CrossMark

ORIGINAL ARTICLE

Mapping landslide susceptibility with frequency ratio, statistical index, and weights of evidence models: a case study in northern Iran

Samaneh Razavizadeh[1] · Karim Solaimani[2] · Matteo Massironi[3] ·
Ataollah Kavian[2]

Abstract The main objective of this study is to investigate potential application of frequency ratio (FR), weights of evidence (WoE), and statistical index (SI) models for landslide susceptibility mapping in a part of Mazandaran Province, Iran. First, a landslide inventory map was constructed from various sources. The landslide inventory map was then randomly divided in a ratio of 70/30 for training and validation of the models, respectively. Second, 13 landslide conditioning factors including slope degree, slope aspect, altitude, plan curvature, stream power index, topographic wetness index, sediment transport index, topographic roughness index, lithology, distance from streams, faults, roads, and land use type were prepared, and the relationships between these factors and the landslide inventory map were extracted by using the mentioned models. Subsequently, the multi-class weighted factors were used to generate landslide susceptibility maps. Finally, the susceptibility maps were verified and compared using several methods including receiver operating characteristic curve with the areas under the curve (AUC), landslide density, and spatially agreed area analyses. The success rate curve showed that the AUC for FR, WoE, and SI models was 81.51, 79.43, and 81.27, respectively. The prediction rate curve demonstrated that the AUC achieved by the three models was 80.44, 77.94, and 79.55, respectively. Although the sensitivity analysis using the FR model revealed that the modeling process was sensitive to input factors, the accuracy results suggest that the three models used in this study can be effective approaches for landslide susceptibility mapping in Mazandaran Province, and the resultant susceptibility maps are trustworthy for hazard mitigation strategies.

Keywords Susceptibility modeling · Geographic information systems (GISs) · Bivariate statistics · Mazandaran Province

Introduction

The world's development infrastructure is at risk from landslides and their consequences in many areas across the globe. There is evidence that landslide disaster risk is increasing in developing countries (Anderson and Holcombe 2013). With rapid urbanization, the growth of densely populated communities in mountainous and hazardous locations is continuing and this leads to the instability of slopes and thus increases the potential for landslides. Therefore, developing landslide susceptibility maps, risk analyzing, and adopting suitable land use policies for different environmental settings are urgently needed.

Landslides occur due to complex and relate to various factors such as geology, topography, hydrogeological conditions, vegetation, rainstorm, and human activities (Cruden 1991; Montgomery and Dietrich 1994; Wu and Sidle 1995; Guzzetti et al. 1999; Gorsevski et al. 2006; Ardizzone et al. 2007).

A wide variety of methods and techniques have been proposed and applied for landslide susceptibility mapping

✉ Samaneh Razavizadeh
razavizadeh@rifr-ac.ir

[1] Research Institute of Forests and Rangelands, Agricultural Research, Education and Extension Organization (AREEO), Tehran, Iran

[2] Department of Watershed Management, Faculty of Natural Resources, Sari University of Agricultural Science and Natural Resources, Sari, Iran

[3] Department of Geosciences, University of Padova, Padova, Italy

(LSM). The most common approaches proposed in the literature are frequency ratio (FR), statistical index (SI), weights of evidence (WoE), logistic regression, multivariate regression, discriminant analysis, index of entropy, spatial multi-criteria evaluation, analytical hierarchy process, evidential belief function, decision tree, artificial neural network, fuzzy logic, neuro-fuzzy, and support vector machines (Gorsevski et al. 2003, 2005; Nefeslioglu et al. 2008; Gorsevski and Jankowski 2010; Pourghasemi et al. 2012, 2013; Ozdemir and Altural 2013; Pradhan 2013; Jebur et al. 2014; Jaafari et al. 2014, 2015a, 2017a; Pham et al. 2016, 2017).

