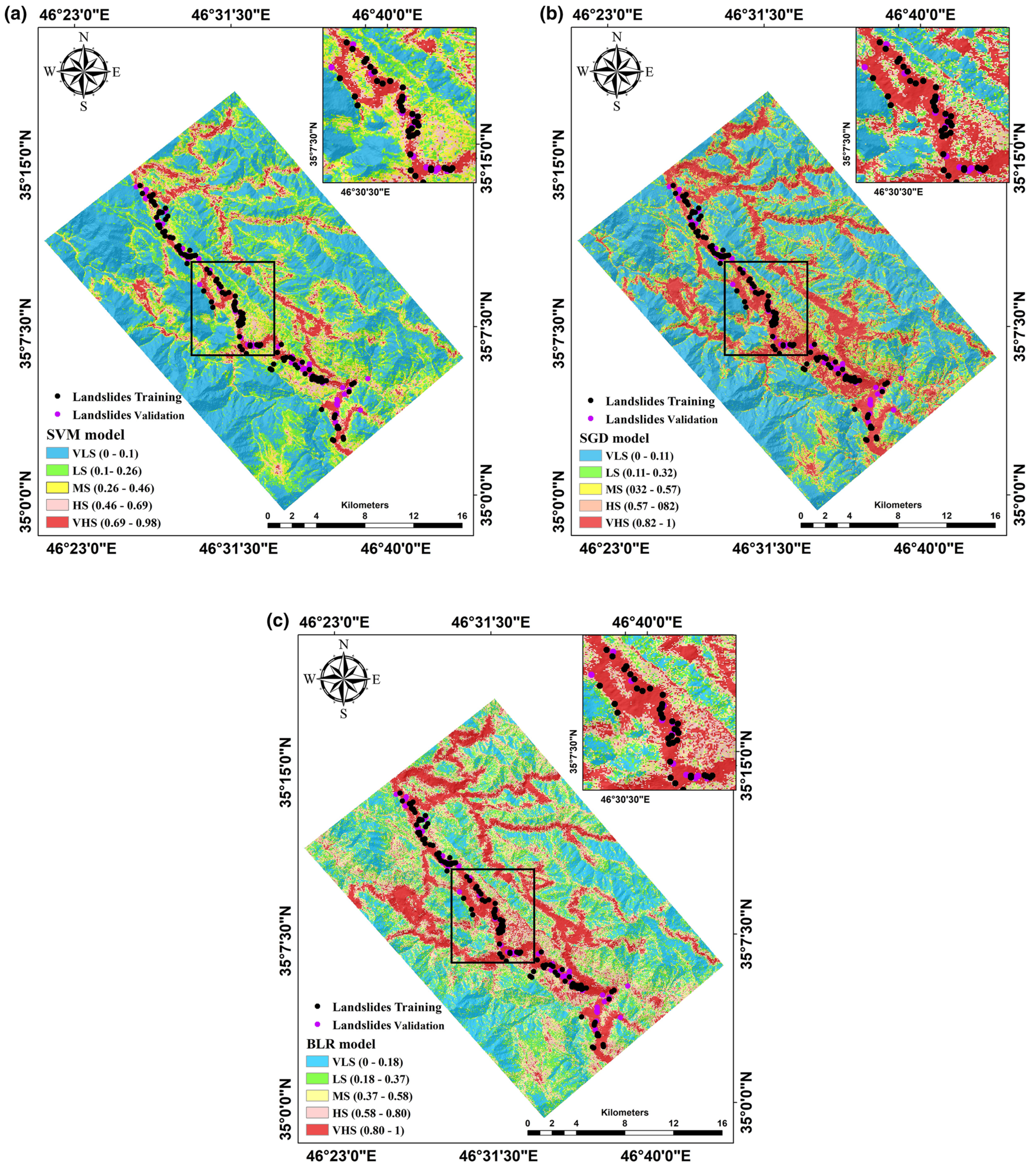


Fig. 6 Landslide susceptibility maps in the study area **a** SVM, **b** SGD, and **c** BLR

factor) these landslides occur in the study area. Saturation of soil under the gravity force on the slopes could facilitate the landslide occurrence in the study area. This is consistent with the results of Pham et al. (2017) and Pham et al. (2016) in terms of greater impact of the road factor than other factors affecting the occurrence of landslides. Landslide susceptibility assessment is one of the most important issues in recent decades, due to the identification of susceptible areas that can be used in decision-making related to land use planning and landslide hazard assessment. Different methods of

Table 2 Performance and prediction accuracy of the algorithms by training and validation datasets

Index	Training dataset			Validation dataset		
	SVM	SGD	BLR	SVM	SGD	BLR
True positive (TP)	110	110	106	46	46	42
True negative (TN)	106	104	100	44	44	45
False positive (FP)	17	19	23	8	8	7
False negative (FN)	13	13	17	6	6	10
Sensitivity (%)	0.894	0.894	0.862	0.885	0.885	0.808
Specificity (%)	0.862	0.846	0.813	0.846	0.846	0.865
Accuracy (%)	0.878	0.870	0.837	0.865	0.865	0.837
Kappa	0.756	0.739	0.675	0.731	0.730	0.675
MAE	0.206	0.207	0.162	0.207	0.134	0.164
RMSE	0.320	0.322	0.403	0.322	0.367	0.404

landslide susceptibility mapping have been suggested by different researchers. However, the prediction accuracy of these methods is still controversial around the world.

