

Novel Ensemble-Based Machine Learning Models Based on The Bagging, Boosting and Random Subspace Methods for Landslide Susceptibility Mapping

Ali Nouh Mabdeh (✉ alinouh@aabu.edu.jo)

Al al-Bayt University

Akif Al-Fugara

Al al-Bayt University

Mohammad Ahmadi

KN Toosi: KN Toosi University of Technology

Biswajeet Pradhan

University of Technology Sydney

Research Article

Keywords: Landslide susceptibility mapping, Ensemble models, GIS, Random subspace, Boosting

Posted Date: June 29th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-649364/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Novel Ensemble-Based Machine Learning Models Based on The Bagging, Boosting and Random Subspace Methods for Landslide Susceptibility Mapping

Ali Nouh Mabdeh^{1*}, A'kif Al-Fugara², Mohammad Ahmadlou³, Biswajeet Pradhan^{4,5,6}

¹ Department of GIS and Remote Sensing, Institute of Earth and Environmental Sciences, Al Al-Bayt University, Mafraq 25113, Jordan

² Department of Surveying Engineering, Faculty of Engineering, Al Al-Bayt University, Mafraq 25113, Jordan

³ GIS Department, Geodesy and Geomatics Faculty, K. N. Toosi University of Technology, Tehran 1996715433, Iran

⁴ Centre for Advanced Modelling and Geospatial Information Systems (CAMGIS), School of Information, Systems & Modelling, Faculty of Engineering and IT, University of Technology Sydney, Sydney, NSW 2007, Australia

⁵ Department of Meteorology, King Abdulaziz University, P.O. Box 80234, Jeddah 21589, Saudi Arabia

⁶ Earth Observation Center, Institute of Climate Change, Universiti Kebangsaan Malaysia, UKM, Bangi 43600, Selangor, Malaysia

* Correspondence: alinouh@aabu.edu

Novel Ensemble-Based Machine Learning Models Based on The Bagging, Boosting and Random Subspace Methods for Landslide Susceptibility Mapping

Abstract

Individual machine learning models show different limitations such as low generalization power for modeling nonlinear phenomena with complex behavior. In recent years, one of the best approaches to this issue is to use ensemble models. The purpose of this paper is to investigate the predictive power and modeling of three novel ensemble models constructed with four machine learning models: Decision Tree (DT), Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Naive Bayes (NB) models based on three approaches of Bagging, boosting and Random Subspace (RS) in landslide susceptibility mapping (LSM) in the Province of Ajloun in Jordan. A total number of 91 landslide locations along with 16 conditioning factors in LSM were identified and used. Also, before modeling, the selection of effective conditioning factors in LSM was done using genetic algorithm and four single models including DT, KNN, NB and SVM. The selected factors were used in modeling with individual and ensemble models. The results show that the area under the receiver operating characteristic curve (AUROC) for ensemble models is significantly higher than the individual models and the AUC for ensemble models was on average 14% higher than individual models. Based on the results, the most accurate models were RS ensemble model (AUROC = 0.850), Boosting (AUROC = 0.848) and Bagging (AUROC = 0.814), respectively. This study showed that by combining the results of simple machine learning models and making ensemble models, models with the desired accuracy can be achieved.

Keywords: Landslide susceptibility mapping; Ensemble models; GIS; Random subspace; Boosting

1. Introduction

Encompassing a wide range of natural and man-made risks, environmental hazards are barriers to the development of various regions (Cutter 1996; Hunter 2005; Paul 2011). Landslides are a prominent environmental hazard with relatively severe impacts on socio-cultural systems and infrastructures so that they are considered as a barrier to socio-economic growth, and development of any region. Annually, this phenomenon accounts for significant damages and loss of lives (Dikshit & Satyam 2018; Parkash 2011). Rapid soil erosion and the destruction of agricultural and residential lands, forests, and roads are amongst such losses (Dolidon et al 2009; Pradhan et al. 2012). Landslide-induced destructive impacts are being increasingly intensified by the urbanization (Cui et al. 2019). Landslides are

characterized by widespread distribution, frequent occurrence, rapid motion, and drastic losses with horrible impact in high-densely populated regions (Bardi et al. 2017; Persichillo et al. 2017).

Despite above mentioned, landslides could be well-managed than other natural disasters. The measures such as modifying slope geometry, installing structures such as piles and anchors, stabilizing structures and rerouting surface and underwater drainage can be adopted for landslide disaster mitigation (Popescu & Sasahara 2009). To this end, identifying susceptible regions is one of the most important steps to preventing resultant losses, and provides great supports to natural resources management and development plans (Kamp et al. 2008; Kavzoglu et al 2014). Identifying the influential factors of landslides in addition to appropriate susceptibility mapping, with the aim of determining high-risk regions, provides planners with essential and beneficial information (Pham et al. 2019). Landslide susceptibility mapping, which represents the spatial distribution of landslides, is an efficient tool for mitigating landslide-induced impacts, and efficient land use planning (Huang & Zhao 2018; Wang et al. 2020). In addition, susceptibility zoning gives a better estimation of a zone's potential capabilities, and divides a susceptible zone to several subzones, enabling planners to work more easily. Furthermore, scientific and accurate zoning is a time and cost effective approach in terms of mitigating resultant risks.

To date, scholars have provided various models for modeling and mapping landslide-susceptible regions. They are generally data-driven or expert-based models, or a combination of them (Arabameri et al. 2018; Zhu et al. 2018). In recent years, data-driven approaches and artificial intelligent algorithms have been in the area of focus due to their higher efficacy. Random forest (RF), Neural Networks (NNs), Support Vector Machine (SVM), Decision Tree (DT), K-nearest neighbors (KNN), Extreme Learning Machine (ELM) and Naïve Bayes (NB) are amongst such techniques used for landslide susceptibility mapping (Boateng et al. 2020; Chen et al., 2019; Dou et al., 2019; Huang et al. 2017).

Different models have diverse capabilities and functions. Therefore, attentions have been recently attracted to integrating the output of multiple models in order to benefit from the advantages of each single model, and eventually to obtain more precise and accurate predictions than could be obtained from any of the constituent models alone (Brown 2010; Sagi & Rokach 2018). Ensemble models are integrated techniques which have been in the area of focus due to their capability of improving the results of learning models and techniques used to solve modeling and classification problems (Arabameri et al, 2020; Pham et al., 2020; Roy et al, 2019). This study proposes and compares the use of three novel ensemble model by bagging, boosting and random subspace (RS) techniques and based on SVM,

KNN, NB, and DT models for landslide susceptibility mapping in Ajloun province of Jordan. The obtained results from ensemble models will be compared by those of the constituent models.

2. Study area

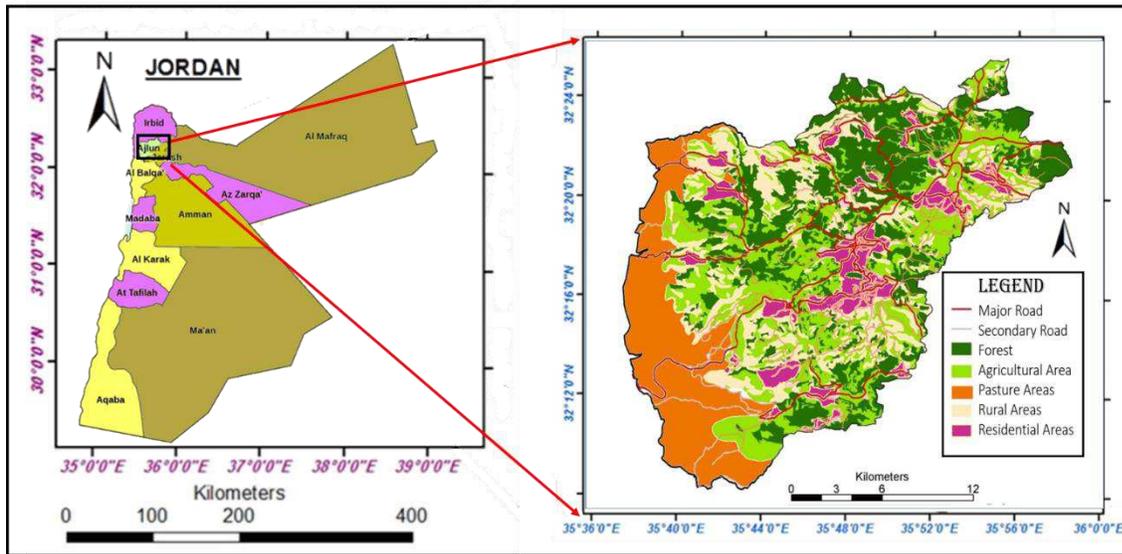


Figure 1. Study area location map

Ajloun Governorate is one of the governorates in northern Jordan, and the most important cities of Ajloun Governorate is Ajloun city, which considered as the economic and administrative center of the governorate located in ($32^{\circ}19'57''$ N , $35^{\circ}45'6''$ E). Although Ajloun Governorate is the second smallest governorate in the Kingdom in terms of area with 420 km^2 , it has the highest governorate in the Kingdom in terms of the percentage of vegetation cover. The topography of the area varies greatly, while some of the heights in the governorate are considered among the highest in the Kingdom with 1247 meters above sea level, some areas of the governorate are located under the sea surface, which means that there are many steep slopes in the study area, and this is what causes many landslides accidents in it. In recent years, the investment from the government and the local community has resulted in the growth of the tourism infrastructure in Ajloun, especially after the locked down caused by the Corona pandemic, which raised the interest in domestic tourism due to the ban on travel outside the country, which made such a study necessary to help decision makers know the places that are Vulnerable to landslides.

3. Methodology

Landslide inventory map along with 16 initial conditioning factors for modeling were first prepared. Then, the relationships between each class of conditioning factors and landslide occurrence were evaluated in the study area using Frequency Ratio (FR) method (Yilmaz 2009). Next, four wrapper-based feature selection algorithms were developed and used to select effective features. The selected optimal features were used in the models and ensemble models were built by bagging, boosting and RS approaches. Finally, the built models were evaluated, and landslide susceptibility maps were developed. Fig. 2 shows the steps of this study.

3.1. Preparation of landslide inventory map

The main types of landslides that occur in the study area are: Rock slides, Slumping, Earth flow, Toppling, Rock spreads, It is noted that most of these landslides occur when the steep areas contain geological structures consisting of Limestone and marly limestone, as same as Weak marl and clay. Multiple field inspections and ground control points (GCPs) were conducted between June 2nd and August 23rd, 2020, using a global positioning system (GPS) to obtain mapping information on landslide locations (Figure 3). These points were utilized in the GIS context to produce a map of the landslide inventory. 64 landslide locations (70 percent) were used to build models and 27 landslide locations (30 percent) were used for validation (Oh & Pradhan 2011).