Each method differs in terms of input process, calculations, output process, and predictive reliability. Although many comparative studies that have been carried out on predictive ability of different methods (e.g., Pradhan 2013; Pourghasemi et al. 2013; Jebur et al. 2014; Hong et al. 2015; Pham et al. 2016), decision makers and engineers involved in slope management and land use planning still need to select the best method for different environmental settings. Among the commonly used geographic information system (GIS)-based models for susceptibility modeling, FR, SI, and WoE have been widely investigated in the literature (e.g., Mohammady et al. 2012; Pourghasemi et al. 2013; Ozdemir and Altural 2013; Regmi et al. 2014; Jaafari et al. 2014; Youssef et al. 2016). While, in these models, input process and calculations are very straightforward and can easily be implemented within a GIS environment (Mohammady et al. 2012; Jaafari et al. 2014), their results prove high level of accuracy for prediction of future landslides that give rise to appealing qualitative and quantitative maps for the landslide-prone areas (Mohammady et al. 2012; Youssef et al. 2016). These advantages have motivated us to use these three models to assess landslide susceptibility for a part of Mazandaran Province of northern Iran. Our study aims to compare the predictive ability of FR, WoE, and SI methods for LSM in the study area. Since this is a prototype study performed in one of the characteristic landslide susceptible regions of northern Iran, the findings can be utilized for regions that show similar geoenvironmental and landslide characteristics.

Study area

The study area is located about 80 km southwest of Sari between 36°11′43″N to 36°18′14″N latitude and 51°14′37″E to 52°58′49″ E longitude in the Mazandaran Province of northern Iran. It covers an area of about 270 km^2 and encompasses farmlands, mountainous terrains of forest, rangelands, and residential areas. The area represents a variable and rough topography, with slope gradients between 0° (flat) and 62° and altitudes between 1173 and 3536 m. The climate in the study area is Mediterranean with a mean annual precipitation of 800 mm that occurs in the form of snow during the winter. The lithology of the study area consists of several geologic units such as dolomite, siltstone, sandstone, marl, and conglomerate (Table 1). The soil texture consists of several types including sand, loamy sand, sandy clay loam, sandy loam, silty clay, silty loam, and silty sand, of which sandy clay loam occupies about 78% of the total area. In this area, deforestation and inappropriate land use practices contribute to natural disasters such as flooding, soil erosion, and landslides in the past several years. The most important fault in this area is the Alborz Fault that is a reverse fault and follows the west–east orientation (Jaafari et al. 2017a). Some reports suggest that this fault is the main source of most earthquakes and landslides that occur in the Mazandaran Province (Darvishzadeh 2004; Jaafari et al. 2017a).

Database construction

Our database for LSM involves an inventory map of landslides occurred within the study area and a set of landslide conditioning factors. The inventory map of existing landslides of the study area was compiled using