Observation of the LSMs based on BLR, SVM and SGD algorithms showed that most landslides occurred in areas with steep slopes. Also, areas with steep slopes and rocky material are less sensitive to landslides, and finally low slopes, which have limited areas due to the mountainous nature of the region, have a much lower sensitivity to landslides. Accordingly, areas with high landslide sensitivity coincide with the middle sections of slopes cut by road construction. Also, the existence of lithological units susceptible to landslides (i.e., marl and shale) in the middle parts of the slopes played an important role in landslide incidence in these areas. The presence of landslide-sensitive formations in this part, including marl and shale, plays the role of rupture surface for saturated topsoil and causes the upper layers to slip. The upper and the higher altitudes of the area are less susceptible to landslides due to the presence of crystalline and basaltic units. In the case of land vegetation, landslides mostly occurred in dry lands that were formerly semi-dense and grassland forests, indicating the impact of forest degradation and land use change on landslides.

The results of this study showed that in addition to the above parameters, geomorphological forms and processes also played an important role in the occurrence of landslides in the study area. High drainage densities, especially in erodible formations such as marl and shale, have led to the creation of river channels with steep walls, resulted in landslide incidence. Also, the presence of many fractures and faults with different aspects played an important role in the infiltration of water and intensification of weathering of rocks and sediments and has finally accelerated the occurrence of landslides.

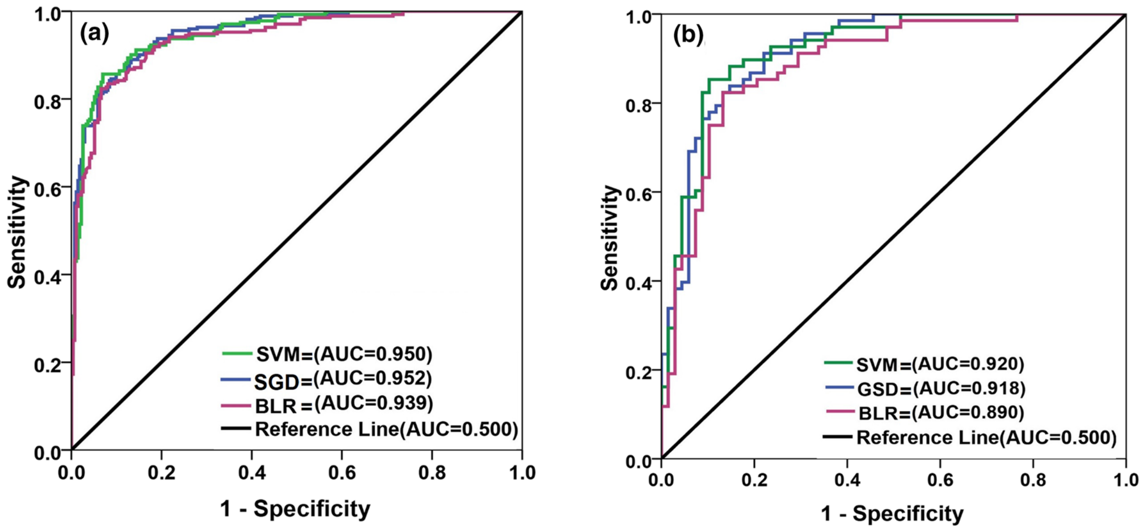


Fig. 7 ROC curve for training **a** and validation **b** datasets

Table 3 Mean ranks of landslide susceptibility algorithms using Friedman test

Algorithm	Mean ranks	Chi-square	Sig
SVM	2.06	900.868	0.000
SGD	3.19		
BLR	3.39		

Table 4 Pairwise comparison of the three algorithms using Wilcoxon signed-rank test

Pairwise comparison	NPD	NND	*Z* value	*P* value	Sig
SVM. Vs. SGD	431	113	− 18.071	0.000	Yes
SVM. Vs. BLR	538	6	− 20.151	0.000	Yes
SGD. Vs. BLR	243	301	−4.925	0.000	Yes

