Landslide susceptibility assessment and mapping using statistical and data mining models in Iran

Abdulaziz Hanifinia
Urmia University

Habib Nazamejad
Urmia University

Saeed Najafi
Urmia University

Aiding Kornejady
Gorgan University of Agricultural Sciences and Natural Resources

Hamid Reza Pourghasemi (✉ hr.pourghasemi@shirazu.ac.ir)
Shiraz University https://orcid.org/0000-0003-2328-2998

Research Article

Keywords: Landslides, GLM, MaxEnt, ANN, SVM

DOI: https://doi.org/10.21203/rs.3.rs-239985/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Abstract

This study applied to evaluate landslide susceptibility using four data mining models including, “Generalized Linear Model (GLM)”, “Maximum Entropy (ME)”, “Artificial Neural Network (ANN)”, and “Support Vector Machine (SVM)” in Cherikabad Watershed in Urmia City, Iran. In particular, Shannon entropy was used to assess the intercomparison of factors’ classes. Eleven factors including, elevation, slope angle, slope aspect, geological formation, annual mean rainfall, land use/land cover, distance to the village, distance to faults, distance to roads, distance to streams, and NDVI used in the current study. Landslide inventory map was identified using Google Earth imagery, extensive field surveys, and scrutinizing archived data. The produced landslide susceptibility maps were evaluated by the AUROC index. The results of performance metrics revealed that the Shannon entropy with an AUROC of 0.879 proved highly reliable and so is the intercomparison analysis of factors’ classes derived from it. Additionally, the goodness-of-fit of the GLM, ME, ANN, and SVM models were 0.763, 0.740, 0.926, and 0.924, while their predictive powers were 0.751, 0.727, 0.917, and 0.935, respectively. Hence, the results indicated that the SVM model can be introduced as the superior model for the study area based on which the most critical factors affecting landslides were found to be elevation, annual mean rainfall, and distance to the village. The results of this work are of great use for land use planning in landslide-prone areas with similar geo-topological, geomorphological, and climatic conditions.

Introduction

Landslides are complex, isolated, and destructive natural hazards by which human lives are threatened, and considerable socioeconomic and environmental losses are incurred (Palmisano et al., 2016; Zhou et al., 2018; Ambrosi et al., 2018; Billah et al., 2019). Landslides occur owing to the interactive effect of different geo-environmental factors including landform (Kim et al., 2015; Liucci et al., 2017; Chudý et al., 2019), geological evolution (Agostini et al., 2014; Niculită et al., 2019; Song et al., 2020), groundwater table oscillation (Peng et al., 2018; Cao et al., 2020), land uses/land covers (Persichillo et al., 2017, Hao et al., 2020), precipitation (Chang et al., 2017; Segoni et al., 2018, Stanley et al., 2020, Lin et al., 2020), excessive irrigation (Hou et al., 2018; Cummins, 2019; Lacroix et al., 2020), seismic activities (Wang et al., 2018; Ojala et al., 2019; Hibert et al., 2019; Shafique, 2020), anthropogenic interferences (Mohammadi and Taiebat, 2016; Singh and Singh, 2020; Wu et al., 2020), and climate changes (Alvioli et al., 2018; Peres and Cancelliere, 2018; Rianna et al., 2019; Niculiță, 2020).

Iran is considered a landslide-prone country due to favorable geographical conditions in which many landslides have occurred in recent years (Moayedi et al., 2020). Evaluation of landslide susceptibility, and hazard with a spatial connotation of landslide occurrence probability has been considered as a practical solution and a pivotal part of the landslide risk mitigation (Hong et al. 2019). In particular, the classification of susceptible areas through landslide spatial modeling (LSM) is an essential step in environmental hazard assessment in watershed management frameworks (Sarkar et al., 1995). Also, landslide sensitivity assessment is a logical approach and a new solution to reduce and minimize landslide damages (Fell et al., 2008).

Several studies have been devoted to map landslide susceptibility and hazard using different statistical algorithms including, “frequency ratio” (Singh and Kumar, 2017; Chen et al., 2017a, c; Sharma and Mahajan, 2019; Khan et al., 2019), “weights of evidence” (Wang et al., 2016; Chen et al., 2016b; Xie et al., 2017; Wang et al., 2019), “logistic regression” (Tsangaratos and Ilia, 2015; Hadjii et al., 2017; Achour et al., 2018; Abedini et al., 2018), “bivariate statistical analysis” (Wang et al., 2015; Youssef et al., 2015; Mersha and Meten, 2020; Ozioko and Igwe, 2020), “multivariate logistic regression” (Lee, 2007; Pradhan, 2010; Pourghasemi and Rahmati, 2018), “analytical hierarchy process-AHP” (Demir et al., 2013; Chen et al., 2016a; Ali et al., 2019; Arabameri et al., 2019a), and “discriminant
analysis" (He et al., 2012; Arabameri et al., 2019b). In recent years, various data mining techniques developed and adopted to assess landslide susceptibility and hazard such as ANNs (Tien Bui et al., 2016; Chen et al., 2017b; Zhu et al., 2018; Soma et al., 2019), fuzzy logic (Mandal and Mondal, 2019; Razifard et al., 2019), “neuro-fuzzy” (Lee et al., 2015; Jaafari and Pourghasemi, 2019; Panahi et al., 2020), “Decision Trees-DTs” (Alkhasawneh et al., 2014; Chen et al., 2019a), “random forest-RF” (Youssef et al., 2016; Chen et al., 2018c; Jaafari and Pourghasemi, 2019; Dang et al., 2020; Shirvani, 2020), SVM (Pourghasemi et al., 2013b; Hong et al., 2015; Chen et al., 2018a,b; Shirzadi et al., 2018; Ada and San, 2018, Huang and Zhao, 2018), “classification and regression tree-CART” (Pham et al., 2018a; Choubin et al., 2019), “boosted regression tree-BRT” (Arabameri et al., 2019b; Park and Kim, 2019), ME (Chen et al., 2017b) and some comparative modeling attempts (Liu and Duane, 2018; Zhang et al., 2018; Chen et al., 2019b,c; Moayedi et al., 2020; ).

The aim of generating a landslide spatial map is to understand the most susceptible areas as well as to predict future landslide-prone areas (Chen et al., 2019c). During the past three decades, the development of novel modeling techniques has been unceasingly progressing, all of which aim at producing a more precise and generalizable landslide susceptibility mapping scheme (Kornejady et al., 2017; Broeckx et al., 2018; Reichenbach et al., 2018). Among different landslide susceptibility assessment and zoning models, data mining models use a wide range of data to discover patterns and predict the occurrence of the phenomenon of interest in such a way that enables them to provide better results compared to their statistical counterparts (Zaki et al., 2014; Zheng, 2015). In light of these premises, we set the objectives of this work as follows: four data mining models, namely GLM, ME, ANN, and SVM), as well as a bivariate statistical model known as Shannon entropy-SE, were used to map landslide susceptibility across the Cherikabad (NW of Iran) and the derived susceptibility maps per each model were analyzed using one cutoff-independent metric termed as the AUCROC index.

Study area
The Cherikabad Watershed is situated in the Northwest of Iran and extends for an area of 51.8 km$^2$ (Fig. 1). Elevation of the Cherikabad Watershed vary from 1,628 to 3,473 m.a.s.l. The average slope is about 18%. The average annual rainfall of this watershed is 430 mm. The maximum and minimum of temperature, respectively, amounts to 13 °C and -7 °C (Mohammadnejad and Asghari, 2016). An arid or semi-arid climate has prevailed upon the study area. Diverse lithological settings exist in the region mainly consist of the ophiolite, metamorphic and meta-sedimentary rocks, limestone, dolomites, shale, and conglomerates, most of which belong to the Late Cretaceous from the Mesozoic, while plains mostly consist of Quaternary sediments.

Methods And Materials
The methodological flowchart adopted in this study is presented in Fig. 2.

Data compilation
Landslide inventory (LI) is known as the first and the most pivotal step for modeling susceptibility and hazard of landslides and, generally, any modeling effort with a spatial connotation (Regmi et al., 2014). In this study, the LI was carried out utilizing a handheld GPS device during extensive field surveys, Google Earth images, and acquiring archived organizational data. In general, 95 landslide locations were recorded and mapped in ArcGIS, which were split into two sets of training (60%) and validation (40%) (Teimouri and Kornejady 2019). The second step regards the selection of the optimal combination of landslide-conditioning factors, which include predisposing and triggering factors (Pourghasemi and Rossi 2017). Drawing on the literature review and scrutinizing our observations during
field inspections, eleven factors were selected, namely elevation, annual mean rainfall, slope aspect, slope degree, distance to streams/village/roads/faults, lithological formation, land use/land cover, and NDVI (Fig. 3). Table 1 show the original sources of the acquired data, their spatial resolution, and causative roles are provided in detail in Table1.