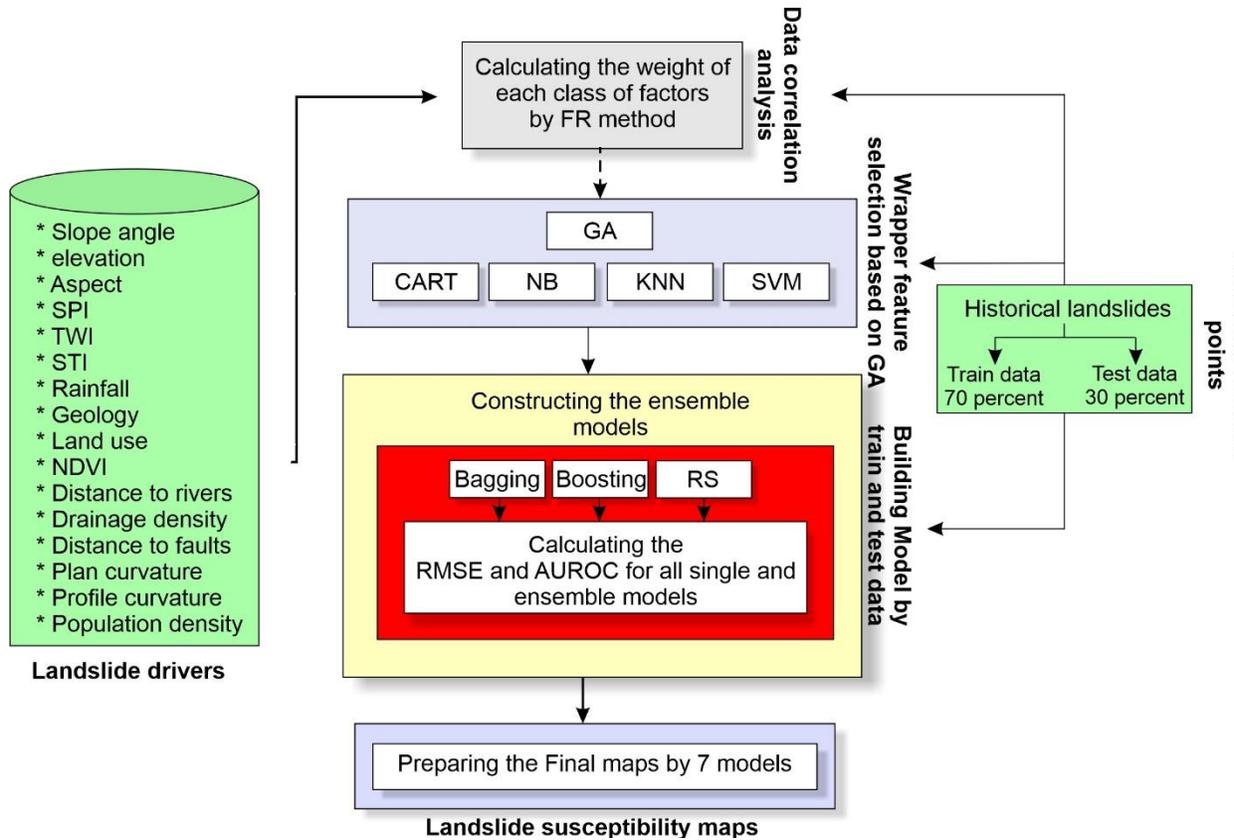


Figure 2. Methodological flow chart of the novel ensemble models for landslide susceptibility mapping

3.2. Landslide conditioning factors

Providing sufficient data and information from earlier landslides and significant landslide drivers is the first and most important step in landslide hazard zonation and analysis. A variety of factors including material, mineral composition, geological structure, water table, earthquakes, distance to faults, vibrations caused by construction machinery, and subsidence may result in slope instability (Rotaru et al. 2007). The present study investigates 16 landslide conditioning factors, including elevation, slope, aspect, land use, distance to faults, geology, plan curvature, profile curvature, topographic wetness index (TWI), sediment transport index (STI), stream power index (SPI), rainfall, distance to roads, distance to drainage, soil texture and normalized difference vegetation index (NDVI) that are amongst the most important causes of slope instability, as the significant landslide factors (Figure 3); and each factor is explained briefly in the following sections. These factors play significant role in the landslide occurrence in the study area.

3.2.1. Elevation

Elevation is a basic factor in most of the spatial modeling landslide susceptibility studies. The amount of precipitations, number of freezing days, and slope steepness usually depend on elevation (Du et al. 2020). High altitudes are believed to be more prone to landslides than low altitudes (Devkota et al. 2013). Rocks at higher altitudes are stronger than rocks at low altitudes or foothills (Chen et al. 2018b). Organic matter in high altitude soils, namely carbon and nitrogen, are responsible for dehydrogenase activity and have a positive impact on the landslide (Błońska et al. 2018). This DEM was obtained from the International Agriculture Research Consortium for Spatial Information Shuttle Radar Topographic Mission (SRTM) data (USGS 2020). The DEM data were processed using the ArcGIS 10.8 to The relevant topographical conditioning factors which include: Slope steepness, slope Aspect, plan curvature, profile curvature, topographic wetness index (TWI), sediment transport index (STI), and stream power index (SPI), (Beven 1997; Beven and Freer 2001).

3.2.2. Slope

In landslide hazard zonation methods, including knowledge-based and data driven approaches and in all scales, slope is regarded as a main factor (Dai et al. 2002; Fall et al. 2006). Slope steepness has a significant effect in occurrence of landslides (Chen et al. 2017). This factor also intensifies the effective role of the other factors in mass movement critically (Vergari et al 2011).

3.2.3. Aspect

Aspect is another factor with a significant effect on slope instability. Although using this factor in landslide zonation models is not as conventional as the slope steepness, in certain areas, there might be a remarkable temperature difference between slopes exposed to the sunlight and those not exposed to the sunlight. This temperature difference might be in the range from above to below the freezing point of water. In such cases, there will be a meaningful relation between distribution of slope instabilities and slope aspects; so that the slope instabilities are more likely to happen in northern slopes (not exposed to the sunlight) and less in the southern slopes (exposed to the sunlight).

3.2.4. Land use

Land use and human activities play a decisive role in changes in the environment. In many cases the land uses are positioned improperly, causing disorders and damages to the natural ecosystems, This is because many of the

infrastructure projects undertaken in the last two decades, such as road construction and new housing projects, were not based on sufficient geological studies. Also, many of the agricultural roads that reach the farms outside the area of the municipal authority area (outside the urban organization), did not take the adequate amount of planning and appropriate design. Improper manipulation of land in land uses like agriculture can increase the potential of landslide, because such land uses prepare soft and unstable soil formations. Figure 3.d. shows the land use map. It has become necessary for the municipalities to extend their authority to more areas than it is now, and to update the maps of land use provided by them to become more responsive to modern requirements such as modern infrastructure and new housing projects.

3.2.5. Distance to faults

Faults have significant effects on development or activation of the areas with landslide potential. Crushing and shearing in fault areas, penetration of water from these areas to the slopes, discontinuities appearing around the faults, and different level of erosion in the slopes are some of such effects.