Table 1 Lithology of the study area

| Geological age	Code	Formation	Lithology
Cretaceous	Ku^2	–	Limestone, sandy limestone, sandstone
Jurassic	J$_s$	Shemshak	Sandstone, shale, and conglomerate
	Jl	Lar	Dolomite, dolomitic limestone, chert nodules
Triassic	TR$_e^1$	Elika	Pale red, thin-bedded to thick-bedded calcareous silty shale, argillaceous limestone, sandy limestone, limestone, and quartzarentic
	TR$_e^2$	Elika	Pale gray to pale yellow, thick-bedded cherty dolomite
	TR$_b$	Elika	Thick-bedded massive dolomites and dolomitic limestone
Permian	P$_r$	Ruteh	Fusulina limestone, dolomitic limestone, with chert in the upper part
Quaternary	Q^s	–	Scree, locally may include undivided alluvial deposits
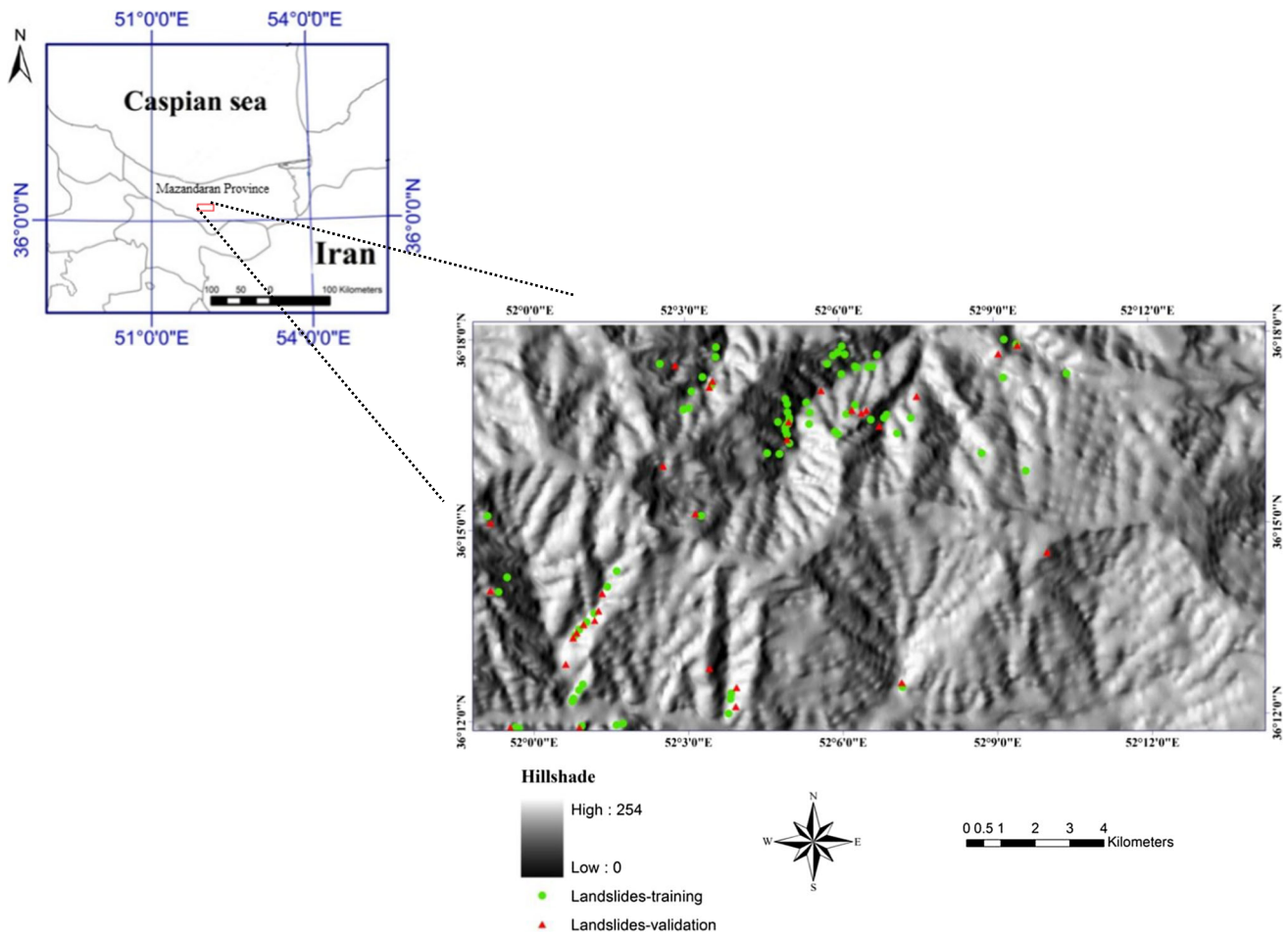
Fig. 1 Location and landslide inventory map of the study area

aerial photography interpretation and extensive field surveys. In total, 105 landslides were mapped as georeferenced points (Fig. 1). The detected landslides were then divided into two subsets: The training dataset that contained 70% of the landslide inventory (74 landslides) was used in the training phase of landslide models, and the validation dataset with 30% of the data (31 landslides) was used for the validation purpose (Pourghasemi et al. 2013; Hong et al. 2015; Jaafari et al. 2015a).