**Statistical model**

In recent decades, different statistical techniques texted to generate landslide susceptibility and hazard maps in divers’ areas. Statistical methods can quantitatively examine the relationship between different variables and landslide locations (Guzzetti et al., 1999; Carrara et al., 1999). Statistical methods have gained due attention from researchers who are in pursuit of assessment of the reciprocal relationship between the landslide locations and effective factors (Aleotti and Chowdhury, 1999). According to the general concusses, statistical methods are divided into two categories (reference): bivariate and multivariate statistical methods. Bivariate statistical methods examine the relationships between each controlling factor and landslide occurrence in an area regardless of the effect of other factors, whereas multivariate statistical methods consider the interaction among different factors and between each factor and landslide occurrence (Wang et al., 2005). Bivariate and multivariate statistical methods have different types. In this study, a bivariate statistical method, namely, Shannon entropy, was used.

**Shannon entropy model**

The concept of entropy has been used in different scientific branches with different meanings, such as disorder, instability, confusion, and uncertainty (Yufeng and Fengxiant, 2009). The theory was first proposed by Boltzmann and expounded by Shannon (1948) later on. Entropy can indirectly determine how to select the most influential factors from the many available factors (Lombardo et al., 2015). Therefore, this theory can have a significant impact on the sieving process of thematic layers contributing to landslide susceptibility modeling (Sharma et al., 2012). Shannon entropy is underpinned by a simple yet practical mathematical equation, which is known as frequency ratio. The FR is a primary statistical model that quantifies the relationship between the factor’s classes within which the landslides have occurred and, accordingly, the relative rate of that particular class and the weights of the factors themselves. In general, the mathematics of Shannon entropy can be presented as follow (Tay et al., 2014; Sharma and Mahajan, 2019):
\[ P_{ij} = \frac{a}{b} \]  \hspace{1cm} (1)

\[ (P_{ij}) \frac{P_{ij}}{\sum_{i=1}^{n} P_{ij}} \quad \text{‘m number of landslides} \]  \hspace{1cm} (2)

\[ H_{j} = -\sum_{i=1}^{n} (P_{ij}) \times \log_{2} P_{ij}, \quad j = 1, \ldots, n \]  \hspace{1cm} (3)

\[ H_{j\max} = \log_{2} S_{j}, \quad S_{j} = \text{number of classes} \]  \hspace{1cm} (4)

\[ I_{j} = \frac{H_{j\max} - H_{j}}{H_{j\max}} \quad I = (0,1), \quad j = 1, \ldots, n \]  \hspace{1cm} (5)

\[ W_{j} = I_{j} P_{ij} \]  \hspace{1cm} (6)

“where a and b are the pixel percentages and landslides percentages, respectively, \((P_{ij})\) is the probability density, \(H_{j}\) and \(H_{j\max}\) represent values of the entropy, \(I_{j}\) is the information coefficient, and \(W_{j}\) represents the resultant weight value for the parameter as a whole (Sharma and Mahajan, 2019)”.

**Machine learning/ data mining models**

Data mining (DM) is known the extraction of information from a big database (Győrödi, 2003). The DM is a branch of computer science that started in the late 1980s by applying the concepts and methods related to artificial intelligence, patterns identification, database systems, and the science of statistics (Chandrasekaran et al., 2019). In this research, some well-known data mining models with previously reported good performances were adopted, namely maximum entropy (ME), SVM, ANN, and GLM, which were all implemented in the ModEco software. The Kappa histogram was plotted and treated as the Jackknife test to determine the most imperative factors that are positively contributing to the modeling process of landslide susceptibility in the Cherikabad Watershed. All data mining models were executed in a software namely the ModEco according to Guo and Liu (2010).

**Support vector machine**

The SVM is a robust data mining technique that is underpinned by theory of statistical learning and was first introduced by Boser et al. (1992). The SVM model pivots on a robust pattern recognition algorithm. In general, in linear problems, the SVM creates a separating line between diver patterns so that maximized the margin between the points (Kecman, 2005). In highly complex nonlinear subjects, the SVM transfers into an n-dimensional space using the kernel trick, where it can better distinguish the patterns and set the separating line termed as the hyperplane (Statnikov, 2011). There are different kernel functions, among which radial basis function (RBF) has been the most popular algorithm and showed outstanding performance in other literature (Aiserman et al., 1964). More details can be found in Yao et al. (2008) and Marjanović et al. (2011).

**Maximum entropy**
The ME is known as a presence-only generative data mining model based on the theory of information (Phillips et al., 2006). Shannon (1948) first coined the term entropy as the information content stored in the data and the degree to which a phenomenon can be relatively unpredictable. A higher degree of relative unpredictability can reciprocally provide higher information content (Shannon, 1951). Hence, Shannon expounded that the ME can increase the content of information and enable the model to better discern the pattern of interest (Kornejady et al., 2017). Phillips et al. (2004) first developed the MaxEnt software for spatial modeling of the present locations of the phenomena, and by walking through the parameters space with a first guess and further modifications, can perpetually distinguish the presences from absences (Kornejady et al., 2017). In this process, the MaxEnt randomly extracts 10,000 pseudo-absence locations that enable the model to better differentiate the pattern. Details can be observed in Phillips et al. (2004, 2006) and Elith et al. (2011).

**Artificial neural network**

The ANN imitates humans’ neural systems in which information is passed through neurons from one synapse to another (Cherkassky et al., 2006). ANNs are machine learning algorithms made of three main layers, namely input, hidden, and output (Saha et al., 2002). Input layers are responsible for transferring the input data to the hidden layer where the input data are processed through different sub-layers and assigned different weights (either positive or negative) by applying sigmoid activation functions (Yesilnacar and Topal, 2005). The output layer analyzes the system’s behavior in response to the gained information and the learning process (Prasad et al., 2012). A specific type of ANNs, named back-propagation neural network (BPNN), executes the learning and error analyzing process through a forward feeding flows of information and backward error propagation cycles (Kumar and Mathur, 2001). ANNs are useful tools once nonlinear data are involved; however, in pursuit of finding the most suitable solution with the least error (global minima), the model may get stuck in local minima, which have been mistakenly recognized by the model as the global minima. Moreover, if the model structure is formed of many layers, the learning process may become cumbersome and prolonged. More details can be found in (Kumar and Mathur, 2001; Yesilnacar and Topal, 2005).

**Generalized linear model**

The GLM simply seeks a linear regression between the landslide locations and effective variables via a link function (Brenning, 2005; Felicísimo et al., 2013; Conoscenti et al., 2014). The latter connects the linear predictors and the mean of an exponential probability distribution function. The GLM was first expounded by Nelder and Wedderburn (1972). As the name of GLM implies, it is a parametric method with limited and disputable assumptions that would not be well-suited to nonlinear predictors. More details are given in Lee and Pradhan (2007).

**Model's validation**

After preparing the landslide susceptibility maps using adopted data mining models, the validation dataset (40% of the total landslide numbers) (Park & Kim, 2019) was used to test the validity of the results. The latter was tested using the ROC curve (i.e., AUROC value). The ROC curve plots the false positives (i.e., 1-specificity, incorrectly predicting the non-landslide locations as landslides) on the x-axis against the true positives (i.e., sensitivity, correctly predicting the landslide locations as landslides).
predicting the landslide locations as observed in nature). The AUC is the representative of the model's performance in which the AUROC value of 1 shows a perfect model, while values equal or close to 0.5 signify a neutral model whose results are derived from pure random chance (Pontius and Schneider, 2001; Yilmaz, 2009). Once the ROC curve is plotted only by the training set, it represents the goodness-of-fit and the learning skill of the model, whereas the validation set-derived ROC curve only states the prediction power and the generalization capacity of the model. For drawing the ROC curve, the Performance Measure Tool (PMT) applied in ArcGIS 10.4 (Rahmati et al., 2019).

Results And Discussion

Validation of the Shannon entropy model

After identifying the factors affecting the landslide occurrence and preparing their classified maps using GIS, the weight of each of these factors was calculated and analyzed following the previously mentioned equations (Table 2). The FR was applied to determine the relationship between the landslides and the most influential variables. Landslide susceptibility map derived from Shannon entropy was categorized into five classes (i.e., very low, low, moderate, high, and very high). The Natural Break method used for this classification (five classes). According to Table (3), portion of high and very high susceptibility zones is about 32% of the Cherikabad Watershed. 