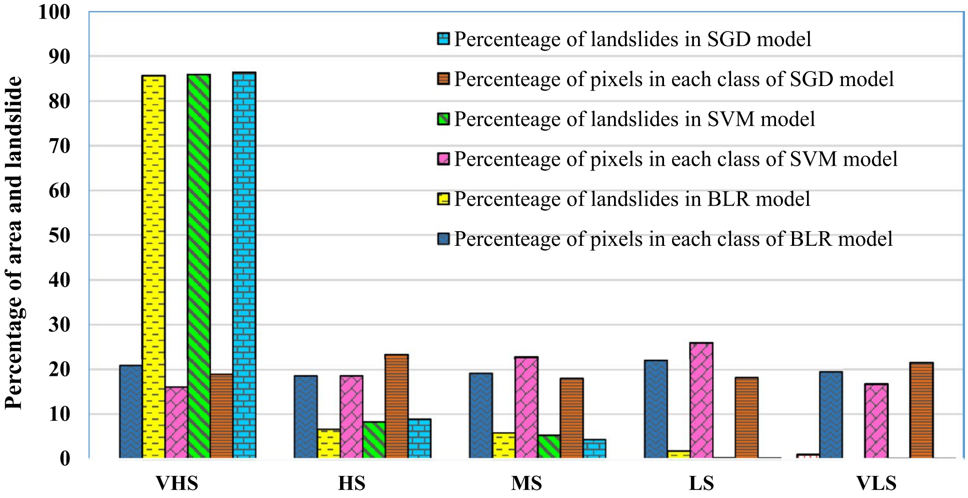
NPD Number of positive differences; *NND* Number of negative differences; *Sig.* Statistical significance

Kappa, TP, specificity, sensitivity, accuracy, and chi-squared statistical measures were used to evaluate the models for both the training and the validation datasets. Finally, the performance of these three models was evaluated through the AUC. Model validation results indicated that SVM model with sub-curve level of 0.920 showed better performance than SGD (AUC = 0.918) and BLR (AUC = 0.890) with minor differences. Nevertheless, the results showed that the performance and prediction accuracy of all three algorithms were validated and confirmed. Also, the performance of these three models was evaluated by the Friedman and Wilcoxon tests and it was found that there was no significant difference at 95% level between the results of these three models. Eventually, the results of the three models can be trusted to identify areas susceptible to landslide incidence in the study area.

Conclusions

In this study, the performance of BLR, SVM and SGD algorithms in order to map landslide susceptibility in Yuzidar-Degaga rout in Kurdistan province was compared. A total of 175 landslides were identified and 20 conditioning factors were employed. Based on the IGR method which was used to show the order and importance of the conditioning factors on landslide occurrence, curvature, plan curvature, profile curvature, SPI, LS, distance to river and river density factors were removed from the final modeling process, because they had not positive roles on landslide occurrence, whereas distance to roads and lithology were the most important factors in landslide modeling. The values obtained from all three algorithms, both in the training and validation datasets, indicated that all three models were confirmed in terms of accuracy and modeling but the SVM model had the highest capability to predict landslide incidence. Finally, the study area was classified into five susceptibility zone. One of the reasons for the success of SVM model based on comparison of results was the strong theoretical assumptions associated with nonlinear algorithm and the ability to obtain parameter values, which made this model superior to others. Also, the results revealed that landslide density increased from VLS classes to VHS. This implied that areas of high susceptibility had a higher incidence of landslides, and the obtained maps corresponded well with areas where landslides had occurred. Given the high effect of the roads in this model, it is suggested that the priority of landslide prevention and control measures should be paid attention to reduce the effect of road construction in the study area. In addition, if the road development operation is planned in the future, this operation must be carried out in strict compliance with the principles of road construction and the stability of the slopes. The proposed landslide sensitivity map can be useful for selecting appropriate management measures and decisions in land use planning, identifying hazard points and

Fig. 8 Bar graphs showing landslide densities in susceptibility classes of SVM, SGD and BLR algorithms



preventing damage related to landslide risk. Also, to minimize the effects of landslides, these results can be used in early warning system strategies, in addition to slope stability models in the Zagros region. The results can be used to assess landslide risk in areas with similar environments and to help improve landslide susceptibility maps.

Acknowledgements The authors would like to thank the University of Kurdistan, Sanandaj, Iran for supplying required data, reports, useful maps, and their nationwide geodatabase for the first author (Mitra Asadi) as Ph.D. student during her passing internal sabbatical under the guidance of Dr. Himan Shahabi in this university. The authors also greatly appreciate the assistance of anonymous reviewers for their constructive comments that helped us to improve the paper.

References

Abdollahizad S, Balafar MA, Feizizadeh B et al (2021) Using hybrid artificial intelligence approach based on a neuro-fuzzy system and evolutionary algorithms for modeling landslide susceptibility in East Azerbaijan Province. Iran. Earth Sci Inform 14:1861–1882. https://doi.org/10.1007/s12145-021-00644-z

Akgün A, Türk N (2011) Mapping erosion susceptibility by a multivariate statistical method: a case study from the Ayvalık region, NW Turkey. Comput Geosci 37:1515–1524

Althuwaynee OF, Pradhan B, Park H-J, Lee JH (2014) A novel ensemble decision tree-based CHi-squared automatic interaction detection (CHAID) and multivariate logistic regression models in landslide susceptibility mapping. Landslides 11:1063–1078