Given the multiplicity of the landslide conditioning factors, they are usually selected based on the landslide types, the failure mechanisms, the map scale of analysis, and the characteristics of the study area and data availability (Glade and Crozier 2005). A total number of 13 landslide conditioning factors were considered to be used in this study. They are slope degree, slope aspect, altitude, plan curvature, stream power index (SPI), topographic wetness index (TWI), sediment transport index (STI), topographic roughness index (TRI), lithology, distance from streams, faults, roads, and land use type (Figs. 2, 3, 4, 5; Table 1). The significance of these factors on landslide

occurrence has explicitly been acknowledged by other authors (e.g., Dai et al. 2001; van Westen et al. 2008; Yalcin et al. 2011; Guillard and Zezere 2012; Mohammady et al. 2012; Kayastha et al. 2012; Ozdemir and Altural 2013; Pourghasemi et al. 2013; Jebur et al. 2014; Jaafari et al. 2014, 2015a; Pham et al. 2017). Slope degree, slope aspect, altitude, and TRI were selected to represent the effect of topographic factors on slope stability analysis. Topographic factors generally control several characteristics (e.g., topographic heterogeneity, land use/cover, shear forces, amount of rainfall, terrain humidity, and erosion–weathering processes) of the landscape that may modulate occurrence of landslides (Guzzetti et al. 1999). Plan curvature and water-related factors (i.e., SPI, STI, and TWI) were chosen due to their influence on hydrogeological conditions that, in turn, exert effect on the surface runoff and infiltration. We also used lithology and distance from faults factors to represent the influence of geomorphological processes and tectonic factors on the occurrence of landslides (Jaafari et al. 2015a; Pham et al. 2017). Since the under-cutting actions and erosion process of streams can
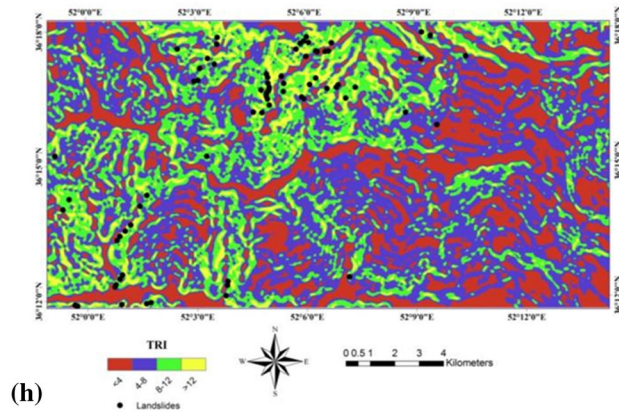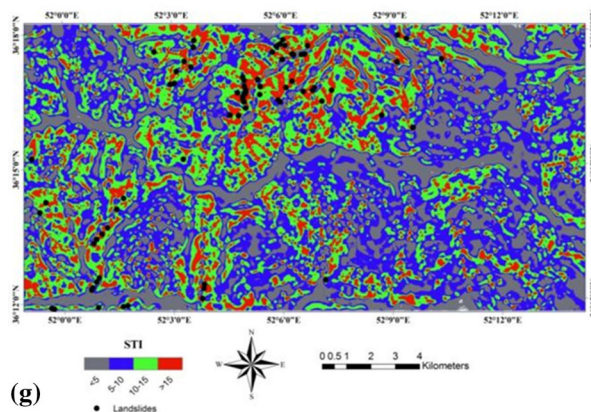
(a) Slope degree
0-15, 15-20, 20-25, 25-30, 30-35, >35
• Landslides