Based on validation's results, the AUROC value of the ME model was found to be about 0.88 (Fig. 4). According to the classification suggested by Hosmer et al. (2013), the results stated that the ME model performs very well in identifying the highest contributing factors to landslide occurrence as well as susceptibility mapping across the basin.

Intercomparison of factors’ classes using Shannon entropy model

Some of the inferences derived from factors and the classes within were expected, while some unexpected results were also discernible. The latter makes any solitary factor analysis infeasible; that is, unexpected results should be scrutinized, and their synergistic interactions with other factors need to be further assessed. By doing so, a more straightforward explanation would emerge, which is provided in detail as follows.

Expected results

It is expected that landslide susceptibility should increase as rainfall increases, as with our results and those reported by Chen et al. (2019a) and Pham et al. (2019). However, such a constant increase is contingent on spatially intact conditions such as homogeneous lithology along the elevation gradient, which is rarely the case. For instance, slopes in higher altitudes of our study area are generally formed of rocks that are highly resistant to landsliding. Hence, the decline of landslide susceptibility in higher rainfall classes (i.e., above 441 mm) is also justified. The same applies to the slope gradient factor. Expectedly, steep slopes exert more weight on the underlying materials, which leads to landslide initiation; however, highly steep slopes are formed of highly weathering-resistant rocks, which manifests in the form of low pedogenic development of soil profiles at upper slope positions.

Moreover, moderate slopes in the study area (i.e., 24-34%) are covered by highly susceptible shale and silts and poorly-managed pastures with signs of overgrazing. Low susceptibility values at lower slope positions are also
justified due mainly to the low and less-effective gravitational pull. These results are similar to Kornejady et al. (2017).

Regarding the lithological formations, the tectonic status of the study area has led to diverse and heterogeneous outcrops that add up to 15 lithological formations, among which ultra-basic rocks, pillow-shaped rock mounds, and pelagic limestone, Mesozoic in age, had the highest rate of landslide density. Mainly, unmetamorphosed pelagic limestone by featuring a chaotic assemblage of clayey-marly sequences with an increase in shale thickness towards upper strata exhibits high susceptibility to different types of landslides, which has been previously reported in different studies such as Ortner and Kilian (2016) and Carabella et al. (2019). Similar results on landslide-prone limestones are also reported by Lau (2018), Liu et al. (2018), and Pham et al. (2019). Regarding the NDVI factor, areas with sparse vegetation gained the highest susceptibility values, which aligns with the role of vegetation canopy and root system in protection of slope surface and reinforcement of subsurface soil particles. The same result has been reported by Yalcin et al. (2008). North-facing slopes in the study area receive more rainfall and accordingly have higher moisture content, which explains the high concentration of landslides in these slopes. This result is in agreement with Zhang et al. (2018) and Chen et al. (2019b).

Changes of the land-use from rangeland to irrigation farming is a well-known driver of landslide occurrence, particularly when it takes place on moderate to steep slopes with susceptible lithological formations. Similarly, irrigation farming land covers in the Cherikabad Watershed gained the highest susceptibility value, which addresses the role of the artificial rise in groundwater table due to the excessive irrigation and the altered natural compactness of soil surface due to the use of agricultural machinery. Similar results have been reported by Pourghasemi et al. (2013) and Sharma and Mahajan (2019). Further, low to moderate altitudes receive more rainfall than higher altitudes due to the prevailing inverse rainfall gradient relationship in the region. Additionally, higher altitudes are comprised of landslide- and erosion-resistant rock formations, which coincides with Jaafari et al. (2014), Pourghasemi and Rossi (2017), Hong et al. (2017), and Kim et al. (2018).

**Unexpected results**

It is often expected to observe a decreasing susceptibility pattern by moving away from linear or point features such as streams, faults, roads, and residential areas, which are mostly mapped under distance functions; however, such an assumption only holds when such factors actively contribute to landsliding process. Nonetheless, even highly important distance-based thematic maps can show asymmetric susceptibility patterns due mainly to the adopted classification method and the pivotal role of factor outweighing effect. The latter regards the fact that factors interact in a complex manner; for instance, lithology and slope as the main predisposing factors can always interfere and outweigh other factors, hence resulting in asymmetric and unexpected results at first glance. Hence, we scrutinized the asymmetries in each factor by investigating the other underlying factors.

As for the distance from streams, most of the landslides and accordingly, the highest susceptibility is observed in the range of 185-310 m from streams, which also contains intense fault systems that exacerbated the predisposing condition for landslide occurrence. Asymmetries in this factor mostly pinpoint the range of 85-185 m; 70% of which have are located on gentle slopes—lower than 24% slope angle. Similarly, most of the roads are constructed in gentle slopes—almost flat. Moreover, as the result of local stakeholders’ participation in soil conservation projects, the slopes nearby roads are protected by afforestation, installment of drainage systems, and construction of concrete
masonry retaining walls. Hence, the role of the road system in landslide occurrence became negligible, which is in line with Regmi et al. (2010).

The role of rural residential areas in landslide occurrence is still controversial; some emphasized the role of landslide redistribution and reactivation of the old dormant landslide, while some simply rejected the idea because the residential fabric is built upon flat slopes. The latter reasoning is also debatable since some local slopes may exist within or in the vicinity of residential fabric, which can be triggered by anthropogenic agents. Nonetheless, it seems that any interpretation regarding the causative role of the village on landslide occurrence, if any, should be concentrated only in areas close to the village, and there is a big chance that in farther areas, other predisposing factors would prevail. As with our results, the effective and interpretable distance reaches 3.37 km from the village where landslide susceptibility decreases concurrently with taking distance from the village, while upper classes are prevailed by other factors such as the presence of irrigation farming.

Distance from faults also seems to be unexpected due to the abundance of fault systems in the region; however, faults require to be seismically active to partake in landslide occurrence, and such an assumption does not always hold. Moreover, factor outweighing is another plausible explanation for such a result based on which by moving away from faults, sparse vegetation, gentle slopes of faults, and the outcrop of susceptible formations. Somewhat similar results have been reported by Pourghasemi et al. (2018) and Devkota et al. (2013).

Data mining/machine learning models

The final maps extracted from four data mining models were classified in ArcGIS10.4 (Fig. 5). The percentages of high and very high classes for the ME, GLM, ANN, and SVM techniques were 13%, 35%, 23.5%, and 29%, respectively (Table 4).

Validation of data mining models

Regarding the AUROC values in the training stage, it is evident that SVM and ANN equally outperformed their counterparts by the respective values of 0.924 and 0.926, indicating a less desirable performance presented by GLM and ME with corresponding AUROC values of 0.763 and 0.74. The results of AUROC values obtained from the validation data for ME, GLM, ANN, and SVM models were 0.727, 0.751, 0.917, and 0.935, respectively. The latter stated that the SVM model with the AUROC value of 0.935 outperforms its counterparts in the Cherikabad Watershed (Fig. 6). Outperformance of SVM has also been reported in some publications (e.g., Hipni et al. 2013; Pham et al. 2016; Chen et al. 2017c; Chen et al. 2018a), where most of its success is known to be related to its robust pattern recognition algorithm. Antithetically, ANNs are more prone to get stuck in local minima in pursuit of finding the global minimum, winding up with a less optimal solution where the most of training problems and generalizations occurred (Chen et al. 2017b). MaxEnt can also be deprived of good differentiation between presence and absence patterns since it only operates on presence location of landslides (Komejady et al. 2017).

Factor importance analysis
According to Fig. 7 and the Jackknife test, three factors, namely altitude, annual mean rainfall, and distance from the village, were ranked as the main contributing variables in the modeling process of landslide susceptibility and hazard in the Cherikabad Watershed. The latter aligns with the observed evidences during field surveys.

**Conclusion**

In this work, the SVM and ANN models showed outstanding performances in terms of high goodness-of-fit and prediction power, outperforming their statistical counterparts. According to the literature review, the ANN can roughly learn and optimize the corresponding learning parameters under limited landslide locations. The latter can lead to a model with somewhat less training skill, which can accordingly wind up in low prediction power and generalization capacity. Among different data mining/machine learning techniques, the SVM model is known to be underpinned by a robust pattern recognition algorithm, which is formulated through the process of setting maximized margins between the support vectors. One of the most important advantages of the SVM in comparison to the ANN, it is that the ANN has some problems with local minima, whereas the SVM solves this subject with a unique power.