Amir Ahmadi A, Kamrani Dalir H, Sadeghi M (2010) Landslide risk zoning using Analytic Hierarchy Process (AHP): Case study of Chalav Amol watershed Geography, 8(27):181–203. https://www.sid.ir/fa/journal/ViewPaper.aspx?id=118670. Accessed 5 Nov 2021

Anbalagan R, Kumar R, Lakshmanan K, Parida S, Neethu S (2015) Landslide hazard zonation mapping using frequency ratio and fuzzy logic approach, a case study of Lachung Valley, Sikkim. Geoenviron Disasters 2:1–17

Arabameri A, Saha S, Roy J, Chen W, Blaschke T, Bui T (2020) Landslide susceptibility evaluation and management using different machine learning methods in the Gallicash river watershed, Iran. Remote Sens 12(3):475. https://doi.org/10.3390/rs12030475

Arjmandzadeh R, Sharifi Teshnizi E, Rastegarnia A et al (2019) GIS-based landslide susceptibility mapping in Qazvin Province of Iran. Iran J Sci Technol Trans Civ Eng 44:619–647. https://doi.org/10.1007/s40996-019-00326-3

Atash Afrooz N, Safaeipour M (2021) Landslide micro-zoning using Demetel and fuzzy AHP techniques (Case study: Dehdez section of Khuzestan province). https://civilica.com/doc/1250893. Accessed 11 Nov 2021

Ayalew L, Yamagishi H (2005) The application of GIS-based logistic regression for landslide susceptibility mapping in the Kakuda-Yahiko Mountains, Central Japan. Geomorphology 65:15–31

Azarafza M, Ghazifard A, Akgün H et al (2018) Landslide susceptibility assessment of South Pars Special Zone, southwest Iran. Environ Earth Sci 77:805. https://doi.org/10.1007/s12665-018-7978-1

Bathrellos GD, Gaki-Papanastassiou K, Skilodimou HD, Papanastassiou D, Chousianitis KG (2012) Potential suitability for urban planning and industry development using natural hazard maps and geological–geomorphological parameters. Environ Earth Sci 66:537–548

Bennett G, Molnar P, McArdell B, Schlunegger F, Burlando P (2013) Patterns and controls of sediment production, transfer and yield in the Illgraben. Geomorphology 188:68–82

Beven KJ, Kirkby MJ (1979) A physically based, variable contributing area model of basin hydrology/Un modèle à base physique de zone d'appel variable de l'hydrologie du bassin versant. Hydrol Sci J 24:43–69

Brodley CE, Friedl MA (1997) Decision tree classification of land cover from remotely sensed data. Remote Sens Environ 61:399–409

Bui DT, Pradhan B, Lofman O, Revhaug I, Dick OB (2012) Spatial prediction of landslide hazards in Hoa Binh province (Vietnam): a comparative assessment of the efficacy of evidential belief functions and fuzzy logic models. CATENA 96:28–40

Bui DT, Pradhan B, Revhaug I, Tran CT (2014) A comparative assessment between the application of fuzzy unordered rules induction algorithm and J48 decision tree models in spatial prediction of shallow landslides at Lang Son City, Vietnam. In: Srivastava P, Mukherjee S, Gupta M, Islam T (eds) Remote sensing applications in environmental research. Springer, pp 87–111

Bui DT, Pradhan B, Revhaug I, Nguyen DB, Pham HVQN (2015) A novel hybrid evidential belief function-based fuzzy logic model in spatial prediction of rainfall-induced shallow landslides in the Lang Son city area (Vietnam) Geomatics. Nat Hazards Risk 6:243–271

Bui DT, Tuan TA, Klempe H, Pradhan B, Revhaug I (2016) Spatial prediction models for shallow landslide hazards: a comparative assessment of the efficacy of support vector machines, artificial neural networks, kernel logistic regression, and logistic model tree. Landslides 13:361–378

Bui DT, Tsangaratos P, Nguyen VT, Liem NV (2020) Comparing the prediction performance of a Deep Learning Neural Network model with conventional machine learning models in landslide susceptibility assessment. CATENA 188:104426

Carletta J (1996) Assessing agreement on classification tasks: the kappa statistic. Comput Linguist 22(2):249–254

Chen W, Xie X, Wang J, Pradhan B, Hong H, Bui DT, Duan Z, 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

Chen W, Pourghasemi HR, Naghibi SA (2018a) A comparative study of landslide susceptibility maps produced using support vector machine with different kernel functions and entropy data mining models in China. Bull Eng Geol Env 77:647–664

Chen W, Shahabi H, Shirzadi A, Li T, Guo C, Hong H, Li W, Pan D, Hui J, Ma M (2018b) A novel ensemble approach of bivariate statistical-based logistic model tree classifier for landslide susceptibility assessment. Geocarto Int 33:1398–1420

Chimidi G, Raghuvanshi TK, Suryabhagavan KV (2017) Landslide hazard evaluation and zonation in and around Gimbi town, western Ethiopia—a GIS-based statistical approach. Appl Geomat (springer) 9(4):219–236