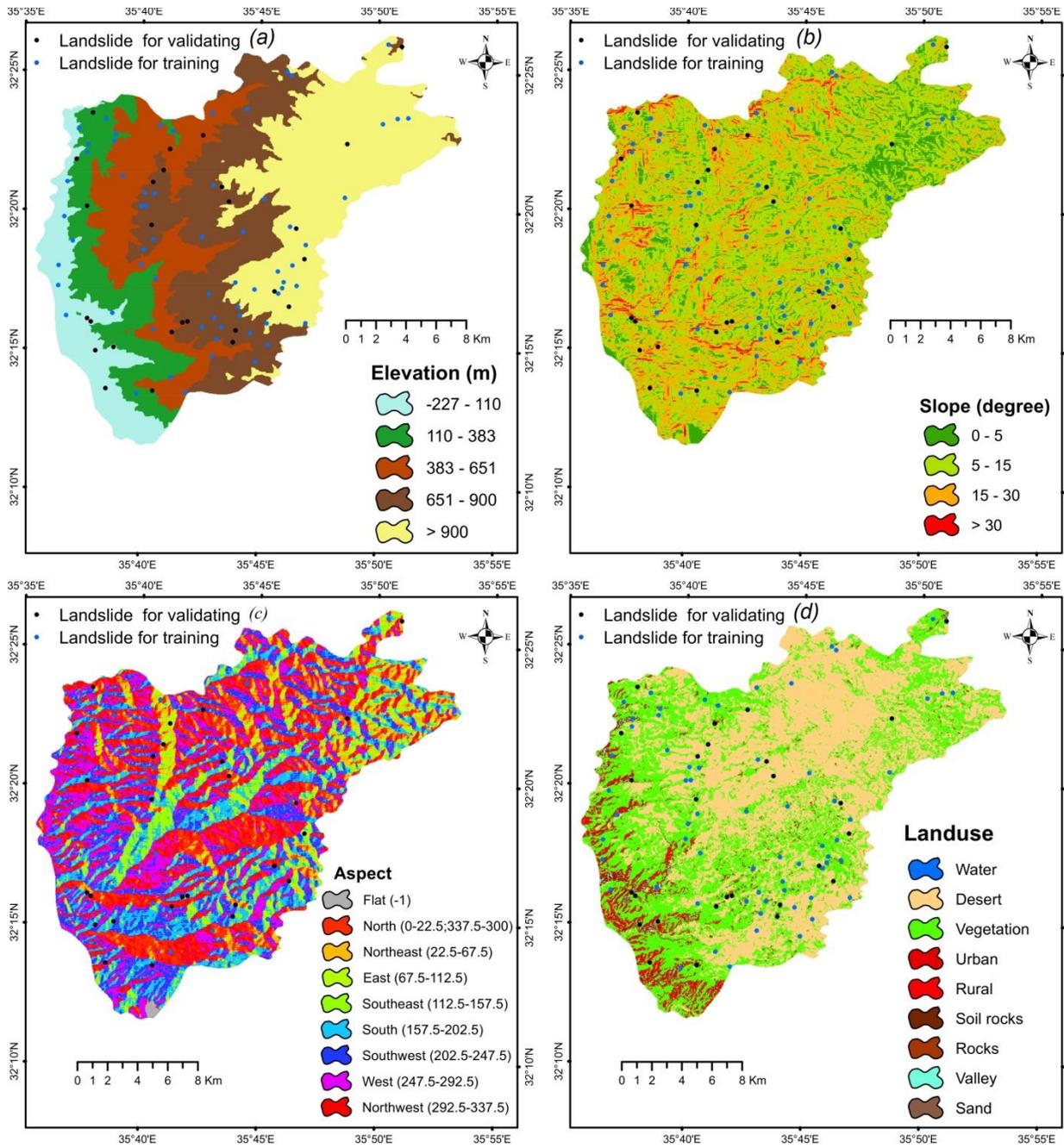


Figure 3. Conditioning factors for landslide susceptibility mapping

3.2.6. Geology

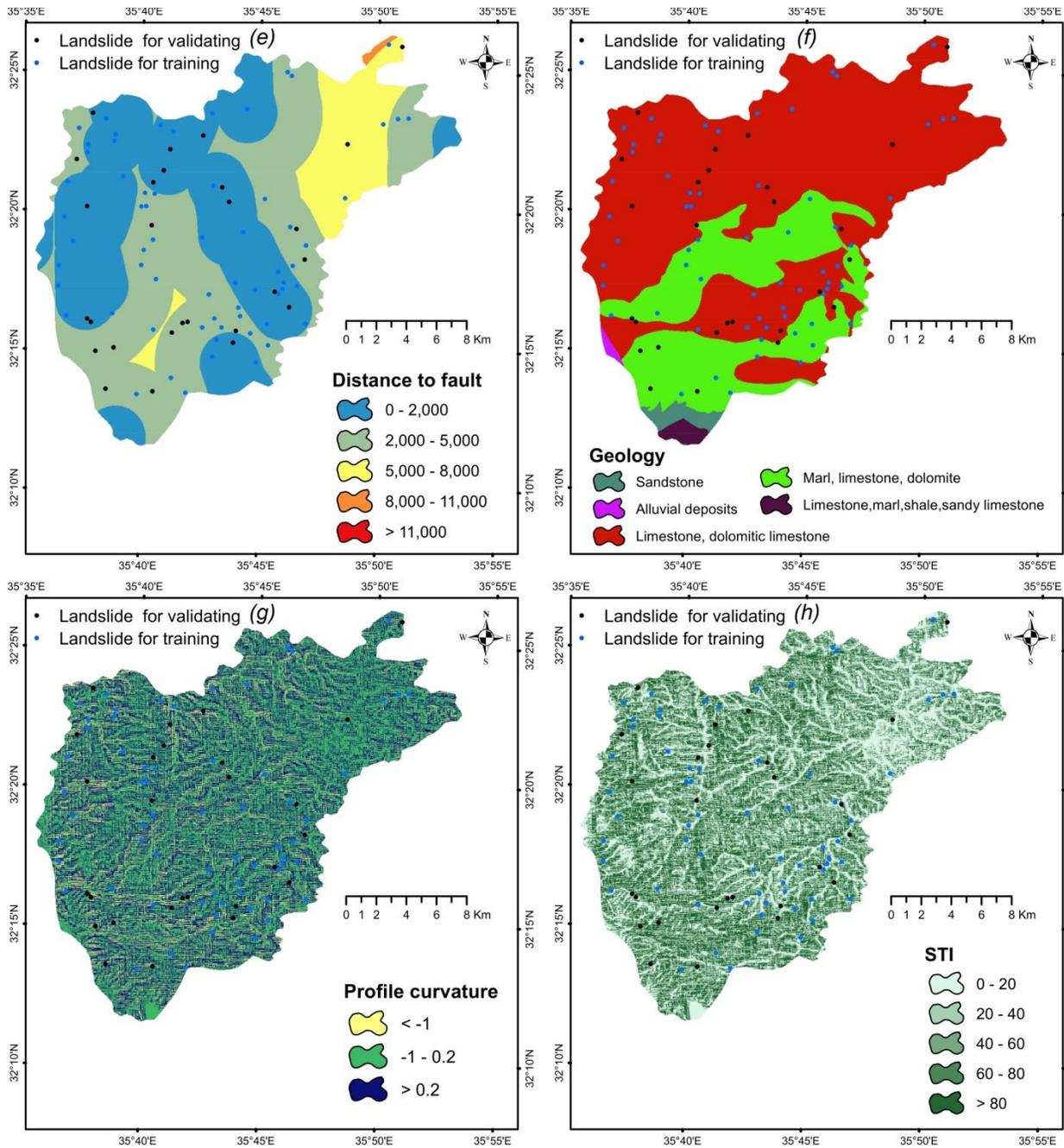


Figure 3. (Continued)

Considering the fact that geology and different geological structures cause different stability and strength of rock structures and diversity of soil textures, this factor is regarded as an important landslide driver. This map was extracted from 1:100,000 map of the Geological Survey and Mineral Explorations of Jordan, which was used after classification and digitalization.

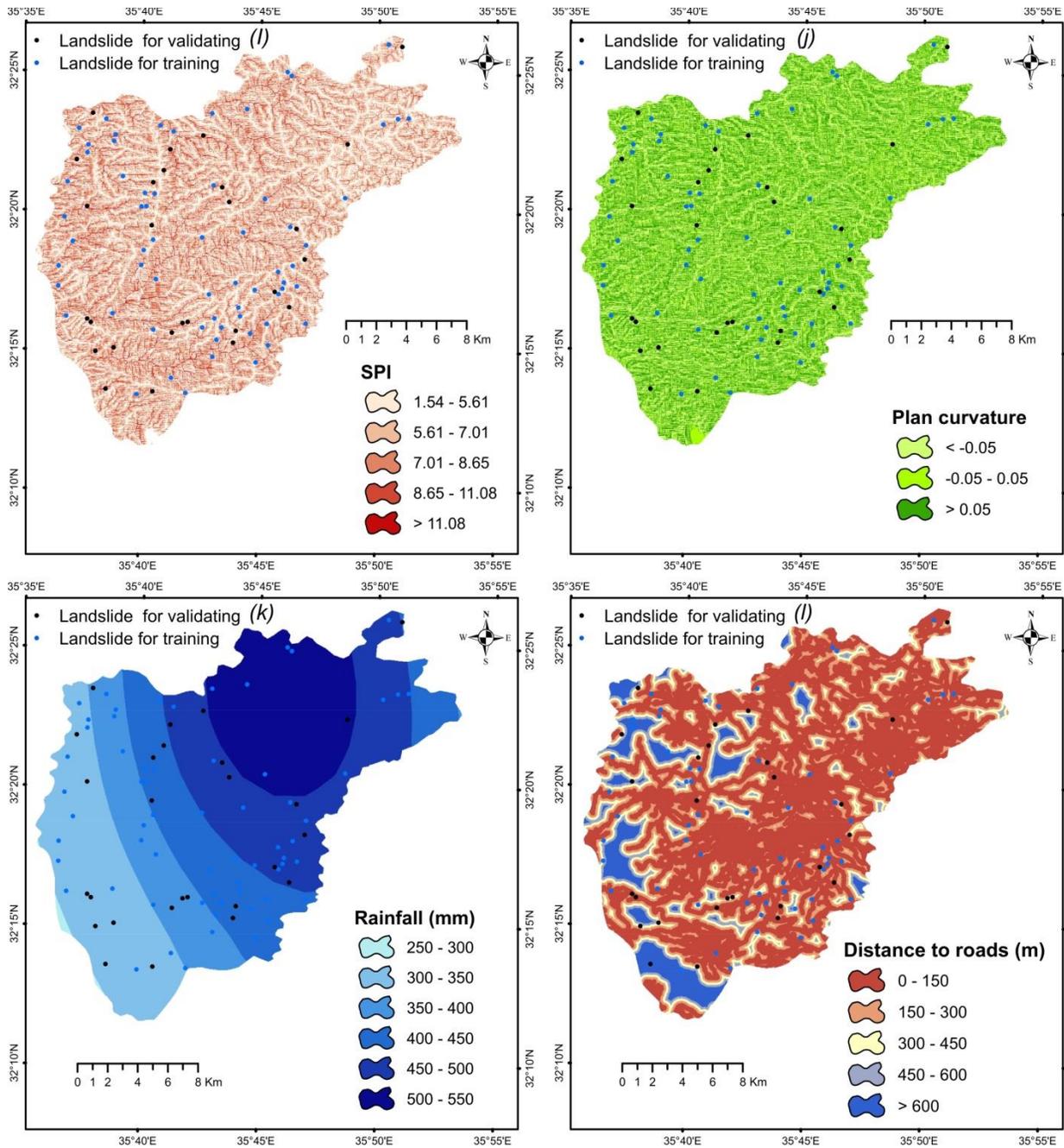


Figure 3. (Continued)

3.2.7. Plan and profile curvature

These two factors are the topographic factors used in many studies for landslide susceptibility mapping. Plan curvature is defined as the curvature of the contours on a topographic map and is classified into convex, concave, and flat categories. A number of studies have explored the relationship between these three categories and occurrence of landslides (Ohlmacher 2007; Saito, Nakayama & Matsuyama 2009).

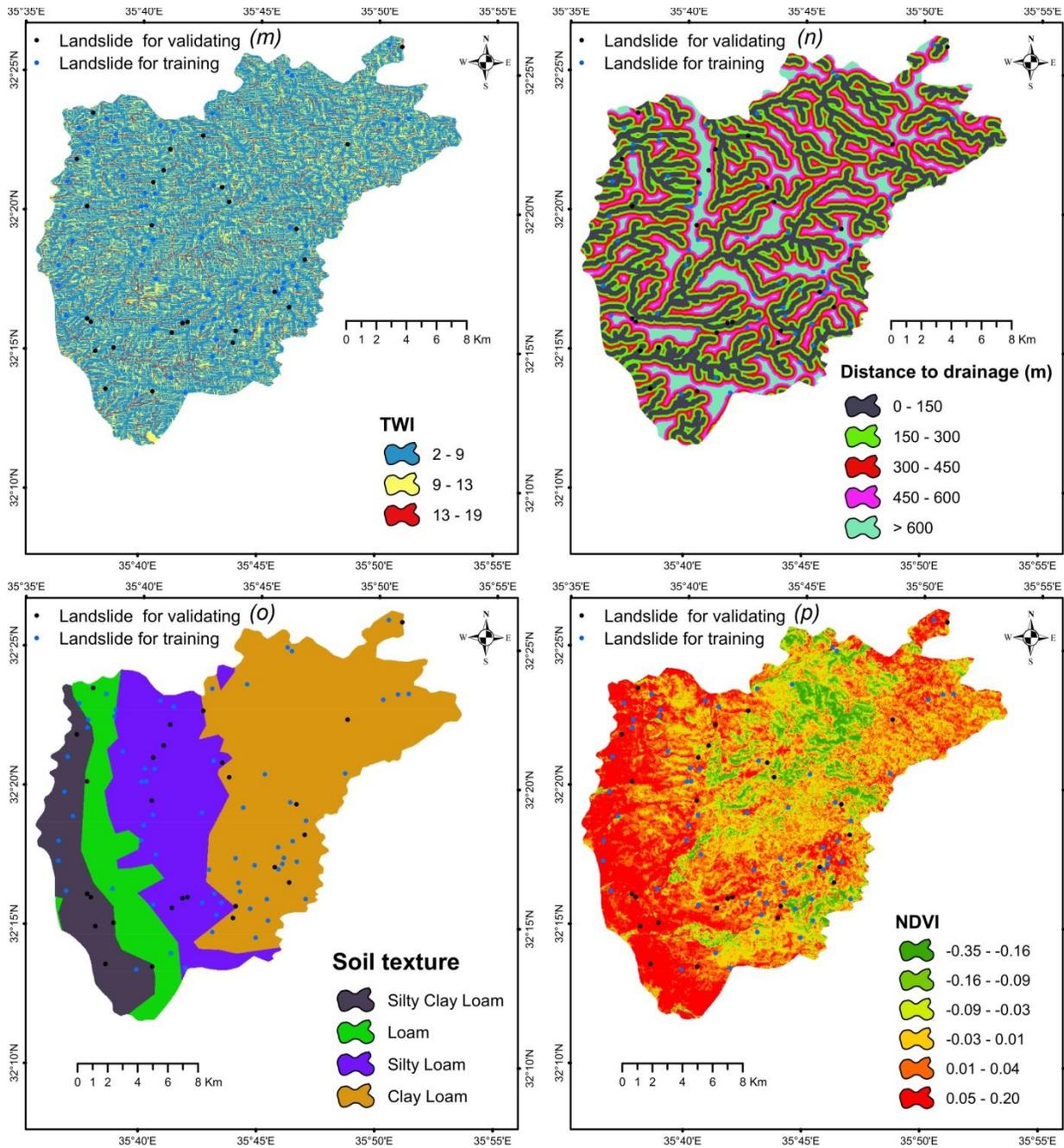


Figure 3. (Continued)

Profile curvature shows curvature of the land surface in the gradient direction. This factor can have negative and positive values. Negative values are indicative of upwardly convex surface of the land, and positive values are indicative of its upwardly concave surface. This factor has been used in the majority of landslide studies (Erener & Düzgün 2010; Saito et al. 2009).