On the other hand, the ANNs often end up in the best local solution (local minima), which is not the final best solution (global minima). Regarding the pitfalls encountered by the MaxEnt algorithms, the main issue regards its presence-only modeling feature. Despite the merits of presence-only modeling schemes such as avoiding to presume the prevalence of the phenomenon, it is noteworthy that MaxEnt inevitably extracts 10,000 pseudo-absence locations to distinguish the presences from absences. Although such a process enables the model to better differentiate the landslide occurrence pattern across the study area, the pseudo-absence locations are randomly selected, and there is a reasonable chance that they may have been selected from areas that have a high susceptibility to landsliding, but the phenomenon is not morphologically emerged due to the lack of triggering factors and the overweening of stability-favoring predisposing factors in the factors’ interaction pool. Moreover, it is known that solving nonlinear relationships among the factors by MaxEnt is somewhat problematic since it operates under some limited quantification equations called features such as linear, quadratic, product, threshold, and hinge, while the SVM and ANN better handle nonlinear relationships.

**References**

Abedini M, Ghasemian, B, Shirzadi A, Shahabi H, Chapi K, Pham B.T, Bin Ahmad B, Tien Bui D (2018) A novel hybrid approach of Bayesian Logistic Regression and its ensembles for landslide susceptibility assessment. Geocarto International, 34(13): 1427–1457

Achour Y, Garçia S, Cavaleiro V (2018) GIS-based spatial prediction of debris flows using logistic regression and frequency ratio models for Zêzere River basin and its surrounding area, Northwest Covilhã, Portugal. Arabian Journal of Geosciences 11(18): 550

Ada M, San BT (2018) Comparison of machine-learning techniques for landslide susceptibility mapping using two-level random sampling (2LRS) in Alakir catchment area, Antalya, Turkey. Natural Hazards 90(1): 237–263

Aditian A, Kubota T, Shinhara Y (2018) Comparison of GIS-based landslide susceptibility models using frequency ratio, logistic regression, and artificial neural network in a tertiary region of ambon, Indonesia. Geomorphology 318: 101–111
Agostini A, Tofani V, Nolesini T, Gigli G, Tanteri L, Rosi A, Cardellini S, Casagli, N (2014) A new appraisal of the Ancona landslide based on geotechnical investigations and stability modelling. Quarterly journal of engineering geology and hydrogeology 47(1): 29–43

Aiserman M.A, Braverman E. M, & Rozonoer L. I (1964) Theoretical foundations of the potential function method in pattern recognition. Avtomat. i Telemeh 25: 917-936

Aleotti P, Chowdhury R (1999) Landslide hazard assessment: summary review and new perspectives. Bulletin of Engineering Geology and the environment 58(1): 21-44

Ali S.A, Khatun R, Ahmad A, Ahmad S.N (2019) Application of GIS-based analytic hierarchy process and frequency ratio model to flood vulnerable mapping and risk area estimation at Sundarban region, India. Modeling Earth Systems and Environment 5(3):1083-1102

Alkhasawneh M.S, Ngah U.K, Tay L.T, Isa N.A.M, Al-Batah M.S (2014) Modeling and Testing Landslide Hazard Using Decision Tree. Journal of Applied Mathematics, pp 1-9.

Alvioli M, Melillo M, Guzzetti F, Rossi M, Palazzi E, von Hardenberg J, Brunetti M.T, Peruccacci S (2018) Implications of climate change on landslide hazard in central Italy. Science of the Total Environment 630: 1528–1543

Ambrosi C, Strozzi T, Scapozza C, Wegmüller U (2018) Landslide hazard assessment in the Himalayas (Nepal and Bhutan) based on Earth-Observation data. Engineering Geology 237: 217–228

Arabameri A, Pradhan B, Rezaei K, Lee S, Sohrabi M (2019b) An Ensemble Model for Landslide Susceptibility Mapping in a Forested Area. Geocarto International, pp1-26

Arabameri A, Pradhan B, Rezaei K, Sohrabi M, Kalantari Z (2019a) GIS-based landslide susceptibility mapping using numerical risk factor bivariate model and its ensemble with linear multivariate regression and boosted regression tree algorithms. Journal of Mountain Science 16(3): 595-618

Billah M, González PA, Castro Delgado R.C (2019) Patterns of Mortality Caused by Natural Disasters and Human Development Level: A South Asian Analysis. Indian Journal of Public Health Research & Development 10(2):312-316

Boser B.E, Guyon I.M, Vapnik, VN (1992) A training algorithm for optimal margin classifiers. In Proceedings of the fifth annual workshop on Computational learning theory, pp 144-152

Brenning A (2005) Spatial prediction models for landslide hazards: review, comparison and evaluation. Natural Hazards and Earth System Sciences 5: 853-862

Broeckx J, Vanmaercke M, Duchateau R, Poesen J, (2018) A data-based landslide susceptibility map of Africa. Earth-Science Reviews 185: 102–121

Cao Y, Yin K, Zhou C, Ahmed B (2020) Establishment of Landslide Groundwater Level Prediction Model Based on GA-SVM and Influencing Factor Analysis. Sensors. 20(3):845

Carabella C, Miccadesi E, Paglia G, Sciarrà N (2019) Post-wildfire landslide hazard assessment: The case of the 2017 Montagna del Morrone fire (Central Apennines, Italy). Geosciences 9(4): pp 175
Carrara A, Guzzetti F, Cardinali M, Reichenbach P (1999) Use of GIS technology in the prediction and monitoring of landslide hazard. Natural hazards 20(2-3): 117-135

Chandrasekaran S. S, Senthilkumar V, Maji V. B (2019) Landslides in Nilgiris: Causal Factors and Remedial Measures. In Geotechnical Design and Practice (pp 183-193). Springer, Singapore

Chang J.M, Chen H, Jou B.J.D, Tsou N.C, Lin G.W (2017) Characteristics of rainfall intensity, duration, and kinetic energy for landslide triggering in Taiwan. Engineering Geology 231: 81–87

Chen W, Ding X, Zhao R, Shi S, (2016b) Application of frequency ratio and weights of study of frequency ratio, statistical index and weights-of-evidence models in landslide susceptibility mapping. Environmental Earth Sciences 75(1): p 64

Chen W, Li W, Chai H, Hou E, Li X, Ding X (2016a) 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(1): p 63

Chen W, Panahi M, Tsangaratos P, Shahabi H, Ilia I, Panahi S., Li, S, Jaafari A, Ahmad B. B (2019c) Applying population-based evolutionary algorithms and a neuro-fuzzy system for modeling landslide susceptibility. Catena 172: 212-231

Chen W, Peng J, Hong H, Shahabi H, Pradhan B, Liu J, Zhu A.X, Pei X, Duan Z (2018b) Landslide susceptibility modelling using GIS-based machine learning techniques for Chongren County, Jiangxi Province, China. Science of the total environment 626: 1121–1135

Chen W, Pourghasemi H.R, Naghibi S.A (2018c) Prioritization of landslide conditioning factors and its spatial modeling in Shangnan County, China using GIS-based data mining algorithms. Bulletin of Engineering Geology and the Environment 77(2): 611–629

Chen W, Pourghasemi H.R, Naghibi S.A, (2018a) A comparative study of landslide susceptibility maps produced using support vector machine with different kernel functions and entropy data mining models in China. Bulletin of Engineering Geology and the Environment 77(2): 647–664

Chen W, Pourghasemi H.R, Zhao Z, (2017a) A GIS-based comparative study of Dempster-Shafer, logistic regression and artificial neural network models for landslide susceptibility mapping. Geocarto International 32 (4): 367–385

Chen W, Pourghasemi H.R, Kornejady A, Xie X (2019a) GIS-Based Landslide Susceptibility Evaluation Using Certainty Factor and Index of Entropy Ensembled with Alternating Decision Tree Models. In Natural Hazards GIS-Based Spatial Modeling Using Data Mining Techniques (pp 225-251). Springer, Cham.