Das I, Stein A, Kerle N, Dadhwal VK (2012) Landslide susceptibility mapping along road corridors in the Indian Himalayas using Bayesian logistic regression models. Geomorphology 179:116–125

Deljooe A, Hossini S, Sadeghi S (2016) Evaluation of different landslide risk zoning methods in forest ecosystems. Ext Dev Watershed Manag 4(13):7–14

Derrac J, García S, Molina D, Herrera F (2011) A practical tutorial on the use of nonparametric statistical tests as a methodology for comparing evolutionary and swarm intelligence algorithms. Swarm Evol Comput 1:3–18

Devkota KC, RegmiA D, Pourghasemi HR, Yoshida K, Pradhan B, Ryu IC, Althuwaynee OF (2013) Landslide susceptibility mapping using certainty factor, index of entropy and logistic regression models in GIS and their comparison at Mugling-Narayanghat road section in Nepal Himalaya. Nat Hazards 65(1):135–165

Dong J-J, Tung Y-H, Chen C-C, Liao J-J, Pan Y-W (2009) Discriminant analysis of the geomorphic characteristics and stability of landslide dams. Geomorphology 110:162–171

Eker AM, Dikmen M, Cambazoğlu S, Düzgün ŞH, Akgün H (2015) Evaluation and comparison of landslide susceptibility mapping methods: a case study for the Ulus District, Bartın, Northern Turkey. Int J Geogr Inf Sci 29(1):132–158

Fang Z, Wang Y, Peng L, Hong H (2020) A comparative study of heterogeneous ensemble-learning techniques for landslide susceptibility mapping. Int J Geogr Inf Sci 35:321–347

Farhadinejad T, Souri S, Lashkaripour Gh, Ghafouri M(2011)Landslide Hazard Zoning in the National Basin (Nojian) Modified by Mora-Warson and Nielsen Method, 6th National Congress of Civil Engineering, Semnan

Farrokhnia A, Pirasteh S, Pradhan B, Pourkermani M, Arian M (2011) A recent scenario of mass wasting and its impact on the transportation in Alborz Mountains, Iran using geo-information technology. Arab J Geosci 4:1337–1349

Felicísimo ÁM, Cuartero A, Remondo J, Quirós E (2013) Mapping landslide susceptibility with logistic regression, multiple adaptive regression splines, classification and regression trees, and maximum entropy methods: a comparative study. Landslides 10:175–189

Friedman N, Geiger D, Goldszmidt M (1997) Bayesian network classifiers. Mach Learn 29:131–163

Gholami M, Ajalloeean R (2017) Comparison of experimental selective methods and statistical methods and artificial neural network for landslide hazard zoning (case study in Beheshtabad Dam Reservoir). J Amirkabir Civ Eng 49:363–437

Goetz J, Brenning A, Petschko H, Leopold P (2015) Evaluating machine learning and statistical prediction techniques for landslide susceptibility modeling. Comput Geosci 81:1–11

Guha-Sapir D, Below R, Hoyois P (2020) EM-DAT: international disaster database. Brussels, Belgium: Université Catholique de Louvain. Available from: http://www.emdat.be. Accessed 3 Mar 2020

Hejazi SA, Najafvand S (2020) Potential assessment of landslide prone areas in Paveh city using Fuzzy logic method. Geogr Hum Relat 2:8

Hong H, Chen W, Xu C, Youssef AM, Pradhan B, Tien Bui D (2017) Rainfall-induced landslide susceptibility assessment at the Chongren area (China) using frequency ratio, certainty factor, and index of entropy. Geocarto Int 32:139–154

Hong H, Tsangaratos P, Ilia I, Loupasakis C, Wang Y (2020) Introducing a novel multi-layer perceptron network based on stochastic gradient descent optimized by a meta-heuristic algorithm for landslide susceptibility mapping. Sci Total Environ 742:140549

Huang Y, Zhao L (2018) Review on landslide susceptibility mapping using support vector machines. CATENA 165:520–529

Jaafari A, Najafi A, Pourghasemi HR, Rezaeian J, Sattarian A (2014) GIS based frequency ratio and index of entropy models for landslide susceptibility assessment in the Caspian forest, northern Iran. Int J Environ Sci Technol 11:909–926

Jaafari A, Panahi M, Pham BT, Shahabi H, Bui DT, Rezaie F, Lee S (2019) Meta optimization of an adaptive neuro-fuzzy inference system with grey wolf optimizer and biogeography-based optimization algorithms for spatial prediction of landslide susceptibility. CATENA 175:430–445

Jamali A (2021) Landslide hazard risk modeling in north-west of Iran using optimized machine learning models. Model Earth Syst Environ. https://doi.org/10.1007/s40808-020-00871-1

Johnson R, Zhang T (2013) Accelerating stochastic gradient descent using predictive variance reduction. NIPS Proc Int Conf Neural Info Process Syst 1:315–323