(b) Aspect
F, N, NE, E, SE, S, SW, W, NW
• Landslides

(c) Altitude (m)
<1500, 1500-2000, 2000-2500, 2500-3000, >3000
• Landslides

(d) Plan Curvature (100/m)
Concave, flat, Convex
• Landslides

(e) SPI
<200, 200-400, >400
• Landslides

(f) TWI
<6, 6-12, 12-18, >18
• Landslides

(g) STI
<5, 5-10, 10-15, >15
• Landslides

(h) TRI
<4, 4-8, 8-12, >12
• Landslides
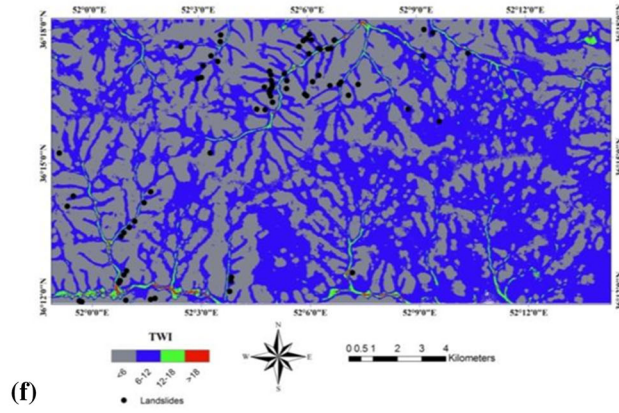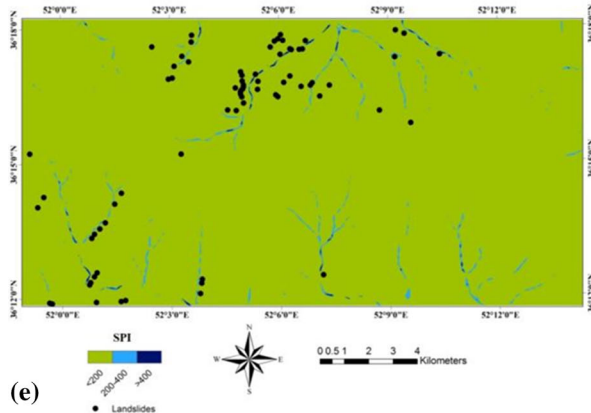
◄**Fig. 2** Topographic parameter maps of the study area; **a** slope degree, **b** slope aspect, **c** altitude, **d** plan curvature, **e** stream power index, **f** topographic wetness index, **g** sediment transport index, **h** topographic roughness index
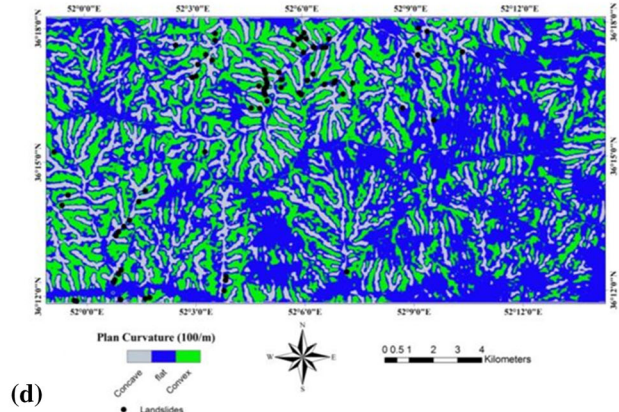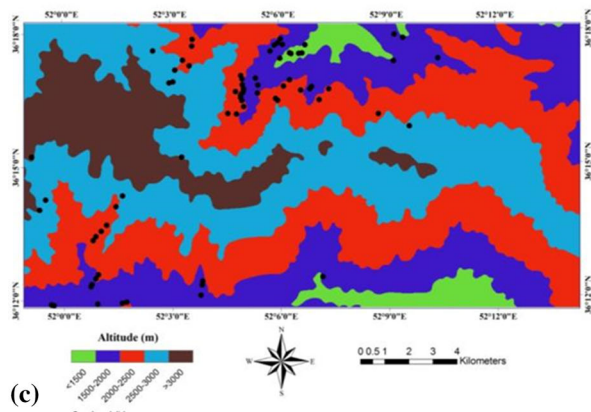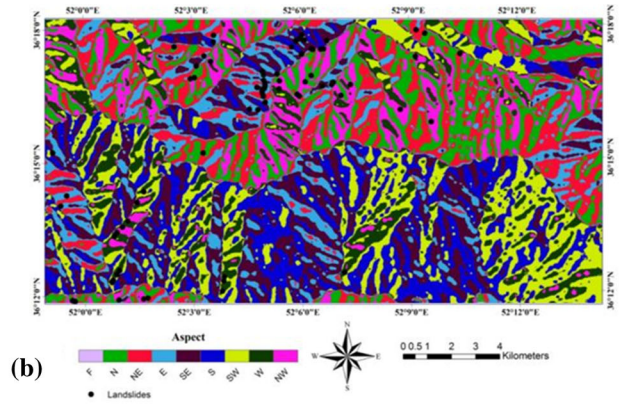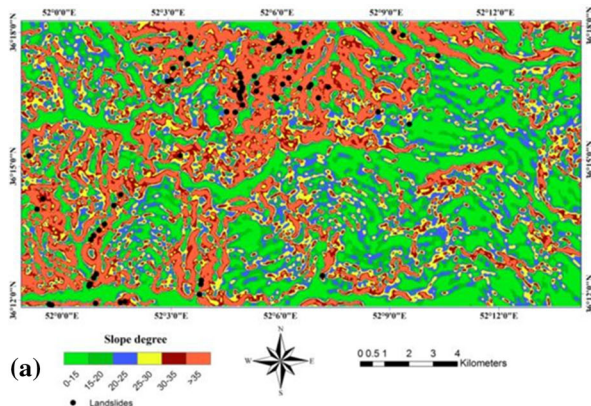
trigger landslides (Xu et al. 2012; Jaafari et al. 2015a), the distance from streams was also included to our analysis. Types of land use/cover, which reflect the human activities on the landscape, can significantly affect the susceptibility of a given landscape (Catani et al. 2013). Road development, especially in mountainous regions, is usually the main cause of erosion and slope failure (Nefeslioglu et al. 2008; Jaafari et al. 2015a, b). Thus, distance from roads is a commonly used conditioning factor in landslide susceptibility modeling (e.g., Pourghasemi et al. 2013; Hong et al. 2015; Pourghasemi and Kerle 2016; Pham et al. 2016).