Chen W, Pourghasemi H.R, Kornejady A, Zhang N, (2017b) Landslide spatial modeling: introducing new ensembles of ANN, MaxEnt, and SVM machine learning techniques. Geoderma 305: 314–327

Chen W, Pourghasemi H.R, Panahi M., Kornejady A, Wang J, Xie X, Cao S (2017c) Spatial prediction of landslide susceptibility using an adaptive neuro-fuzzy inference system combined with frequency ratio, generalized additive model, and support vector machine techniques. Geomorphology 297: 69–85
Chen W, Sun Z, Han J (2019b) Landslide Susceptibility Modeling Using Integrated Ensemble Weights of Evidence with Logistic Regression and Random Forest Models. Applied Sciences 9(1): 171

Cherkassky V, Krasnopolsky V, Solomatine D.P, Valdes J (2006) Computational intelligence in earth sciences and environmental applications: Issues and challenges. Neural Networks 19(2): 113-121

Choubin B, Moradi E, Golshan M, Adamowski J, Sajedi-Hosseini F, Mosavi A (2019) 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

Chudý F, Slámová M, Tomaštík J, Prokešová R, Mokroš M (2019) Identification of micro-scale landforms of landslides using precise digital elevation models. Geosciences. 9(3):117

Conoscenti C, Ciaccio M, Caraballo-Arias N.A, Gómez-Gutiérrez Á, Rotigliano E, Agnesi V (2015) Assessment of susceptibility to earth-flow landslide using logistic regression and multivariate adaptive regression splines: a case of the Belice River basin (western Sicily, Italy). Geomorphology 242, 49-64

Cummins PR (2019) Irrigation and the Palu landslides. Nature Geoscience, 12(11):881-882

Dang V. H, Hoang N. D, Nguyen L. M. D, Bui D. T, Samui P (2020) A Novel GIS-Based Random Forest Machine Algorithm for the Spatial Prediction of Shallow Landslide Susceptibility. Forests 11(1): p 118.

Demir G, Aytekin M, Akgün A, İkizler S.B, Tatar O (2013) A comparison of landslide susceptibility mapping of the eastern part of the north Anatolian fault zone (Turkey) by likelihood-frequency ratio and analytic hierarchy process methods. Natural Hazards 65(3): 1481–1506

Devkota K.C, Regmi A.D, Pourghasemi H.R, Yoshida K, Pradhan B, Ryu I.C, Dhital M.R, Althuwaynee O.F (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. Natural Hazards 65(1): 135–165

Elith J, Phillips S.J, Hastie T, Dudík M, Chee Y.E, Yates C.J (2011) A statistical explanation of MaxEnt for ecologists. Diversity and distributions 17(1), 43-57

Felícísimo Á.M, 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(2), 175-189

Fell R, Corominas J, Bonnard C, Cascini L Leroi E, (2008b) 350 Savage WZ on behalf of the JTC-1 Joint Technical Committee on Landslides and Engineered Slopes Guidelines for landslide susceptibility, hazard and risk zoning for land use planning. Engineering Geology 102: 85-98

Guo Q, Liu Y (2010) ModEco: an integrated software package for ecological niche modeling. Ecography 33(4): 637-642

Guzzetti F, Carrara A, Cardinali M, Reichenbach P (1999) Landslide hazard evaluation: a review of current techniques and their application in a multi-scale study, Central Italy. Geomorphology 31(1-4): 181-216

Győrődi R.S (2003) A comparative study of iterative algorithms in association rules mining. Studies in Informatics and Control 12(3): p 205
Hadji R, Achour Y, Hamed Y (2017) Using GIS and RS for slope movement susceptibility mapping: comparing AHP, LI and LR methods for the Oued Mullah Basin, NE Algeria. In Euro-Mediterranean Conference for Environmental Integration (pp 1853-1856). Springer, Cham.

Hao, L., Westen, C. V., KS, S., Martha, T. R., Jaiswal, P., & McAdoo, B. (2020). Constructing a complete landslide inventory dataset for the 2018 Monsoon disaster in Kerala, India, for land use change analysis. Earth System Science Data Discussions, 1-32.

He S, Pan P, Dai L, Wang H, Liu J, (2012) Application of kernel-based Fisher discriminant analysis to map landslide susceptibility in the Qinggan River delta, Three Gorges, China. Geomorphology, 171: 30–41

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

Hibert C, Michéa D, Provost F, Malet JP, Geertsema M (2019) Exploration of continuous seismic recordings with a machine learning approach to document 20 yr of landslide activity in Alaska. Geophysical Journal International. 219(2):1138-47

Hipni A, El-shafie A, Najah A, Karim OA, Hussain A, Mukhlisin M (2013) Daily forecasting of dam water levels: comparing a support vector machine (SVM) model with adaptive neuro fuzzy inference system (ANFIS). Water resources management. 27(10):3803-3823

Hong H, Miao Y, Liu J, Zhu AX, (2019) Exploring the effects of the design and quantity of absence data on the performance of random forest-based landslide susceptibility mapping. Catena 176:45–64

Hong H, Chen W, Xu C, Youssef A.M, 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 International 32(2): 139–154

Hong H, Pradhan B, Xu C, Bui D.T (2015) Spatial prediction of landslide hazard at the Yihuang area (China) using two-class kernel logistic regression, alternating decision tree and support vector machines. Catena 133: 266-281

Hosmer Jr DW, Lemeshow S, Sturdivant R.X (2013) Applied logistic regression. John Wiley & Sons.

Hou X, Vanapalli S.K, Li, T (2018) Water infiltration characteristics in loess associated with irrigation activities and its influence on the slope stability in heifangtai loess highland, china. Engineering Geology 234: 27–37

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

Jaafari A, Pourghasemi H.R (2019) Factors Influencing Regional-Scale Wildfire Probability in Iran: An Application of Random Forest and Support Vector Machine. In Spatial Modeling in GIS and R for Earth and Environmental Sciences (pp 607-619). Elsevier.

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. 175:430-45
Kalantar B, Pradhan B, Naghibi S.A, Motevalli A, Mansor S (2018) Assessment of the effects of training data selection on the landslide susceptibility mapping: A comparison between support vector machine (SVM), logistic regression (LR) and artificial neural networks (ANN). Geomat. Natural Hazards Risk 9(1): 49–69

Kecman V (2005) Support vector machines—an introduction. In Support vector machines: theory and applications (pp. 1-47). Springer, Berlin, Heidelberg.

Khan H, Shafique M, Khan M.A, Bacha M.A, Shah S.U, Calligaris C (2019) Landslide susceptibility assessment using Frequency Ratio, a case study of northern Pakistan. The Egyptian Journal of Remote Sensing and Space Science 22(1), 11-24

Kim J.C. Lee S, Jung H.S, Lee S (2018) Landslide susceptibility mapping using random forest and boosted tree models in Pyeong-Chang, Korea. Geocarto international 33(9): 1000-1015

Kim M.S, Onda Y, Kim J.K, Kim S.W (2015) Effect of topography and soil parameterization representing soil thicknesses on shallow landslide modelling. Quaternary International 384: 91–106

Kornejady A, Ownegh M, Rahmati O, Bahremand A, (2017) Landslide susceptibility assessment using three bivariate models considering the new topo-hydrological factor: HAND. Geocarto International 33(11), 1155-1185

Kumar Arora M, Mathur S (2001) Multi-source classification using artificial neural network in a rugged terrain. Geocarto International 16(3), 37-44

Lacroix P, Dehecq A, Taipe E (2020) Irrigation-triggered landslides in a Peruvian desert caused by modern intensive farming. Nature Geoscience. 13(1):56-60

Lau N.N (2018) Determination of ground displacement of 25 April 2015 Nepal earthquake by GNSS precise point positioning Vietnam. Journal of Earth Sciences 40:17–25

Lee M.J, Park I, Lee S (2015) Forecasting and validation of landslide susceptibility using an integration of frequency ratio and neuro-fuzzy models: a case study of Seorak mountain area in Korea. Environmental earth sciences 74(1): 413-429

Lee S (2007) Comparison of landslide susceptibility maps generated through multiple logistic regression for three test areas in Korea. Earth Surface Process and Landforms: The Journal of the British Geomorphological Research Group 32(14): 2133–2148

Lee S, Pradhan B (2007) Landslide hazard mapping at Selangor, Malaysia using frequency ratio and logistic regression models. Landslides 4(1), 33-41

Lin GW, Kuo HL, Chen CW, Wei LW, Zhang JM. Using a Tank Model to Determine Hydro-Meteorological Thresholds for Large-Scale Landslides in Taiwan. Water. 2020 Jan;12(1):253.