3.2.8. TWI, SPI and STI

There are two groups of topographic attributes. The first group consists of primary topographic attributes, including slope and aspect. The second group consists of secondary topographic attributes including TWI, STI, and SPI indicators. TWI is used to show wetness conditions at the catchment scale and is calculated using the Eq. 1 (Moore et al 1991):

$$TWI = \ln\left(\frac{A_s}{\tan\sigma}\right) \quad (1)$$

SPI is used to investigate and describe the erosion potential of the terrain. This indicator is calculated using Eq. 2 (Wilson 2000):

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

STI is the result of erosion and sedimentation in the region, and is calculated using the Eq. 3 (Moore et al. 1993).

$$STI = \left(\frac{\text{Flow Accumulation} \times \text{cell size}}{22.13}\right)^{0.6} \cdot \left(\frac{\sin\sigma \times 0.01745}{0.0896}\right)^{1.3} \quad (3)$$

In Equations 1, 2, and 3 A_s is the specific area of the basin based on the total upstream area and σ is slope gradient in degrees. In order to calculate flow accumulation, first the flow direction map is calculated using DEM in GIS environment, and then the flow accumulation map is calculated using the flow direction map.

3.2.10. Rainfall

Atmospheric precipitations in forms of rain or snow play a significant role in formation of mass movements. It is therefore essentially important to know characteristics of the precipitations. Intrinsic conditions prone to landslide, such as type of the material, increase the possibility of landslide with increasing rainfall and humidity (Kayastha et al. 2013).

3.2.11. Distance to roads

In many places, development of road networks on the slopes requires removing the hills. On the other hand, the construction of roads on top of the slopes increases the overhead and mechanical driving force. As a result, the

probability of landslide occurrence increases. This factor, which is caused by the human activities, results in increased instability of slopes and potential of landslide (Juliev et al. 2019).

3.2.13. Distance to drainage

Flowing waters are amongst the major landslide factors in slopes. There are more waterways in less steep slopes, which results in an increased number of landslides. This is indicative of the significant role of waterways in formation of landslides (Lee & Sambath 2006; Pradhan, & Buchroithner 2010). In the steeper slopes, there are fewer waterways, as a result of which less mass movements happen. Waterways usually experience more landslides due to drainage of water and steep walls (Lee & Sambath 2006; Pradhan et al. 2010).

3.2.14. Soil texture

This factor is another important landslide factor (Temme 2021). Lots of studies are conducted on the role of soil texture in occurrence of landslides (Lee et al. 2017; Rahim et al. 2018). The studies show that soils with higher silt content are more likely to erode compared to clay and sandy soils (Fonseca et al. 2017). This factor has 4 classes including Silty Clay Loam, Loam, Silty Loam, and Clay Loam.

3.2.15. NDVI

This factor delineates green vegetation in the study area. NDVI is based on the idea that chlorophyll in the plant structure is capable of absorbing red radiations, and the mesophyll layer is capable of reflecting the infrared radiations (Goward, Markham, Dye, Dulaney, & Yang, 1991; Pettorelli, 2013). This index ranges between -1 for minimum and +1 for maximum photosynthesis activity. NDVI is defined by Eq. 4 (Goward et al. 1991):

$$NDVI = \frac{NIR - RED}{NIR + RED} \quad (4)$$

In this index, for the vegetation areas, red and near infrared (NIR) wavelengths are marked by high absorption and low reflection, respectively. Since this index is somehow indicative of changes in vegetation density, it can be used for modeling the landslides.

3.3. Spatial correlation between landslides and each class of factors

Having general knowledge of the number of occurred landslides in each class is a conventional step in landslide susceptibility mapping with FR as one of the most typical methods. FR is derived from Eq. 5 (Lee & Sambath, 2006):

$$FR = \frac{A/B}{C/D} \quad (5)$$

Where A is the number of landslide points in each class of conditioning factors, B is the total number of landslide points in the study area, C is the area of each class of conditioning factors and D is the area of the whole studied area.

3.4. Feature selection process using Genetic Algorithm (GA)

Feature selection is an important step in the modeling process through the machine learning algorithms (Balachandran et al., 2018). Various methods have been proposed so far for feature selection. In general, they can be classified as two main categories, i.e. filter and wrapper methods (Talavera 2005). The filter methods are independent of the machine learning algorithms and calculate the credibility of every feature by using the resultant criteria (Sánchez-Marono et al. 2007). The features are then sorted in terms of their ranks, and the features whose ranks are smaller than a predetermined threshold will be removed. These methods are very fast and also appropriate for high-dimensional data; however, they are less accurate than the wrapper methods in which the learning algorithm feedback is used for feature selection (Das 2001; Solorio-Fernández et al. 2020). In other words, the feature selection criterion is the classification algorithm accuracy. Hence, these methods are more accurate. The wrapper approach was adopted in this study for feature selection through the GA based on the DT, KNN, NB, and SVM algorithms. In fact, the accuracy of these four models was used as the criteria for selecting the effective features for modeling.

For feature selection, a population of solutions (i.e. a set of features) was first generated randomly. Each solution or chromosome is defined to consist of 16 genes, each of which indicates one feature or factor of the problem. These genes can be either 0 or 1 at random. If a gene is 0, it shows the absence of the corresponding feature; however, if a gene is 1, it shows the presence of the corresponding feature. After that, the modeling process was performed through the DT, NB, KNN, and SVM algorithms based on the features selected from each chromosome. The outputs of the models were then compared to the real values from the training set to determine the RMSE, which was used as the fitness function. Afterwards, the roulette wheel was used for selection in the genetic algorithm based on the problem solutions and the fitness values. The chromosomes entered into the second generation through selection and two-point

mutation. The algorithm continues, and the features with the best fitness values are selected as the optimal features. There were 20 chromosomes in each generation, whereas the number of generations was 200.

3.5. Decision Tree (DT)

DTs belong to the new generation of data mining techniques and have seen considerable development in the last two decades. These models can be used to extract knowledge from a database, or alternatively to create a predictive model (Kotsiantis,2013). Unlike artificial neural networks, the DTs can generate interpretable rules (Kotsiantis 2013). In fact, a DT describes its prediction as a set of rules. In addition, non-numerical (categorical) variables can also be used in these models. In this research, the Classification and Regression Tree (CART) model (Breiman et al. 1984) is employed to map the landslide susceptibility. The DT used in this research is based on three main principles (Breiman et al. 1984):

- A set of questions in the general form of $x < d$, where x is an independent variable and d is a constant, and each question is answered with yes or no.
- Selecting the best criterion for determining the best independent variables that can be used to create the tree branches.
- Providing a summary statistic for the final node

Various metrics can be used to create a branch and build the DT. However, several studies have shown that the Gini index is more suitable for most of the data set (Daniya et al. 2020). This index can be computed using Eq. 6 (Daniya et al. 2020).

$$Gini(t) = \sum_{i \neq j} P(w_i) \times P(w_j) \quad (6)$$

where $P(W_i)$ is the relative frequency of class i and $Gini(t)$ is the gini index at node i . The tree grows until the maximum purity (i. e., minimum impurity index) at the leaf nodes is achieved.

3.6. K-nearest neighbors (KNN)

The KNN algorithm finds the k nearest samples and predicts the value of the desired sample based on their value (Kramer 2013). These k samples are the k most similar samples to the examined sample. The mechanism of this algorithm is quite simple. It relies on the distance between samples to find the k nearest neighbors. Once the k nearest samples are determined, the algorithm can simply return a weighted combination of their values as the model's prediction for the desired sample (Kramer 2013).

The algorithm consists of the following steps (Kramer 2013):

- 1) Preparing the training samples (X_j)
- 2) Determining the neighborhood size (k)
- 3) Calculating the distance of the input sample (X_i) to each category or group of samples using the Euclidean distance presented in Eq. 7:

$$d(x_i, x_j) = \sqrt{\sum_{a \in A} d_a(x_i, x_j)^2} \quad (7)$$

- 4) Determining the state of the input sample and the class that it belongs to.

3.7. Support Vector Machine (SVM)

SVM is a supervised machine learning algorithm and is usually used to solve classification and regression problems. It is a relatively new algorithm and has shown better performance compared to the previous classification algorithms, including the perceptron neural networks, in recent years. The equation that determines the optimal hyperplane separating the different classes is solved using quadratic programming (QP), which is a well-known technique for solving constrained equations (Vapnik et al. 1995). To handle the data complexity, the model uses special functions called kernels to transfer the data to a higher-dimensional space, in which the transferred data can be linearly separated (Vapnik et al. 1995). Radial basic function (RBF) is the most well-known and commonly used kernel function and is reported in various works as the optimal option (Patle & Chouhan 2013).

3.8. Naïve Bayes (NB)

NB is a simple, widely used model that has been used in various domains. This method can be used in problems where sample X can be described as conjugation of conditionally independent features (Leung 2007). Bayesian learning, which is founded on Bayesian probability theory, allows us to determine the posterior probability based on the prior probabilities (Kelly & Kolstad 1999):

$$P(h|D) = \frac{P(D|h)P(h)}{P(D)} \quad (8)$$

$P(h)$ is the initial probability of *the* h hypothesis before observing the training sample D (prior probability). If this probability is not known, one can assign the same probability to all hypotheses. $P(D)$ is the prior probability of observing the training sample D . $P(D|h)$ is the probability of observing the training sample D given that the h hypothesis is true.

3.9. The Proposed Bagging, Boosting and Random Subspace Schemes Ensemble models

In recent years, ensemble models have been used extensively by researchers. These models combine the outputs of multiple models and utilize their various capabilities to improve modeling efficiency. Bagging is one of the techniques that can be used to build an ensemble model. To create an ensemble bagging model, one can combine the bootstrap sampling method with a single model (for example, a decision tree) or use multiple models built using the same training dataset (Bühlmann 2012). The current study has used the second approach to combine the results of five different models, namely, DT, SVM, KNN and NB, to create an ensemble model. The averaging method is used to combine the outputs of these models.

In the boosting method, the Classifier 1 is built by a group of samples from main data sets. Then, the samples that were mis-classified are given to Classifier 2 to perform the learning process better (Bühlmann 2012); therefore, this method acts like the bagging method; however, the mis-classified samples are entered into the next model in each step to better perform the learning process on them. Finally, the RS model operates on features instead of samples. In other words, models are developed from features through random selection (Tian & Feng 2020).