A digital elevation model (DEM) with 20×20 m grid size (Projection: UTM 39N; Datum: WGS1984) and the data prepared by Geological Survey of Iran were used as the main input data to generate the aforementioned conditioning factors (Table 2). All the calculations and data processing were hosted in ArcGIS 9.3 and SAGA GIS 2.1.4. The map of all of the conditioning factors was generated and converted into raster format with the size of 20×20 m (Kayastha et al. 2012; Ozdemir and Altural 2013; Jaafari et al. 2014, 2015a, b; Pham et al. 2017). To implement the FR, WoE, and SI models, the conditioning factors were classified into different classes. The different classes were selected based on previous landslide studies (e.g., Pourghasemi et al. 2013; Jaafari et al. 2014) that were further informed by our surveys and observations in the study area.

## Multi-collinearity analysis of landslide conditioning factor

In the final step of preparing a spatial database for landslide modeling, we examined whether any of the selected conditioning factors exhibited multi-collinearity. Tolerance and variance inflation factors (VIF) (Hair et al. 2006) are the most common metrics to check for multi-collinearity among factors (e.g., Hong et al. 2015; Jaafari et al. 2017b) that a violation of their critical values (VIF > 5 and tolerance <0.2) indicates a potential problem with multi-collinearity (Hair et al. 2006). In our study, the computation of these metrics showed that no significant multi-collinearity existed among the conditioning factors (Table 3) and therefore all factors can be used in the modeling process.