Kavzoglu T, Sahin EK, Colkesen I (2014) Landslide susceptibility mapping using GIS-based multi-criteria decision analysis, support vector machines, and logistic regression. Landslides 11:425–439

Kayastha P, Dhital MR, De Smedt F (2012) Landslide susceptibility mapping using the weight of evidence method in the Tinau watershed, Nepal. Nat Hazards 63:479–498

Khezri S, Rustaei Sh, Rajaei Asl A (2006) Assessment and zoning of slope instability risk in the central part of Zab basin (Sardasht city) by Anbalagan method. Lecturer of Humanities, 10 (48 consecutive) special issue of Geography), pp 49–80. https://www.sid.ir/fa/journal/ViewPaper.aspx?id=71065. Accessed 12 Oct 2020

Lee S, Choi J, Min K (2002) Landslide susceptibility analysis and verification using the Bayesian probability model. Environ Geol 43:120–131

Lee S, Hong S-M, Jung H-S (2017) A support vector machine for landslide susceptibility mapping in Gangwon Province, Korea. Sustainability 9:48

Liao K, Wu Y, Miao F, Li L, Xue Y (2020) Using a kernel extreme learning machine with grey wolf optimization to predict the displacement of step-like landslide. Bull Eng Geol Env 79:673–685

Lin M-L, Tung C-C (2003) A GIS-based potential analysis of the landslides induced by the Chi-Chi earthquake. Eng Geol 71(1–2):63–77

Mansoori M, Shirani K (2016) Landslide risk zoning by entropy methods and control weight : Case study Doab Samsami area of Chaharmahal and Bakhtiari province. Earth Sci 26(102):267–280. https://www.sid.ir/fa/journal/ViewPaper.aspx?id=299961. Accessed 23 Sept 2021

Marcot BG, Steventon JD, Sutherland GD, McCann RK (2006) Guidelines for developing and updating Bayesian belief networks applied to ecological modeling and conservation. Can J for Res 36:3063–3074

Marjanović M, Kovačević M, Bajat B, Voženílek V (2011) Landslide susceptibility assessment using SVM machine learning algorithm. Eng Geol 123:225–234

Martínez-Álvarez F, Reyes J, Morales-Esteban A, Rubio-Escudero C (2013) Determining the best set of seismicity indicators to predict earthquakes. Two case studies: Chile and the Iberian Peninsula. Knowl-Based Syst 50:198–210

Mila AL, Yang XB, Carriquiry AL (2003) Bayesian logistic regression of Soyabean Sclerotinia stem rot prevalence in the U.S. north-central region: accounting for uncertainty in parameter estimation. Phytopathology 93:758–763

Moore ID, Wilson JP (1992) Length-slope factors for the revised universal soil loss equation: simplified method of estimation. J Soil Water Conserv 47:423–428

Mou N, Wang C, Yang T, Zhang L (2020) Evaluation of development potential of ports in the Yangtze river delta using FAHP-entropy model. Sustainability 12:1–24

Muthu K, Petrou M, Tarantino C, Blonda P (2008) Landslide possibility mapping using fuzzy approaches. IEEE Trans Geosci Remote Sens 46:1253–1265

Naemitabar M, Zanganeh Asadi M (2021) Landslide zonation and assessment of Farizi watershed in northeastern Iran using data mining techniques. Nat Hazards 108:2423–2453. https://doi.org/10.1007/s11069-021-04805-7

Narimani S (2016) Evaluation of artificial intelligence model and multi criteria decision modeling in landslide risk mapping (case study: Idoghmush Chai Basin), Master's thesis, University of Tabriz, Tabriz, Iran

Nhu V-H, Mohammadi A, Shahabi H, Ahmad BB, Al-Ansari N, Shirzadi A, Clague JJ, Jaafari A, Chen W, Nguyen H (2020a) Landslide susceptibility mapping using machine learning algorithms and remote sensing data in a tropical environment. Int J Environ Res Public Health 17:4933

Nhu V-H, Mohammadi A, Shahabi H, Ahmad BB, Al-Ansari N, Shirzadi A, Geertsema M, Kress VR, Karimzadeh S, Valizadeh Kamran K (2020b) Landslide detection and susceptibility modeling on Cameron highlands (Malaysia): a comparison between random forest, logistic regression and logistic model tree algorithms. Forests 11:830

Nhu V-H, Shirzadi A, Shahabi H, Chen W, Clague JJ, Geertsema M, Jaafari A, Avand M, Miraki S, Talebpour Asl D (2020c) Shallow landslide susceptibility mapping by random forest base classifier and its ensembles in a semi-arid region of Iran. Forests 11:421

Nhu V-H, Zandi D, Shahabi H, Chapi K, Shirzadi A, Al-Ansari N, Singh SK, Dou J, Nguyen H (2020d) Comparison of support vector machine, Bayesian logistic regression, and alternating decision tree algorithms for shallow landslide susceptibility mapping along a mountainous road in the west of Iran. Appl Sci 10:5047