4. Results

After the necessary data are prepared and created, the FR method was adopted to identify important classes in each of the designated factors. Table 1 shows the results of this method. The important classes of every factor can be seen clearly. Before the modeling process is performed, the optimal features were obtained using GA and the KNN, DT, NB, and SVM models based on the wrapper approach. Table 2 reports the feature selection results. Accordingly, the outputs of all feature selection methods show the presence of slope, STI, distance to drainage, and NDVI factors. The TWI, plan and profile curvature, and geology factors are also present in at least three out of four feature selection algorithms. Moreover, the distance to faults and aspect factors were selected as important factors in only one out of four algorithms. Figures 4 to 7 show the status of each generation of the GA based on the RMSE. Accordingly, the best value of the fitness function (the best RMSE) and the mean fitness were obtained 0 and 0.0049, respectively, from the GA-DT algorithm after 200 generations. At the same time, these values were reported 0.3558 and 0.3586; 0.4256 and 0.4269; and 0.3768 and 0.3896 for the GA-KNN, GA-NB, and GA-SVM algorithms respectively.

Table 1. Correlation analysis for each class of variables

Factors	Classes	No. of Pixels	No. of landslides	FR weights
Elevation (m)	< 110	51357	7	0.98
	110 – 383	61088	8	0.95
	383 – 651	84365	10	0.85
	651 – 900	122128	22	1.30
	> 900	143149	17	0.86
Aspect	Flat	1564	0	0
	North	68854	3	0.31
	North east	31066	2	0.46
	East	26478	6	1.64
	Southeast	33880	3	0.64
	South	73606	15	1.47
	Southeast	750567	10	1.02
	West	81947	13	1.15
Slope (degree)	Northwest	74125	12	1.17
	0 – 5	64728	8	0.89
	5 – 15	256933	29	0.81
	15 – 30	127871	23	1.30
Rainfall (mm)	> 30	12555	4	2.30
	250 – 300	106671	12	0.81
	300 – 350	65119	12	1.33
	350 – 400	1296	0	0
	400 – 450	86638	17	1.42
	450 – 500	109552	18	1.19
NDVI	500 – 550	92811	5	0.39
	-0.35 – -0.16	17720	0	0
	-0.16 – -0.09	29280	1	0.25
	-0.09 – -0.03	53071	2	0.27
	-0.03 – 0.01	104827	10	0.69
	0.01 – 0.04	131959	22	1.20
TWI	0.04 – 0.20	125230	29	1.67
	< 9	124682	45	2.61
	9 – 13	230849	15	0.47
Distance to rivers (m)	>13	106556	4	0.27
	0 – 150	144332	19	0.95
	150 – 300	116568	16	0.99
	300 – 450	93536	14	1.08
	450 – 600	63418	11	1.25
Distance to faults (m)	> 600	44233	4	0.65
	0 – 2000	201136	32	1.15
	2000 – 5000	201277	29	1.04
	5000 – 8000	58360	3	0.37
	8000 – 11000	1314	0	0
STI	> 110000	0	0	0
	0 – 20	162868	22	0.98
	20 – 40	61308	8	0.94
	40 – 60	51842	6	0.84
	60 – 80	38230	7	1.32
Profile curvature	> 80	147839	21	1.03
	< -1	48665	6	0.89
	-1 – 0.2	236966	30	0.91
	> 0.2	176456	28	1.15

Plan curvature	< -0.05	207614	25	0.87
	-0.05 – 0.05	45426	5	0.79
	> 0.05	209047	34	1.14
SPI	1.54 – 5.61	115911	12	0.75
	5.61 – 7.01	148493	30	1.46
	7.01 – 8.65	98893	12	0.88
	8.65 – 11.08	63305	9	1.03
	> 11.08	35485	1	0.20
Distance to roads (m)	0 – 150	277073	37	0.96
	150 – 300	89637	15	1.21
	300 – 450	42651	8	1.35
	450 - 600	22674	1	0.32
	> 600	30052	3	0.72
LULC	Water	10	0	0
	Desert	225729	17	0.54
	Vegetation	200719	36	1.29
	Urban	4052	0	0
	Rural	20820	7	2.43
	Soil rocks	5949	0	0
	Rocks	1549	1	4.66
	Valley	2275	1	3.17
Geology	Sand	984	2	14.68
	Sandstone	6384	0	0
	Alluvial deposits	1333	0	0
	Limestone, dolomitic limeston	326237	47	1.04
	Marl, limestone, dolomite	124099	17	1.00
	Limestone, marl, shale, sandy limestone	4034	0	0
Soil texture	Silty Clay Loam	60930	8	0.95
	Loam	60555	7	0.83
	Silty Loam	118497	19	1.16
	Clay Loam	222105	30	0.98

Table 2. Optimal features selected by GA-KNN, GA-NB, GA-SVM and GA-DT models

Explanatory variables	GA-KNN	GA-NB	GA-SVM	GA-DT
Elevation	-	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	-
Slope aspect	-	-	<input checked="" type="checkbox"/>	-
Slope	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Land use	<input checked="" type="checkbox"/>	-	-	<input checked="" type="checkbox"/>
TWI	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	-	<input checked="" type="checkbox"/>
NDVI	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Distance to drainage	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Distance to roads	-	-	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Soil texture	<input checked="" type="checkbox"/>	-	<input checked="" type="checkbox"/>	-
Distance to faults	-	-	<input checked="" type="checkbox"/>	-
SPI	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Rainfall	-	-	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Plan curvature	<input checked="" type="checkbox"/>	-	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Profile curvature	-	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
STI	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Geology	<input checked="" type="checkbox"/>	-	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>