Liu J, Duan Z (2018) Quantitative assessment of landslide susceptibility comparing statistical index, index of entropy, and weights of evidence in the Shangnan area, China. Entropy 20(11): p 868
Liu X, Zhao C, Zhang Q, Peng J, Zhu W, Lu Z (2018) Multi-Temporal Loess Landslide Inventory Mapping with C-, X- and L-Band SAR Datasets- A Case Study of Heifangtai Loess Landslides, China. Remote Sensing 10(11): 1756-1765

Liucci L, Melelli L, Suteau C, Ponziani F (2017) The role of topography in the scaling distribution of landslide areas: A cellular automata modeling approach. Geomorphology 290: 236–249

Lombardo L, Cama M, Conoscenti C, Märker M, Rotigliano EJNH (2015) Binary logistic regression versus stochastic gradient boosted decision trees in assessing landslide susceptibility for multiple-occurring landslide events: application to the 2009 storm event in Messina (Sicily, southern Italy). Natural Hazards 79(3): 1621-1648

Mandal S, Mondal S (2019) Knowledge-Driven Statistical Approach for Landslide Susceptibility Assessment Using GIS and Fuzzy Logic (FL) Approach. In Statistical Approaches for Landslide Susceptibility Assessment and Prediction (pp. 163-180). Springer, Cham.

Marjanović M, Kovačević M, Bajat B, Voženilek V (2011) Landslide susceptibility assessment using SVM machine learning algorithm. Engineering Geology 123(3), 225-234

Mersha T, Meten M (2020) GIS-based landslide susceptibility mapping and assessment using bivariate statistical methods in Simada area, northwestern Ethiopia. Geoenvironmental Disasters.7(1):1-22

Moayedi H, Khari M, Bahiraei M, Foong L.K, Bui D.T (2020) Spatial assessment of landslide risk using two novel integrations of neuro-fuzzy system and metaheuristic approaches; Ardabil Province, Iran. Geomatics, Natural Hazards and Risk 11(1): 230-258

Mohammadi S, Taiebat H (2016) Finite element simulation of an excavation-triggered landslide using large deformation theory. Engineering Geology, 205: 62–72

Mohammadnejad V, Asghari, S (2016) Landslide Risk Assessment using statistical methods in Barandoz Chai basin. Quantitative Geomorphological Research, 4(4): 181-191

Nelder J.A, Wedderburn R.W (1972). Generalized linear models. Journal of the Royal Statistical Society: Series A (General) 135(3), 370-384

Niculiță M, Mărgărint MC, Cristea AI (2019) Using archaeological and geomorphological evidence for the establishment of a relative chronology and evolution pattern for Holocene landslides. Plos one.14(12):e0227335

Niculiță M (2020) Landslide hazard induced by climate changes in North-Eastern Romania. In Climate Change, Hazards and Adaptation Options. 245-265. Springer, Cham

Ortner H, Kilian S (2016) Sediment creep on slopes in pelagic limestones: Upper Jurassic of Northern Calcareous Alps, Austria. Sedimentary Geology 344: 350-363

Palmisano F, Vitone C, Cotecchia F (2016) Methodology for landslide damage assessment. Procedia Engineering 161:511–515

Ojala AE, Mattila J, Markovaara-Koivisto M, Ruskeeniemi T, Palmu JP, Sutinen R (2019) Distribution and morphology of landslides in northern Finland: an analysis of postglacial seismic activity. Geomorphology. 326:190-201
Ozioko OH, Igwe O (2020) GIS-based landslide susceptibility mapping using heuristic and bivariate statistical methods for Iva Valley and environs Southeast Nigeria. Environmental monitoring and assessment. 192(2):1-9

Panahi M, Gayen A, Pourghasemi HR, Rezaie F, Lee S (2020) Spatial prediction of landslide susceptibility using hybrid support vector regression (SVR) and the adaptive neuro-fuzzy inference system (ANFIS) with various metaheuristic algorithms. Science of The Total Environment. 7:139937

Park S, Kim J (2019) Landslide Susceptibility Mapping Based on Random Forest and Boosted Regression Tree Models, and a Comparison of Their Performance. Applied Sciences 9(5): 942

Peng D, Xu Q, Liu F, He Y, Zhang S, Qi X, Zhao K, Zhang X (2018) Distribution and failure modes of the landslides in heitai terrace, china. Engineering. Geology 236: 97–110

Peres D.J, Cancelliere A (2018) Modeling impacts of climate change on return period of landslide triggering. Journal of Hydrology 567: 420–434

Persichillo M.G, Bordoni M, Meisina C (2017) The role of land use changes in the distribution of shallow landslides. Science of the total environment 574: 924–937

Pham B.T, Jaafari A, Prakash I, Bui D.T (2019) A novel hybrid intelligent model of support vector machines and the MultiBoost ensemble for landslide susceptibility modeling. Bulletin of Engineering Geology and the Environment 78(4): 2865-2886

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

Pham B.T, Shirzadi A, Tien Bui D, Prakash I, Dholakia MB (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 os Sediment Research 33(2): 157–170

Phillips S. J, Anderson R. P, Schapire R. E (2006) Maximum entropy modeling of species geographic distributions. Ecological modelling, 190(3-4), 231-259.

Phillips S.J, Dudik M, Schapire RE (2004) A maximum entropy approach to species distribution modeling. In Proceedings of the twenty-first international conference on Machine learning (p. 83)

Pontius Jr R. G, Schneider, L.C (2001) Land-cover change model validation by an ROC method for the Ipswich watershed, Massachusetts, USA. Agriculture, Ecosystems & Environment 85(1-3): 239-248

Pourghasemi H.R, Rahmati O (2018) Prediction of the landslides susceptibility: which algorithm, which precision? Catena 162: 177-192

Pourghasemi H.R, Rossi M (2017) Landslide susceptibility modeling in a landslide prone area in Mazandaran Province, north of Iran: a comparison between GLM, GAM, MARS, and M-AHP methods. Theoretical and Applied Climatology 130(1-2): 609-633

Pourghasemi H.R, (2014) Landslide hazard prediction using data mining methods in the North of Tehran City. Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy (Ph.D.) in
Watershed Management Engineering and Sciences, Department of Watershed Management Engineering, Faculty of Natural Resources, Tarbiat Modarres University. 143 PP.

Pourghasemi H.R, Jirandeh A.G, Pradhan B., Xu C, Gokceoglu C (2013b) Landslide susceptibility mapping using support vector machine and GIS at the Golestan Province, Iran. Journal of Earth System Science 122(2): 349-369

Pradhan B (2010) Landslide susceptibility mapping of a catchment area using frequency ratio, fuzzy logic and multivariate logistic regression approaches. Journal of Indian Society of Remote Sensing 38(2): 301–320

Pradhan B (2013) A comparative study on the predictive ability of the decision tree, support vector machine and neuro-fuzzy models in landslide susceptibility mapping using GIS. Computers & Geosciences 51:350-365

Pradhan B, Sameen M.I (2019) Modeling Traffic Accident Severity Using Neural Networks and Support Vector Machines. In Laser Scanning Systems in Highway and Safety Assessment (pp. 111-117). Springer, Cham.

Prasad R, Pandey A, Singh K.P, Singh V.P, Mishra R.K, Singh D (2012). Retrieval of spinach crop parameters by microwave remote sensing with back propagation artificial neural networks: A comparison of different transfer functions. Advances in space research 50(3): 363-370

Rahmati O, Kornejady A, Samadi M, Deo R C, Conoscenti C, Lombardo L, ... Bui D.T (2019) PMT: New analytical framework for automated evaluation of geo-environmental modelling approaches. Science of the total environment 664: 296-311

Razifard M, Shoaei G, Zare M (2019) Application of fuzzy logic in the preparation of hazard maps of landslides triggered by the twin Ahar-Varzeghan earthquakes (2012). Bulletin of Engineering Geology and the Environment 78(1): 223-245

Regmi A. D, Yoshida K, Pourghasemi H.R, DhitaL M. R, Pradhan B (2014) Landslide susceptibility mapping along Bhalubang–Shiwapur area of mid-Western Nepal using frequency ratio and conditional probability models. Journal of Mountain Science 11(5), 1266-1285

Regmi N. R, Giardino J. R, Vitek J. D, Dangol V (2010) Mapping landslide hazards in western Nepal: Comparing qualitative and quantitative approaches. Environmental & Engineering Geoscience, 16(2): 127-142.