Methodology

Frequency ratio model

As a bivariate statistical technique, FR model is a robust geospatial assessment tool for computing the probabilistic relationship between dependent and independent variables (Oh et al. 2011). For the purpose of LSM, FR considers the impact of each conditioning factor on landsliding and assigns the weights very precisely (Lee and Pradhan 2007). If the weight be more than 1, it means a greater correlation, whereas the weights less than 1 represent a minor correlation (Lee and Min 2001). The calculation process of FR model is very straightforward and can be
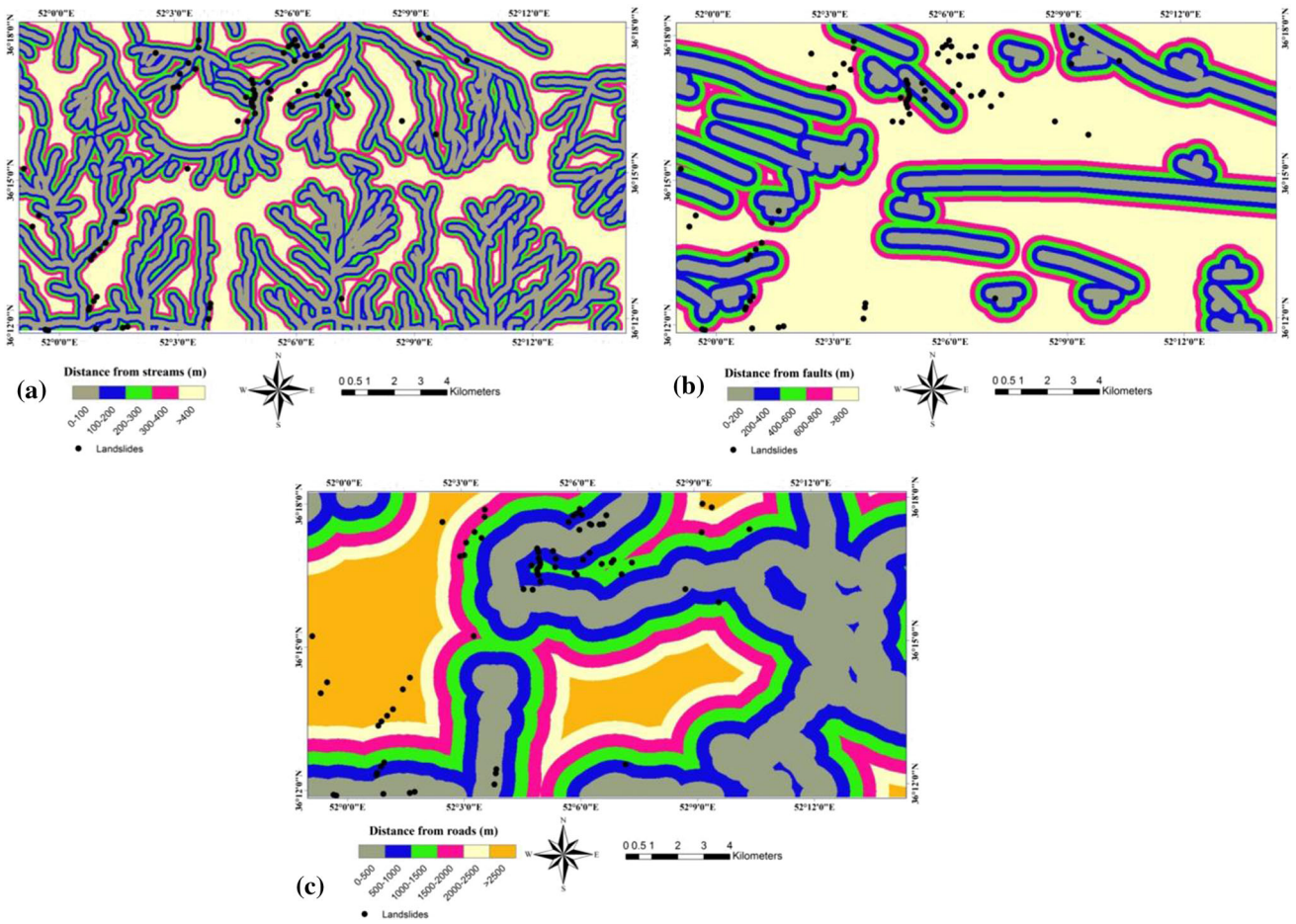
Fig. 3 Lithology map of the study area
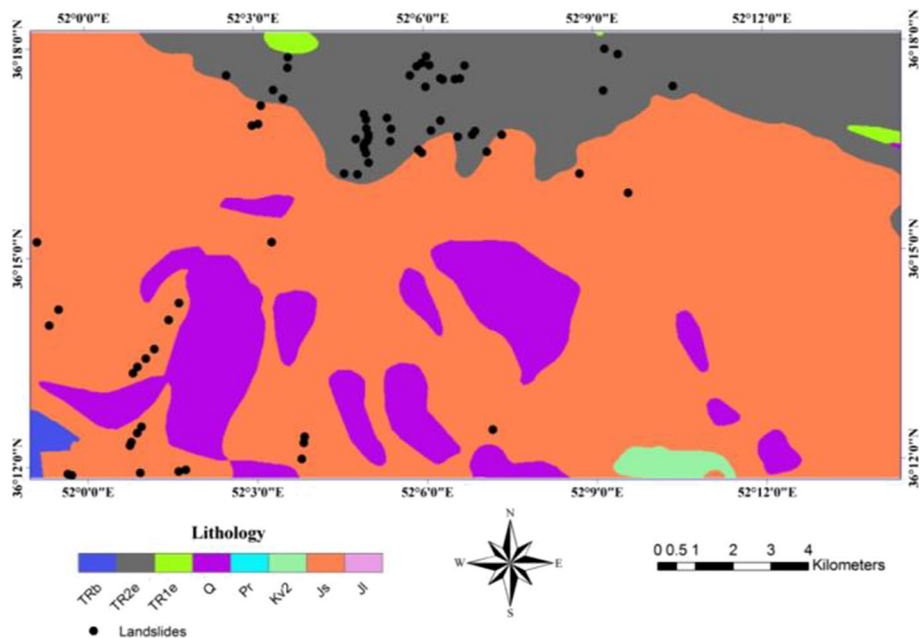
Fig. 4 Distance factors; a distance from streams, b distance from faults, c distance from roads