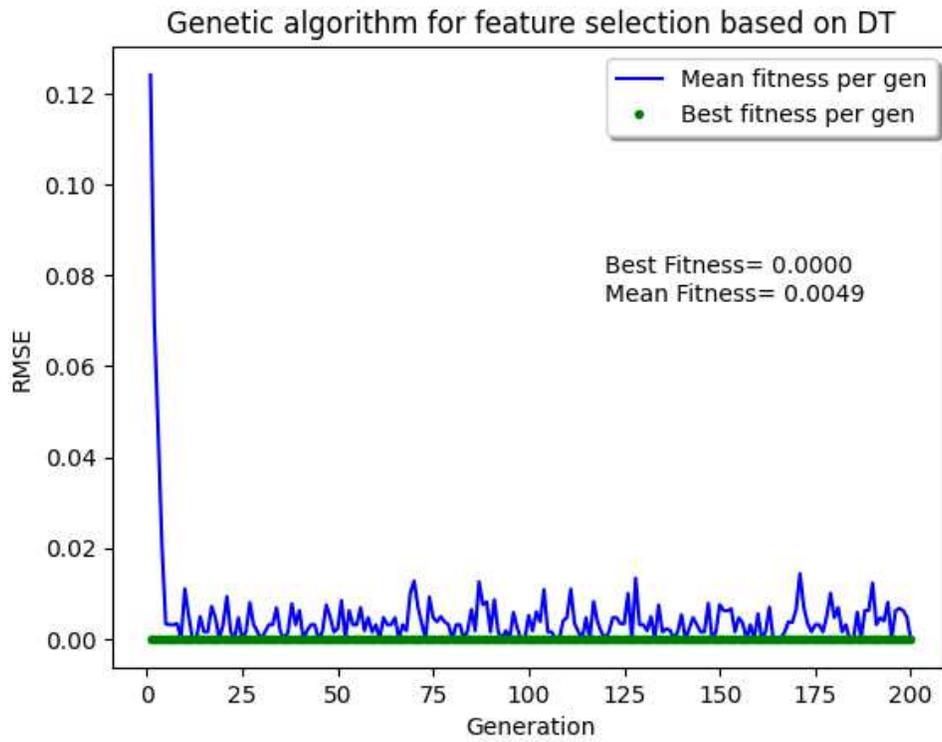


Figure 4. Convergence plot of DT based GA for feature selection

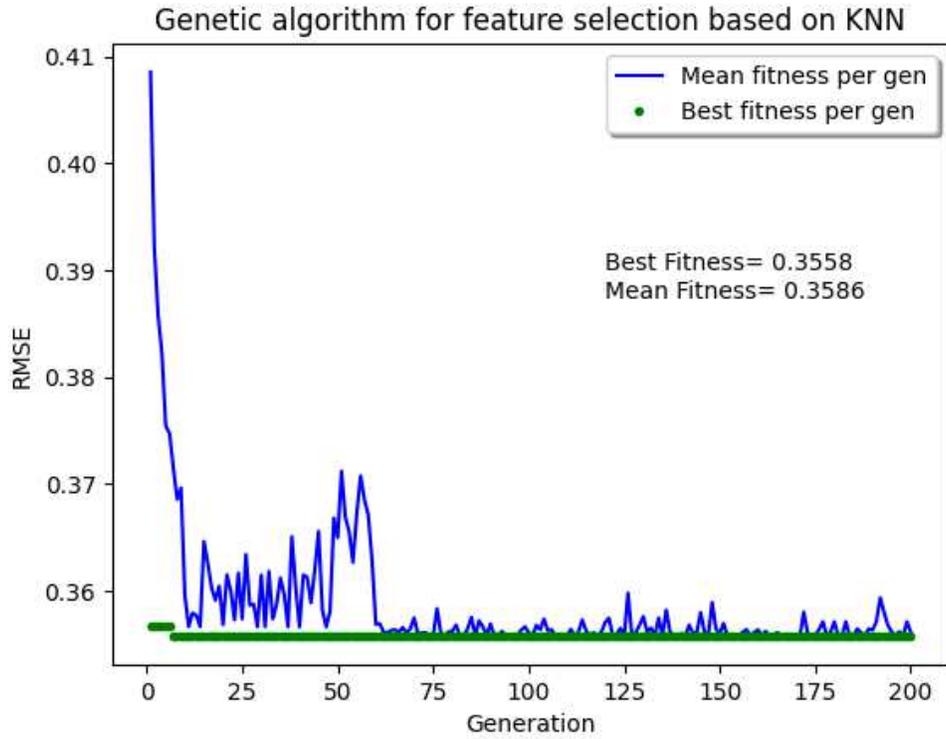


Figure 5. Convergence plot of KNN based GA for feature selection

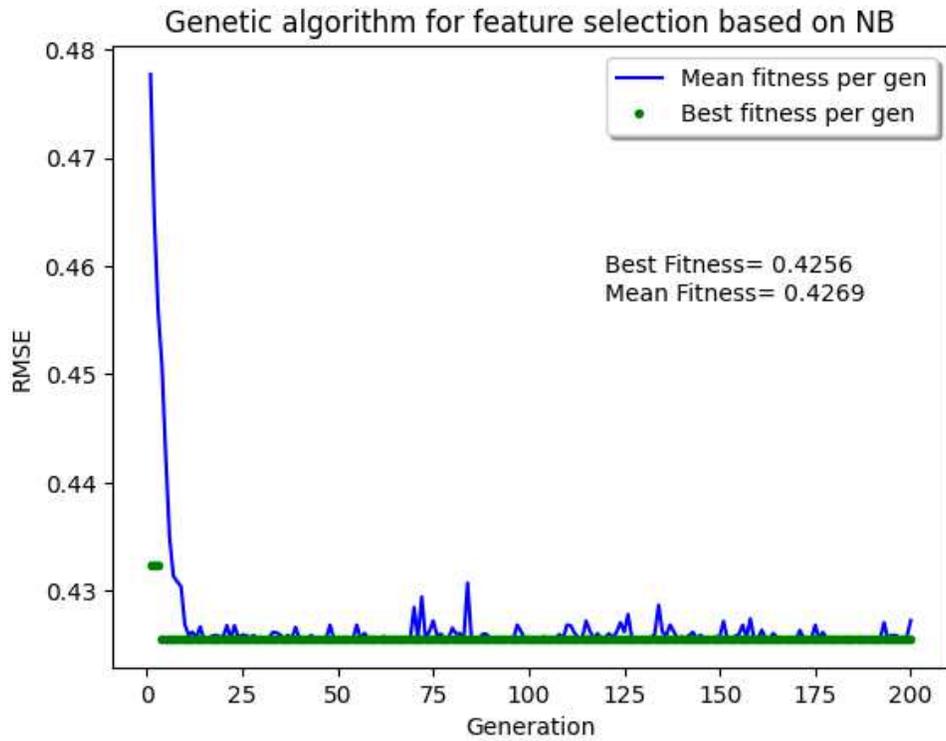


Figure 6. Convergence plot of NB based GA for feature selection

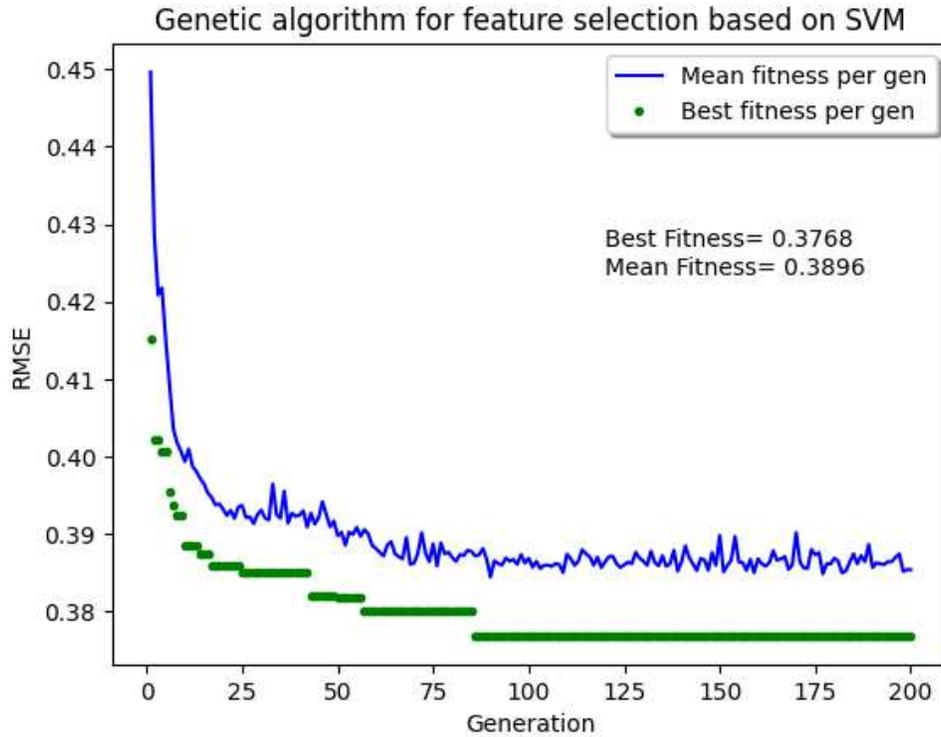


Figure 7. Convergence plot of SVM based GA for feature selection

After the optimal features are found for the four models, the ensemble models were developed. Figure 8 demonstrates the ROC curves of all models in the training and test phases. Accordingly, the ensemble algorithms obtained the best AUROC in the learning phase, and the boosting, RS, and bagging models had the highest AUROC values, whereas the NB, SVM, KNN, and DT algorithms obtained the lowest AUROC values, respectively. In the validation phase, the ensemble models obtained higher accuracies than the single models. In other words, the highest values of AUROC resulted from the RS, boosting, and bagging models, respectively; therefore, the ensemble models achieved desired accuracy by integrating less accurate models.

After these models were validated, all samples of the study area were entered into the seven developed models based on the selected features in order to prepare landslide susceptibility maps. The outputs of these models were then produced in the form of landslide susceptibility

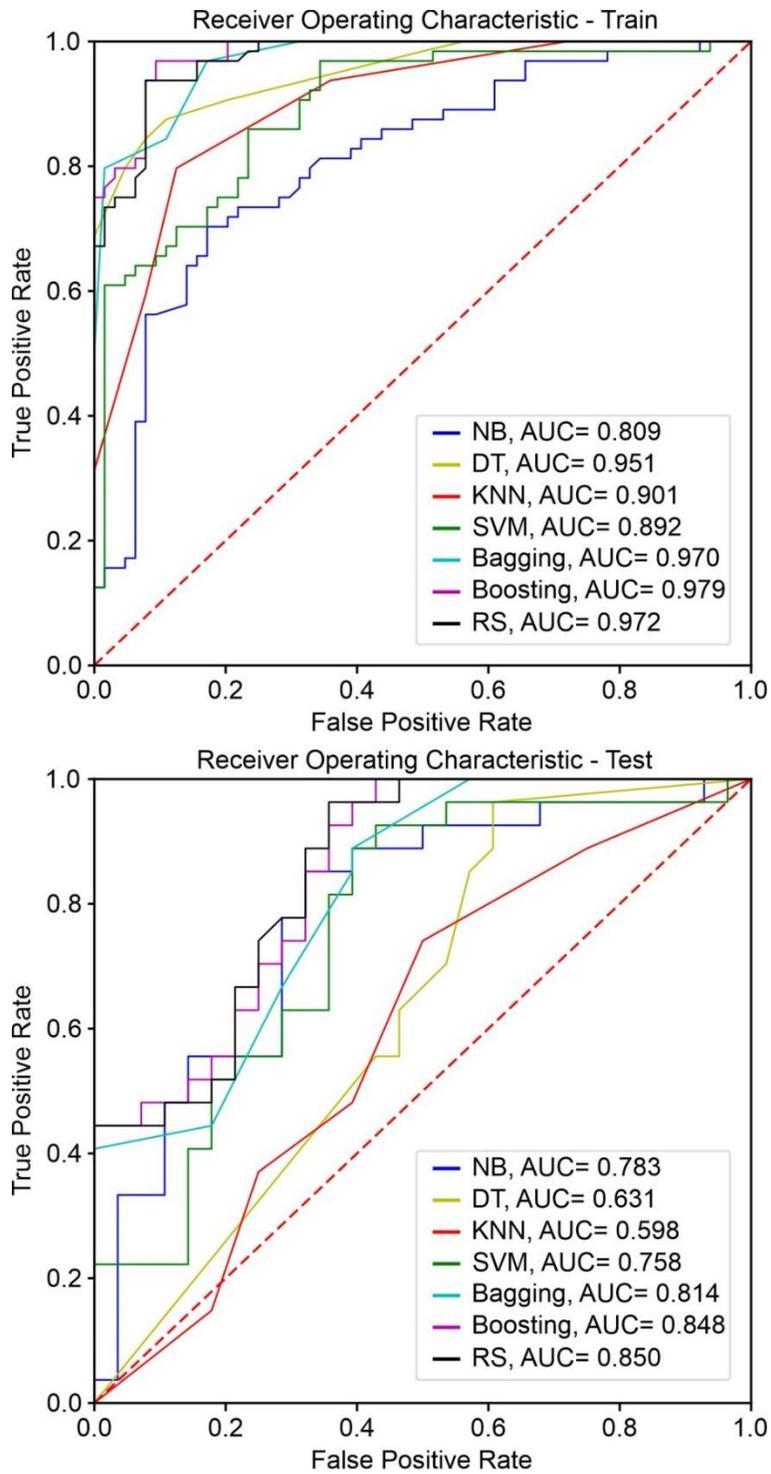


Figure 8. The ROC curves for seven models in the train and test phase

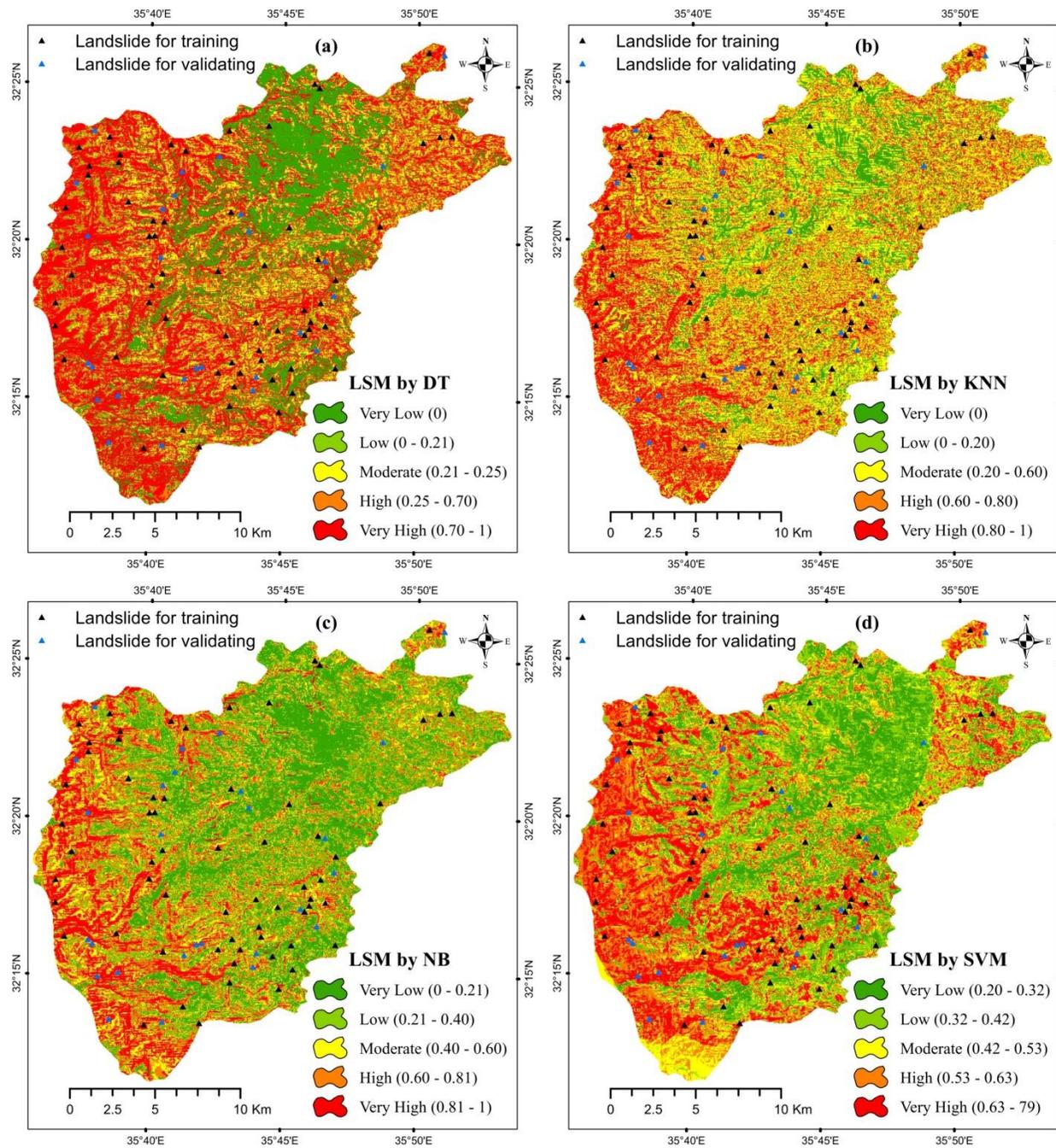


Figure 9. Landslide susceptibility maps obtained from the single and ensemble models

maps in GIS environment. These maps were classified as very high, high, moderate, low, and very low categories based on the natural break method (Figure 9). Also, figure 9 shows the approximate percentage of each class based on the seven adopted models. Nearly 50% of the entire region can be classified as