Reichenbach P, Rossi M, Malamud B.D, Mihir M, Guzzetti F (2018) A review of statistically-based landslide susceptibility models. Earth-Science Reviews 180: 60–91

Rianna G, Reder A, Pagano L, Mercogliano P (2019) Assessing future variations in landslide occurrence due to climate changes: insights from an Italian test case. InNational Conference of the Researchers of Geotechnical Engineering. 255-264. Springer, Cham

Saha A.K, Gupta R.P, Arora M.K (2002) GIS-based landslide hazard zonation in the Bhagirathi (Ganga) valley, Himalayas. International journal of remote sensing 23(2), 357-369

Sarkar S, Kanungo D.P, Mehrotra G.S (1995) Landslide hazard zonation: a case study in Garhwal Himalaya, India. Mountain Research and Development 15(4): 301-309

Segoni S, Rosi A, Lagomarsino D, Fanti R, Casagli, N (2018) Brief communication: Using averaged soil moisture estimates to improve the performances of a regional-scale landslide early warning system. Natural Hazards and
Shafique M (2020) Spatial and temporal evolution of co-seismic landslides after the 2005 Kashmir earthquake. Geomorphology. 27:107228

Shannon C.E (1948). A mathematical theory of communication. Bell system technical journal 27(3): 379-423

Shannon C.E (1951). Prediction and entropy of printed English. Bell system technical journal, 30(1): 50-64

Sharma L.P, Patel N, Ghose M.K Debnath P (2012) Influence of Shannon's entropy on landslide-causing parameters for vulnerability study and zonation-a case study in Sikkim, India. Arabian Journal of Geosciences 5(3): 421-431

Sharma S, Mahajan A.K (2019) A comparative assessment of information value, frequency ratio and analytical hierarchy process models for landslide susceptibility mapping of a Himalayan watershed, India. Bulletin of Engineering Geology and the Environment 78(4): 2431-2448

Shirvani Z (2020) A Holistic Analysis for Landslide Susceptibility Mapping Applying Geographic Object-Based Random Forest: A Comparison between Protected and Non-Protected Forests. Remote Sensing 12(3): 434

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

Singh K, Kumar V (2017) Landslide hazard mapping along national highway-154A in Himachal Pradesh, India using information value and frequency ratio. Arabian Journal of Geosciences 10(24): 539

Singh K, Singh MP (2020) Causes and remedial measures for rockfall and landslides in Naini lake basin: Uttarakhand, India. Environment Conservation Journal. 21(1&2):95-102

Soma A.S, Kubota T, Mizuno H (2019) Optimization of causative factors using logistic regression and artificial neural network models for landslide susceptibility assessment in Ujung Loe Watershed, South Sulawesi Indonesia. Journal of Mountain Science 16(2): 383-401

Song J, Alves TM, Omosanya KO, Hales TC, Ze T (2020) Tectonic evolution of strike-slip zones on continental margins and their impact on the development of submarine landslides (Storegga Slide, northeast Atlantic). Geological Society of America Bulletin

Stanley T, Kirschbaum DB, Pascale S, Kapnick S (2020). Extreme Precipitation in the Himalayan Landslide Hotspot. In Satellite Precipitation Measurement.1087-1111. Springer, Cham.

Statnikov A, Aliferis C.F, Hardin D.P, Guyon I (2011) A gentle introduction to support vector machines in biomedicine, Vol.1: Theory and methods. World scientific.

Tay L.T, Lateh H, Hossain M.K, Kamil AA (2014) Landslide hazard mapping using a Poisson distribution: a case study in Penang Island, Malaysia. In Landslide Science for a Safer Geoenvironment (pp. 521-525). Springer, Cham.

Teimouri M, Kornejadi A (2019) The dilemma of determining the superiority of data mining models: optimal sampling balance and end users’ perspectives matter. Bulletin of Engineering Geology and the Environment. 79: 1707-1720.
Tien Bui D, 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(2): 361–378

Tsangaratos P, Iliadis R (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

Wang H, Liu G, Xu W, Wang G, (2005) GIS-based landslide hazard assessment: an overview. Progress in physical Geography 29(4): 548- 567

Wang L. J, Guo M, Sawada K, Lin J, Zhang J (2016) A comparative study of landslide susceptibility maps using logistic regression, frequency ratio, decision tree, weights of evidence and artificial neural network. Geosciences Journal 20(1): 117–136

Wang L.J, Guo M, Sawada K, Lin J, Zhang J (2015) Landslide susceptibility mapping in Mizunami City, Japan: a comparison between logistic regression, bivariate statistical analysis and multivariate adaptive regression spline models Catena 135: 271–282

Wang Q, Guo Y, Li W, He J, Wu Z (2019) Predictive modeling of landslide hazards in Wen County, northwestern China based on information value, weights-of-evidence, and certainty factor. Geomatics, Natural Hazards and Risk 10(1): 820-835

Wang T, Wu S.R, Shi J.S, Xin P, Wu L.Z (2018) Assessment of the effects of historical strong earthquakes on large-scale landslide groupings in the wei river midstream. Engineering Geology 235: 11–19

Wu Y, Ke Y, Chen Z, Liang S, Zhao H, Hong H (2020) Application of alternating decision tree with AdaBoost and bagging ensembles for landslide susceptibility mapping. Catena. 87:104396

Xie Z, Chen G, Meng X, Zhang Y, Qiao L, Tan L (2017) A comparative study of landslide susceptibility mapping using weight of evidence, logistic regression and support vector machine and evaluated by SBAS-InSAR monitoring: Zhouqu to wudu segment in bailing river basin, china. Environmental Earth Sciences 76(8): 313

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

Yao X, Tham L.G, Dai F.C (2008) Landslide susceptibility mapping based on support vector machine: a case study on natural slopes of Hong Kong, China. Geomorphology 101(4): 572-582.

Yosilnacar E, Topal T (2005) Landslide susceptibility mapping: a comparison of logistic regression and neural networks methods in a medium scale study, Hendek region (Turkey). Engineering Geology 79(3-4): 251-266

Yilmaz I (2009) Landslide susceptibility mapping using frequency ratio, logistic regression, artificial neural networks and their comparison: a case study from Kat landslides (Tokat-Turkey). Computers & Geosciences 35(6): 1125-1138

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(5): 839–856
Youssef A.M, Pradhan B, Jebu, M.N, El-Harbi H.M (2015) Landslide susceptibility mapping using ensemble bivariate and multivariate statistical models in Fayfa area, Saudi Arabia. Environmental Earth Sciences 73(7): 3745–3761

Yufeng S, Fengxiant J. (2009) Landslide Stability Analysis Based on Generalized Information Entropy. In: 2009 International Conference on Environmental Science and Information Application Technology 2: 83-85. IEEE.

Zaki M.J, Meira Jr W (2014) Data mining and analysis: fundamental concepts and algorithms. Cambridge University Press.

Zhang T, Han L, Chen, W, Shahabi H (2018) Hybrid integration approach of entropy with logistic regression and support vector machine for landslide susceptibility modeling. Entropy 20(11): p 884

Zheng Y (2015) Trajectory data mining: an overview. ACM Transactions on Intelligent Systems and Technology (TIST) 6(3):1-14

Zhou C, Yin,K, Cao Y, Ahmed B, Li Y, Catani F, Pourghasemi H.R (2018) Landslide susceptibility modeling applying machine learning methods: A case study from Longju in the Three Gorges Reservoir area, China. Computers & Geosciences 112: 23–37

Zhu A.X, Miao Y, Wang R, Zhu T, Deng Y, Liu J, Yang L, Qin C.Z, Hong H (2018) A comparative study of an expert knowledge-based model and two data-driven models for landslide susceptibility mapping. Catena 166: 317–327

Tables

Table 1 Description of the opted landslide-controlling factors in the study area
| Controlling factor         | Scale      | Data source                                                                 | Causative role                                                                 |
|---------------------------|------------|----------------------------------------------------------------------------|--------------------------------------------------------------------------------|
| Elevation                 | 1:50,000   | UAF\(^1\) (12.5×12.5 m grid)                                               | A proxy to rock resistance, vegetation cover, and precipitation regime          |
| Slope degree              | 1:50,000   | DEM-derived                                                               | Gravitational pull                                                            |
| Slope aspect              | 1:50,000   | DEM-derived                                                               | A proxy to soil moisture and vegetation cover                                 |
| Annual Rainfall           | 30 m * 30 m| Regional Water Company of West Azerbaijan                               | Climatic agent altering soil moisture content and causing erosion          |
| NDVI                      | 1:25,000   | Sentinel2 red and near-infrared bands (USGS)                            | Vegetation cover                                                             |
| Land use/land cover       | 30 m * 30 m| Landsat 8 RGB bands (USGS) & Google earth                             | Vegetation cover/ land conversion as an anthropogenic intervention          |
| Lithological formation    | 1:100,000  | GSI\(^2\)                                                              | Resistance to weathering                                                      |
| Distance from the village | 1:25,000   | NCCI\(^3\) & Google earth                                               | Anthropogenic intervention causing landslide redistribution and reactivation |
| Distance from stream      | 1:25,000   | NCCI & Google earth                                                      | Erosional process, slope undercut                                             |
| Distance from roads       | 1:25,000   | NCCI & Google earth                                                      | Anthropogenic intervention causing slope undercut                             |
| Distance from faults      | 1:100,000  | GSI                                                                     | Seismic activity                                                             |