Nsengiyumva JB, Luo G, Nahayo L, Huang X, Cai P (2018) Landslide susceptibility assessment using spatial multi-criteria evaluation model in Rwanda. Int J Environ Res Public Health 15:243

OFDA/CRED (2018) International Disaster Database. Brussels: Université Catholique de Louvain. www.emdat.be. Accessed 9 Aug 2018

Ohlmacher GC (2007) Plan curvature and landslide probability in regions dominated by earth flows and earth slides. Eng Geol 91:117–134

Pang P. K, Tien L. T, Lateh H (2012) Landslide hazard mapping of penang island using decision tree model, in Proceedings of the International Conference on Systems and Electronic Engineering (ICSEE '12), Phuket, Thailand, December.

Pearl J (2014) Probabilistic reasoning in intelligent systems: networks of plausible inference. Elsevier

Pham BT, Bui DT, Prakash I, Dholakia M (2016) Rotation forest fuzzy rule-based classifier ensemble for spatial prediction of landslides using GIS. Nat Hazards 83:97–127

Pham BT, Bui DT, Pourghasemi HR, Indra P, Dholakia M (2017) Landslide susceptibility assesssment in the Uttarakhand area (India) using GIS: a comparison study of prediction capability of naïve bayes, multilayer perceptron neural networks, and functional trees methods. Theoret Appl Climatol 128:255–273

Pham BT, Prakash I, Bui DT (2018) Spatial prediction of landslides using a hybrid machine learning approach based on random subspace and classification and regression trees. Geomorphology 303:256–270

Pham BT, Prakash I, Singh SK, Shirzadi A, Shahabi H, Bui DT (2019) Landslide susceptibility modeling using Reduced Error Pruning Trees and different ensemble techniques: hybrid machine learning approaches. CATENA 175:203–218

Pourghasemi HR, Kerle N (2016) Random forests and evidential belief function-based landslide susceptibility assessment in Western Mazandaran Province, Iran. Environ Earth Sci 75:185

Pourghasemi HR, Pradhan B, Gokceoglu C, Moezzi KD (2012) Landslide susceptibility mapping using a spatial multi criteria evaluation model at Haraz Watershed, Iran. Terrigenous mass movements. Springer, Berlin, Heidelberg, pp 23–49

Pourghasemi HR, Pradhan B, Gokceoglu C, Mohammadi M, Moradi HR (2013a) Application of weights-of-evidence and certainty factor models and their comparison in landslide susceptibility mapping at Haraz watershed, Iran. Arab J Geosci 6(7):2351–2365

Pourghasemi H, Moradi H, Aghda SF (2013b) Landslide susceptibility mapping by binary logistic regression, analytical hierarchy process, and statistical index models and assessment of their performances. Nat Hazards 69:749–779

Pourtaghi ZS, Pourghasemi HR (2014) GIS-based groundwater spring potential assessment and mapping in the Birjand Township, southern Khorasan Province, Iran. Hydrogeol J 22:643–662

Pradhan B (2010) Manifestation of an advanced fuzzy logic model coupled with geoinformation techniques for landslide susceptibility analysis. Environ Ecol Stat 18:471–493

Pradhan B, Lee S (2010) Landslide susceptibility assessment and factor effect analysis: backpropagation artificial neural networks and their comparison with frequency ratio and bivariate logistic regression modelling. Environ Model Softw 25:747–759

Pradhan SP, Vishal V, Singh TN (eds) (2019) Landslides: theory, practice and modelling. Springer, p 50

Qasemian B, Abedini M, Rustaei Sh, Shirzadi A (2018) Comparative study of vector support machine models and tree logistics to evaluate landslide sensitivity, Case study: Kamyaran city, Kurdistan province. Nat Geogr 11(1(39 consecutive)):47–68

Quinlan J (1993) Programs for machine learning (Morgan Kaufmann series in machine learning). Morgan Kaufmann, p 302

Razavizadeh S, Solaimani K, Massironi M et al (2017) Mapping landslide susceptibility with frequency ratio, statistical index, and weights of evidence models: a case study in northern Iran. Environ Earth Sci 76:499. https://doi.org/10.1007/s12665-017-6839-7

Regmi NR, Giardino JR, Vitek JD (2010) Assessing susceptibility to landslides: using models to understand observed changes in slopes. Geomorphology 122:25–38

Rozos D, Bathrellos G, Skillodimou H (2011) Comparison of the implementation of rock engineering system and analytic hierarchy process methods, upon landslide susceptibility mapping, using GIS: a case study from the Eastern Achaia County of Peloponnesus, Greece. Environ Earth Sci 63:49–63

Schilirò L, Montrasio L, Mugnozza GS (2016) Prediction of shallow landslide occurrence: validation of a physically-based approach through a real case study. Sci Total Environ 569:134–144

Schölkopf B, Smola AJ, Williamson RC, Bartlett PL (2000) New support vector algorithms. Neural Comput 12:1207–1245