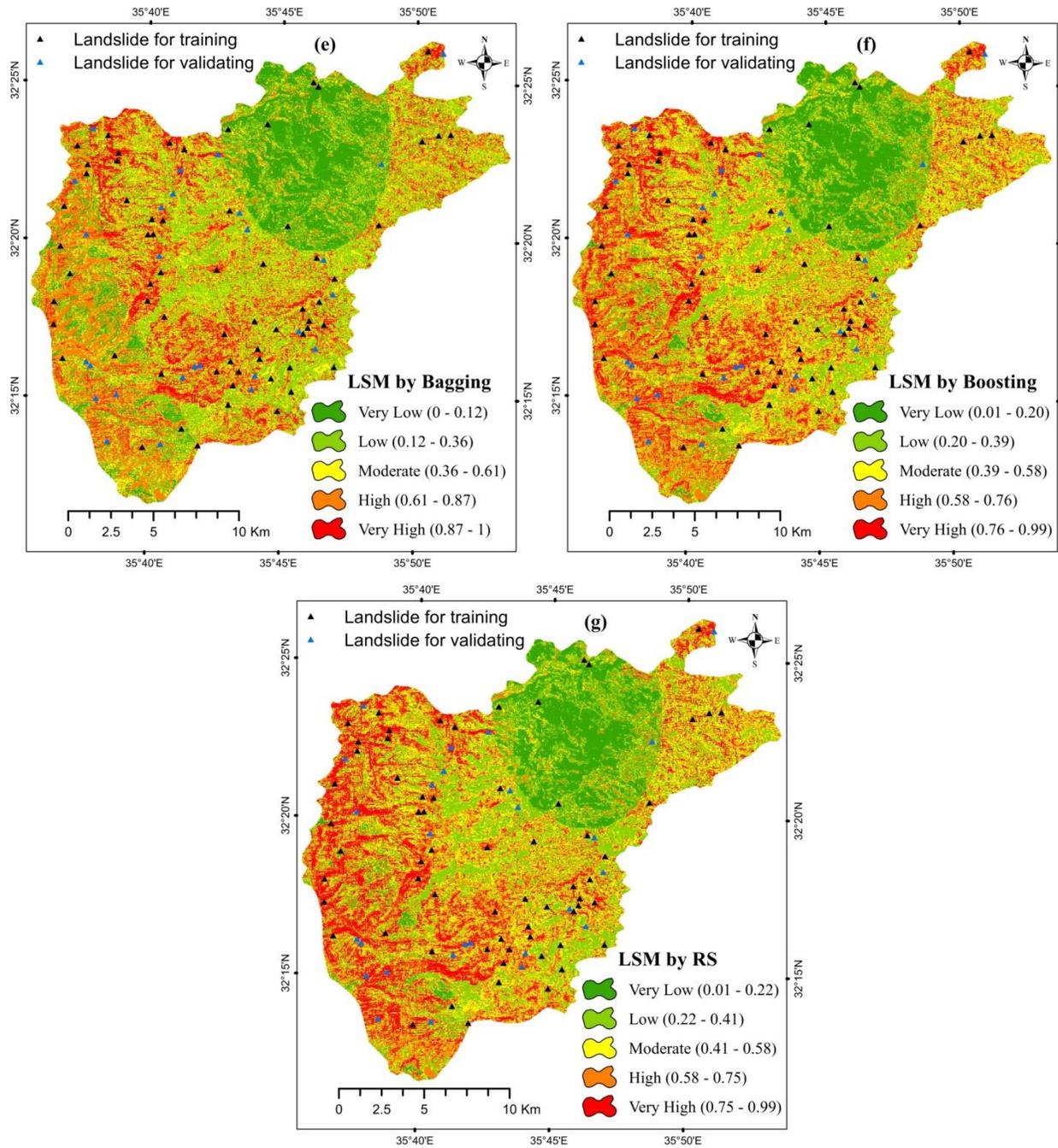


Figure 9. (Continued)

very high and high categories by using the best model, i.e. RS. These two classes are mostly observed in the western half of the region.

5. Discussion

Landslides occur due to various natural and man-made factors where each factor plays a key role in landslide occurrence. This study used 16 conditioning factors for landslide susceptibility mapping in Ajloun, Jordan. The first section discusses the factors which are present at least in 3 of 4 feature selection models used to select optimal factors. It should be noted that when the factors are evaluated independently and without considering other factors, it may appear that the studied independent factor leaves a significant impact on modeling, whereas this may be completely different when all factors are evaluated as whole. For example, some studies consider distance to fault as an important factor of landslide susceptibility mapping while this was selected as an effective factor in none of the four feature selection algorithms. A total of 8 features, namely slope, NDVI, plan curvature, profile curvature, geology, distance to drainage, STI and TWI were selected at least in 3 feature selection algorithms as important optimal modeling factors. According to results, almost 81% of mass movement occurred in the slope classes of 5 to 30 degree. Instability of slope, taking the form of slope failure, accelerates due to sliding but decreases after a certain slope, and the cause behind this lies in the hardness of the formations that constitute the slope and also due to damages that occurred during soil formation on the steeper parts of the slope. This hypothesis acknowledges the results obtained by Ahmadi et al., (2011). Out of various geological formations, the ones which are prone to landslides are those that comprise limestone, dolomitic limestone and Marl, and dolomite units. Such formations, being highly porous, increase instability of the slope due to their water permeability, thus causing movement of mass along the direction of the slope. Canals are also responsible for landslides, as the slope bases experience piping phenomenon due to the presence of canals, causing disturbances in the stabilizing forces that hold the slopes, increasing the risk of landslides. Also, it has been found that, 94% of landslides have occurred within a distance of less than 600 meters from canals. Highest landslide sensitivity is depicted by higher values of TWI and STI, where TWI gives the value of the content of moisture in the soil, whereas, increase in STI value indicates the accumulation of higher quantity of moisture in downstream slope directions, which influences landslide occurrence. In this study, the value of TWI ranges between 2 and 1, showing highest sensitivity to landslides, while also agreeing with studies conducted by Ahmadi et al. (2019). This study indicated that the highest number of landslides occurred in 20-0 and <80 classes in STI factor, as inferred by Huwang et al. (2018). Also, plan Curvature and profile curvature are land curvature factors which are two other vital conditioning factors.

Selecting effective features is among the most important activities in developing machine learning and data mining models. Regarding feature selection, some factors have no effects on the final output and produce random values for

every sample. At the same time, some factors affect the model output and cannot be considered among the features having no effects on the modeling process. Many studies have proven that it is fundamentally important to select relevant and necessary features in the preprocessing step. Therefore, feature selection was performed through all models in this study, and the effective features were selected through the genetic algorithm. Hence, researchers and model developers should take this step extremely seriously in the modeling process, before which it should be performed as an essential preprocessing step.

In recent years, many studies have been conducted on the ensemble classifiers to develop a final classifier. According to various studies, the resultant classifiers are more accurate than the initial classifiers. This study adopted bagging, boosting, and random subspace approaches to develop ensemble models. The KNN, NB, DT, and SVM algorithms were also employed to develop these models. This study drew a comprehensive comparison between these four models and the three ensemble models. The accuracy of every ensemble model was tangibly higher than the single models, and the RS model was more accurate than the other three ensemble models. The same finding was also obtained in other studies of spatial modeling. Another study used ensemble, bagging, and boosting models as well as the random forest to classify the land cover and compared them with the DT model. The results of ensemble models were more accurate than that of the DT model. Regarding hazard mapping, the ensemble algorithms were more accurate than the single models.

6. Conclusion

This paper conducted a comparative study using three ensemble approaches, namely bagging, boosting and RS constructed with four well-known machine learning algorithm including DT, KNN, NB and SVM for landslide susceptibility mapping in the Ajloun province of Jordan. Although various landslide zonation techniques and methods have been developed in recent two decades, ensemble models have been adopted and compared in fewer studies. This study selected effective features based on GA and DT, NB, SVM and KNN algorithms. The results of these algorithms showed that 8 factors, namely slope, plan curvature, profile curvature, TWI, STI, distance to drainage, geology, and NDVI are the main drivers of landslide in the study area. AUROC was used to compare the models. The value of this index was higher in all three ensemble models than individual ones, and the three models showed a desired modeling accuracy. Landslide susceptibility maps are considered as a beneficial tool in land use planning, disaster management

and hazard risk reduction. This study showed that ensemble models can serve as landslide susceptibility mapping models with a desired accuracy by integrating low-accurate individual models.