1 Data Set of University of Alaska Fairbanks
2 Geological Survey of Iran (GSI)
3 National Cartographic Center of Iran

**Table 2** Weight of factors affecting the landslide occurrence
| Landslide Conditioning Factors | Class     | Pįj | (Pįj)  | Hįj | Hįj(max) | Iįj | Wįj |
|-------------------------------|-----------|-----|--------|-----|----------|-----|-----|
| slope (%)                     | 0-10      | 0.479 | 0.114  |     |          |     |     |
|                              | 10-18     | 1.103 | 0.263  |     |          |     |     |
|                              | 18-24     | 1.637 | 0.390  | 1.883| 2.322    | 0.189| 0.158|
|                              | 24-36     | 0.976 | 0.233  |     |          |     |     |
|                              | > 36      | 0.000 | 0.000  |     |          |     |     |
| Aspect                        | Flat      | 0.000 | 0.000  |     |          |     |     |
|                              | North     | 0.660 | 0.101  |     |          |     |     |
|                              | North West| 1.729 | 0.265  |     |          |     |     |
|                              | West      | 1.186 | 0.181  |     |          |     |     |
|                              | South West| 0.430 | 0.066  | 2.705| 3.170    | 0.147| 0.107|
|                              | South     | 0.128 | 0.020  |     |          |     |     |
|                              | South East| 0.293 | 0.045  |     |          |     |     |
|                              | East      | 1.186 | 0.181  |     |          |     |     |
|                              | East North| 0.924 | 0.141  |     |          |     |     |
|                              | 1648 – 1998| 1.180 | 0.235  |     |          |     |     |
|                              | 1998 – 2283| 0.207 | 0.041  |     |          |     |     |
| Elevation (m)                 | 2283 – 2561| 1.187 | 0.236  | 2.121| 2.322    | 0.087| 0.087|
|                              | 2561-2871 | 1.653 | 0.329  |     |          |     |     |
|                              | 2871 <    | 0.794 | 0.158  |     |          |     |     |
|                              | 0-160     | 1.025 | 0.235  |     |          |     |     |
| Distance to fault (m)         | 160-340   | 1.256 | 0.288  |     |          |     |     |
|                              | 340-540   | 0.776 | 0.178  | 1.608| 2.322    | 0.307| 0.278|
|                              | 540-790   | 1.306 | 0.299  |     |          |     |     |
|                              | 790 <     | 0.000 | 0.000  |     |          |     |     |
|                              | 0-85      | 1.031 | 0.227  |     |          |     |     |
| Distance to streams (m)       | 85-185    | 0.789 | 0.173  |     |          |     |     |
|                              | 185-310   | 1.631 | 0.359  | 1.603| 2.322    | 0.309| 0.282|
|                              | 310-500   | 1.097 | 0.241  |     |          |     |     |
|                              | 500 <     | 0.000 | 0.000  |     |          |     |     |
|                              | 0-1804    | 0.794 | 0.164  |     |          |     |     |
| Distance to villages (m)      | 1804-3370 | 0.658 | 0.136  |     |          |     |     |
| Distance to roads (m) | 3370-4868 | 1.120 | 0.231 | 2.256 | 2.322 | 0.028 | 0.027 |
|----------------------|-----------|-------|-------|-------|-------|-------|-------|
|                      | 4868-6365 | 1.485 | 0.306 |
|                      | 6365<     | 0.789 | 0.163 |
|                      | 0-170     | 0.692 | 0.129 |
|                      | 170-380   | 1.059 | 0.197 |
|                      | 380-600   | 1.182 | 0.220 | 2.256 | 2.322 | 0.028 | 0.030 |
|                      | 600-880   | 0.807 | 0.150 |
|                      | 880<      | 1.633 | 0.304 |
|                      | 413-424   | 1.044 | 0.216 |
|                      | 424-432   | 1.081 | 0.223 |
| Annual mean rainfall (mm) | 432-441 | 1.521 | 0.314 | 2.225 | 2.322 | 0.042 | 0.040 |
|                      | 441-450   | 0.697 | 0.144 |
|                      | >450      | 0.499 | 0.103 |
|                      | Rangeland | 1.236 | 0.152 |
|                      | garden    | 2.627 | 0.322 |
| Land Use/ land cover | Irrigation agriculture | 4.290 | 0.526 |
|                      | dry farming | 0.000 | 0.000 | 1.427 | 2.585 | 0.448 | 0.609 |
|                      | Rock Outcrop | 0.000 | 0.000 |
|                      | Water     | 0.000 | 0.000 |
| NDVI                 | -0.12-0.15 | 6.795 | 0.659 |
|                      | 0.15-0.25  | 2.064 | 0.200 | 2.089 | 2.322 | 0.100 | 0.207 |
|                      | 0.25-0.32  | 0.840 | 0.081 |
|                      | 0.32-0.45  | 0.387 | 0.037 |
|                      | >0.45      | 0.229 | 0.022 |
|                      | Ms         | 0.000 | 0.000 |
|                      | PC-Cs      | 0.000 | 0.000 |
|                      | Cbt        | 3.838 | 0.397 |
|                      | Msc        | 1.485 | 0.154 |
|                      | Qt2        | 0.000 | 0.000 |
| Lithological formation | Qal      | 0.000 | 0.000 |
|                      | Qt         | 0.000 | 0.000 | 2.355 | 3.907 | 0.397 | 0.256 |
|                      | om         | 1.138 | 0.118 |
Table 3  Areal percentage of landslide susceptibility classes generated by Shannon entropy model

| Landslide susceptibility class | Area (ha) | Percentage of landslide numbers |
|-------------------------------|-----------|---------------------------------|
| Very Low                      | 704       | 13                              |
| Low                           | 1655      | 30                              |
| Moderate                      | 1113      | 25                              |
| High                          | 917       | 17                              |
| Very High                     | 772       | 15                              |
| Total                         | 5180      | 100                             |

Table 4  Area percentage of landslide susceptibility classes generated by data mining/ machine learning models

| Landslide susceptibility classes | GLM Area (ha) | GLM Landslide Counts (%) | ME Area (ha) | ME Landslide Counts (%) | ANN Area (ha) | ANN Landslide Counts (%) | SVM Area (ha) | SVM Landslide Counts (%) |
|---------------------------------|---------------|--------------------------|--------------|--------------------------|---------------|--------------------------|---------------|--------------------------|
| Very Low                        | 2590.98       | 50                       | 1005.93      | 19                       | 2250.31       | 43                       | 2137.57       | 41                       |
| Low                             | 1219          | 23                       | 1210.69      | 23.5                     | 939.69        | 18                       | 839.16        | 16                       |
| Medium                          | 733.89        | 14                       | 1192.84      | 23                       | 803.06        | 15.5                     | 721.46        | 14                       |
| High                            | 439.34        | 8                        | 1084.06      | 21                       | 645.31        | 12.5                     | 668.68        | 13                       |
| Very High                       | 15192         | 5                        | 679.19       | 13.5                     | 538.3         | 11                       | 805.39        | 16                       |
| Total                           | 5180          | 100                      | 5180         | 100                      | 5180          | 100                      | 5180          | 100                      |
Figure 1

Location map of the study area (a) Figures of landslide extracted by camera (b) Figures of landslide extracted from Google Earth professional satellite image.
Figure 2

Flow chart of the overall methodology.
Figure 3

Landslide conditioning factors used for LSM in the study area: (a) Elevation, (b) Distance to streams, (c) Rainfall (d) slope aspect, (e) Distance to the village, (f) slope degree (g) Distance to faults, (h) Lithological formation, (i) Land cover, (j) NDVI, (k) distance to roads.
Figure 4
The ROC curve of the LSM derived from Shannon entropy model in the Cherikabad Watershed (plotted by PMT).

Figure 5
LSMs derived from different data mining models in the Cherikabad Watershed (a: ME; b: GLM; c: ANN; d: SVM).
Figure 6

ROC curves plotted for the implemented data mining models in the Cherikabad Watershed.
Figure 7

Kappa values presented in the jackknife histogram, identifying the most important factors affecting landslide occurrence in Cherikabad Watershed