Shadman Roodposhti M, Aryal J, Shahabi H, Safarrad T (2016) Fuzzy shannon entropy: a hybrid gis-based landslide susceptibility mapping method. Entropy 18:343

Shirzadi A, Bui DT, Pham BT, Solaimani K, Chapi K, Kavian A, Shahabi H, Revhaug I (2017) Shallow landslide susceptibility assessment using a novel hybrid intelligence approach. Environ Earth Sci 76:60

Skilodimou HD, Bathrellos GD, Chousianitis K, Youssef AM, Pradhan B (2019) Multi-hazard assessment modeling via multi-criteria analysis and GIS: a case study. Environ Earth Sci 78:47

Song Y, Gong J, Gao S, Wang D, Cui T, Li Y, Wei B (2012) Susceptibility assessment of earthquake-induced landslides using Bayesian network: a case study in Beichuan, China. Comput Geosci 42:189–199

Tamene L, Abegaz A, Aynekulu E, Woldearegay K, Vlek PL (2011) Estimating sediment yield risk of reservoirs in northern Ethiopia using expert knowledge and semi-quantitative approaches. Lakes Reserv Res Manag 16:293–305

Tazeh M, Taghizadeh Mehrjerdi R, Fathabadi A, Kalantari S (2016) Model of landslide hazard zonation and its effective factors using quantitative geomorphology (Case Study: Sanich region, Yazd). Environ Erosion Res 6:15–1

Tien Bui D, Shahabi H, Omidvar E, Shirzadi A, Geertsema M, Clague JJ, Khosravi K, Pradhan B, Pham BT, Chapi K (2019a) Shallow landslide prediction using a novel hybrid functional machine learning algorithm. Remote Sens 11:931

Tien Bui D, Shirzadi A, Shahabi H, Geertsema M, Omidvar E, Clague JJ, Thai Pham B, Dou J, Talebpour Asl D, Bin Ahmad B (2019b)

New ensemble models for shallow landslide susceptibility modeling in a semi-arid watershed. Forests 10:743

Tiranti D, Cremononi D (2019) Editorial: landslide hazard in a changing environment. Front Earth Sci. https://doi.org/10.3389/feart.2019.00003

Tsangaratos P, Benardos A (2014) Estimating landslide susceptibility through a artificial neural network classifier. Nat Hazards 74:1489–1516

Tsangaratos P, Ilia I (2016) Comparison of a logistic regression and Naïve Bayes classifier in landslide susceptibility assessments: the influence of models complexity and training dataset size. CATENA 145:164–179

Turner AK, Shuster R L (1996) Landslide; investigation and mitigation, Special report (National Research Council (U.S) Transportation Research Board, Ch.9: 199-209

Vapnik V (1999) The nature of statistical learning theory. Springer, New York

Varnes DJ (1958) Landslide types and processes. Landslides Eng Pract 24:20–47

Wang Yt, Seijmonsbergen AC, Bouten Wt, Chen Q (2015) Using statistical learning algorithms in regional landslide susceptibility zonation with limited landslide field data. J Mt Sci 12:268–288

Wang G, Lei X, Chen W, Shahabi H, Shirzadi A (2020) Hybrid computational intelligence methods for landslide susceptibility mapping. Symmetry 12:325

Wilson JP, Gallant JC (2000) Terrain analysis: principles and applications. John Wiley & Sons

Wu YP, Chen L, Cheng C, Yin KL, Török Á (2014) GIS-based landslide hazard predicting system and its realtime test during a typhoon, Zhejiang Province, Southeast China. Eng Geol 175:9–21

Xu C, Xu X, Dai F, Xiao J, Tan XXuC, Xu X, Dai F, Xiao J, Tan X, Yuan R (2012) Landslide hazard mapping using GIS and weight of evidence model in Qingshui river watershed of 2008 Wenchuan earthquake struck region. J Earth Sci 23:97–120

Xuegong Z (2000) Introduction to statistical learning theory and support vector machines. Acta Automatica Sinica 26:32–42

Yalcin A (2008) GIS-based landslide susceptibility mapping using analytical hierarchy process and bivariate statistics in Ardesen (Turkey): comparisons of results and confirmations. CATENA 72:1–12

Yamani M, Ahmadabadi A, Zare Gh (2012) Application of vector support machine algorithm in landslide risk zoning : Case study Darkeh catchment. Geography Environ Hazards 1(3):125–142. https://www.sid.ir/fa/journal/ViewPaper.aspx?id=189677. Accessed 4 Apr 2019

Zhang G, Cai Y, Zheng Z, Zhen J, Liu Y, Huang K (2016) Integration of the statistical index method and the analytic hierarchy process technique for the assessment of landslide susceptibility in Huizhou, China. CATENA 142:233–244

Zhao Sh, Zhou Z (2021) A comparative study of landslide susceptibility mapping using SVM and PSO-SVM models based on grid and slope units. Math Probl Eng 2021:1–15

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