7. References

- Arabameri, A., Pradhan, B., Rezaei, K., Lee, S., & Sohrabi, M. (2020). An ensemble model for landslide susceptibility mapping in a forested area. *Geocarto International*, 35(15), 1680-1705.
- Arabameri, A., Rezaei, K., Pourghasemi, H. R., Lee, S., & Yamani, M. (2018). GIS-based gully erosion susceptibility mapping: a comparison among three data-driven models and AHP knowledge-based technique. *Environmental Earth Sciences*, 77(17), 1-22.
- Balachandran, P. V., Xue, D., Theiler, J., Hogden, J., Gubernatis, J. E., & Lookman, T. (2018). Importance of feature selection in machine learning and adaptive design for materials *Materials discovery and design* (pp. 59-79): Springer.
- Bardi, F., Raspini, F., Frodella, W., Lombardi, L., Nocentini, M., Gigli, G., . . . Casagli, N. (2017). Monitoring the rapid-moving reactivation of Earth flows by means of GB-InSAR: the April 2013 Capriglio Landslide (Northern Apennines, Italy). *Remote Sensing*, 9(2), 165.
- even K (1997) TOPMODEL: a critique. *Hydrol Process* 11:1069–1085
- Beven K, Freer J (2001) A dynamic topmodel. *Hydrol Process* 15:1993–2011
- Boateng, E. Y., Otoo, J., & Abaye, D. A. (2020). Basic Tenets of Classification Algorithms K-Nearest-Neighbor, Support Vector Machine, Random Forest and Neural Network: A Review. *Journal of Data Analysis and Information Processing*, 8(4), 341-357.
- Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1984). *Classification and regression trees*: CRC press.
- Brown, G. (2010). Ensemble Learning. *Encyclopedia of machine learning*, 312, 15-19.
- Bühlmann, P. (2012). Bagging, boosting and ensemble methods *Handbook of computational statistics* (pp. 985-1022): Springer.
- Chen, W., Yan, X., Zhao, Z., Hong, H., Bui, D. T., & Pradhan, B. (2019). Spatial prediction of landslide susceptibility using data mining-based kernel logistic regression, naive Bayes and RBFNetwork models for the Long County area (China). *Bulletin of engineering geology and the environment*, 78(1), 247-266.
- Cui, Y., Cheng, D., Choi, C. E., Jin, W., Lei, Y., & Kargel, J. S. (2019). The cost of rapid and haphazard urbanization: lessons learned from the Freetown landslide disaster. *Landslides*, 16(6), 1167-1176.
- Cutter, S. L. (1996). Vulnerability to environmental hazards. *Progress in human geography*, 20(4), 529-539.
- Dai, F., Lee, C., & Ngai, Y. Y. (2002). Landslide risk assessment and management: an overview. *Engineering Geology*, 64(1), 65-87.
- Daniya, T., Geetha, M., & Kumar, K. S. (2020). Classification And Regression Trees with Gini Index. *Advances in Mathematics: Scientific Journal*, 9(10), 8237-8247.
- Das, S. (2001). *Filters, wrappers and a boosting-based hybrid for feature selection*. Paper presented at the Icml.
- Dikshit, A., & Satyam, D. N. (2018). Estimation of rainfall thresholds for landslide occurrences in Kalimpong, India. *Innovative Infrastructure Solutions*, 3(1), 1-10.
- Dolidon, N., Hofer, T., Jansky, L., & Sidle, R. (2009). Watershed and forest management for landslide risk reduction *Landslides—disaster risk reduction* (pp. 633-649): Springer.
- Dou, J., Yunus, A. P., Bui, D. T., Merghadi, A., Sahana, M., Zhu, Z., . . . Pham, B. T. (2019). Assessment of advanced random forest and decision tree algorithms for modeling rainfall-induced landslide susceptibility in the Izu-Oshima Volcanic Island, Japan. *Science of the Total Environment*, 662, 332-346.
- Du, J., Glade, T., Woldai, T., Chai, B., & Zeng, B. (2020). Landslide susceptibility assessment based on an incomplete landslide inventory in the Jilong Valley, Tibet, Chinese Himalayas. *Engineering Geology*, 270, 105572.
- Erener, A., & Düzgün, H. S. B. (2010). Improvement of statistical landslide susceptibility mapping by using spatial and global regression methods in the case of More and Romsdal (Norway). *Landslides*, 7(1), 55-68.
- Fall, M., Azzam, R., & Noubactep, C. (2006). A multi-method approach to study the stability of natural slopes and landslide susceptibility mapping. *Engineering Geology*, 82(4), 241-263.
- Fonseca, L. d. A. M., Lani, J. L., Santos, G. R. d., Ferreira, W. P. M., & Santos, A. M. R. T. (2017). Variability in soil physical properties in landslide-prone areas. *Acta Scientiarum. Agronomy*, 39(1), 109-118.

- Goward, S. N., Markham, B., Dye, D. G., Dulaney, W., & Yang, J. (1991). Normalized difference vegetation index measurements from the Advanced Very High Resolution Radiometer. *Remote sensing of environment*, 35(2-3), 257-277.
- Huang, F., Yin, K., Huang, J., Gui, L., & Wang, P. (2017). Landslide susceptibility mapping based on self-organizing-map network and extreme learning machine. *Engineering Geology*, 223, 11-22.
- Huang, Y., & Zhao, L. (2018). Review on landslide susceptibility mapping using support vector machines. *Catena*, 165, 520-529.
- Hunter, L. M. (2005). Migration and environmental hazards. *Population and environment*, 26(4), 273-302.
- Juliev, M., Mergili, M., Mondal, I., Nurtaev, B., Pulatov, A., & Hübl, J. (2019). Comparative analysis of statistical methods for landslide susceptibility mapping in the Bostanlik District, Uzbekistan. *Science of the Total Environment*, 653, 801-814.
- Kamp, U., Growley, B. J., Khattak, G. A., & Owen, L. A. (2008). GIS-based landslide susceptibility mapping for the 2005 Kashmir earthquake region. *Geomorphology*, 101(4), 631-642.
- Kavzoglu, T., Sahin, E. K., & Colkesen, I. (2014). Landslide susceptibility mapping using GIS-based multi-criteria decision analysis, support vector machines, and logistic regression. *Landslides*, 11(3), 425-439.
- Kayastha, P., Dhital, M. R., & De Smedt, F. (2013). Application of the analytical hierarchy process (AHP) for landslide susceptibility mapping: A case study from the Tinau watershed, west Nepal. *Computers & Geosciences*, 52, 398-408.
- Kelly, D. L., & Kolstad, C. D. (1999). Bayesian learning, growth, and pollution. *Journal of economic dynamics and control*, 23(4), 491-518.
- Kotsiantis, S. B. (2013). Decision trees: a recent overview. *Artificial Intelligence Review*, 39(4), 261-283.
- Kramer, O. (2013). K-nearest neighbors *Dimensionality reduction with unsupervised nearest neighbors* (pp. 13-23): Springer.
- Lee, S., Hong, S.-M., & Jung, H.-S. (2017). A support vector machine for landslide susceptibility mapping in Gangwon Province, Korea. *Sustainability*, 9(1), 48.
- Lee, S., & Sambath, T. (2006). Landslide susceptibility mapping in the Damrei Romel area, Cambodia using frequency ratio and logistic regression models. *Environmental geology*, 50(6), 847-855.
- Leung, K. M. (2007). Naive bayesian classifier. *Polytechnic University Department of Computer Science/Finance and Risk Engineering*, 2007, 123-156.
- Moore, I. D., Gessler, P. E., Nielsen, G., & Peterson, G. (1993). Soil attribute prediction using terrain analysis. *Soil science society of america journal*, 57(2), 443-452.
- Moore, I. D., Grayson, R., & Ladson, A. (1991). Digital terrain modelling: a review of hydrological, geomorphological, and biological applications. *Hydrological processes*, 5(1), 3-30.
- Oh, H.-J., & Pradhan, B. (2011). Application of a neuro-fuzzy model to landslide-susceptibility mapping for shallow landslides in a tropical hilly area. *Computers & Geosciences*, 37(9), 1264-1276.
- Ohlmacher, G. C. (2007). Plan curvature and landslide probability in regions dominated by earth flows and earth slides. *Engineering Geology*, 91(2-4), 117-134.
- Parkash, S. (2011). Historical records of socio-economically significant landslides in India. *J. South Asia Disaster Stud*, 4(2), 177-204.
- Patle, A., & Chouhan, D. S. (2013). *SVM kernel functions for classification*. Paper presented at the 2013 International Conference on Advances in Technology and Engineering (ICATE).
- Paul, B. K. (2011). *Environmental hazards and disasters: contexts, perspectives and management*: John Wiley & Sons.
- 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.
- Pettorelli, N. (2013). *The normalized difference vegetation index*: Oxford University Press.
- Pham, B. T., Prakash, I., Khosravi, K., Chapi, K., Trinh, P. T., Ngo, T. Q., . . . Bui, D. T. (2019). A comparison of Support Vector Machines and Bayesian algorithms for landslide susceptibility modelling. *Geocarto International*, 34(13), 1385-1407.
- Pham, B. T., Van Phong, T., Nguyen-Thoi, T., Trinh, P. T., Tran, Q. C., Ho, L. S., . . . Le, H. Q. (2020). GIS-based ensemble soft computing models for landslide susceptibility mapping. *Advances in Space Research*, 66(6), 1303-1320.
- Popescu, M. E., & Sasahara, K. (2009). Engineering measures for landslide disaster mitigation *Landslides–Disaster Risk Reduction* (pp. 609-631): Springer.

- Pradhan, B., Chaudhari, A., Adinarayana, J., & Buchroithner, M. F. (2012). Soil erosion assessment and its correlation with landslide events using remote sensing data and GIS: a case study at Penang Island, Malaysia. *Environmental monitoring and assessment*, 184(2), 715-727.
- Pradhan, B., Oh, H.-J., & Buchroithner, M. (2010). Weights-of-evidence model applied to landslide susceptibility mapping in a tropical hilly area. *Geomatics, Natural Hazards and Risk*, 1(3), 199-223.
- Rahim, I., Ali, S. M., & Aslam, M. (2018). GIS Based landslide susceptibility mapping with application of analytical hierarchy process in District Ghizer, Gilgit Baltistan Pakistan. *Journal of Geoscience and Environment Protection*, 6(2), 34-49.
- Rotaru, A., Oajdea, D., & Răileanu, P. (2007). Analysis of the landslide movements. *International journal of geology*, 1(3), 70-79.
- Roy, J., Saha, S., Arabameri, A., Blaschke, T., & Bui, D. T. (2019). A novel ensemble approach for landslide susceptibility mapping (LSM) in Darjeeling and Kalimpong districts, West Bengal, India. *Remote Sensing*, 11(23), 2866.
- Sagi, O., & Rokach, L. (2018). Ensemble learning: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 8(4), e1249.
- Saito, H., Nakayama, D., & Matsuyama, H. (2009). Comparison of landslide susceptibility based on a decision-tree model and actual landslide occurrence: the Akaishi Mountains, Japan. *Geomorphology*, 109(3-4), 108-121.
- Sánchez-Marono, N., Alonso-Betanzos, A., & Tombilla-Sanromán, M. (2007). *Filter methods for feature selection—a comparative study*. Paper presented at the International Conference on Intelligent Data Engineering and Automated Learning.
- Solorio-Fernández, S., Carrasco-Ochoa, J. A., & Martínez-Trinidad, J. F. (2020). A review of unsupervised feature selection methods. *Artificial Intelligence Review*, 53(2), 907-948.
- Talavera, L. (2005). *An evaluation of filter and wrapper methods for feature selection in categorical clustering*. Paper presented at the International Symposium on Intelligent Data Analysis.
- Temme, A. J. (2021). Relations between soil development and landslides. *Hydrogeology, chemical weathering, and soil formation*, 177-185.
- Tian, Y., & Feng, Y. (2020). RaSE: Random subspace ensemble classification. *arXiv preprint arXiv:2006.08855*.
- Vapnik, V., Guyon, I., & Hastie, T. (1995). Support vector machines. *Mach. Learn*, 20(3), 273-297.
- Vergari, F., Della Seta, M., Del Monte, M., Fredi, P., & Palmieri, E. (2011). Landslide susceptibility assessment in the Upper Orcia Valley (Southern Tuscany, Italy) through conditional analysis: A contribution to the unbiased selection of causal factors. *Natural Hazards and Earth System Sciences*, 11(5), 1475–1497. <https://doi.org/10.5194/nhess-11-1475-2011>
- Wang, Y., Fang, Z., Wang, M., Peng, L., & Hong, H. (2020). Comparative study of landslide susceptibility mapping with different recurrent neural networks. *Computers & Geosciences*, 138, 104445.
- Wilson, J. (2000). Digital terrain analysis. In 'Terrain analysis: principles and applications' (Eds JP Wilson, JC Gallant) pp. 1–27: John Wiley & Sons Inc.: New York.
- 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.
- Zhu, A.-X., Miao, Y., Wang, R., Zhu, T., Deng, Y., Liu, J., . . . 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.