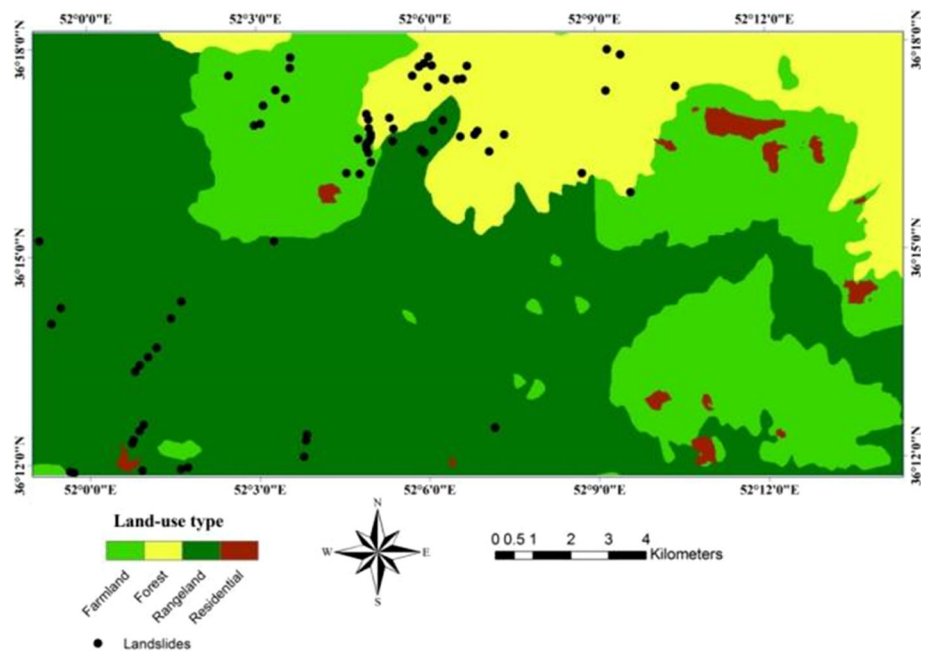

Fig. 5 Land use types of the study area

Table 2 Data source for the landslide conditioning factors

Factor	Source
DEM and its derived factors (slope degree, slope aspect, altitude, plan curvature, TRI, SPI, STI, and TWI)	ASTER Global DEM (http://earthexplorer.usgs.gov)
Lithology	Geological map at 1:100,000 scale (www.gsi.ir)
Distance from streams	Stream network map
Distance from faults	Geological map at 1:100,000 scale (www.gsi.ir)
Distance from roads	Road network map
Land use type	Landsat 7 ETM+ images (http://earthexplorer.usgs.gov)

Table 3 Multi-collinearity analysis for the conditioning factors

Factor	Collinearity statistics	
	Tolerance	VIF
Slope degree	0.528	2.128
Slope aspect	0.719	3.153
Altitude	0.924	2.243
Plan curvature	0.869	2.111
SPI	0.799	1.712
TWI	0.87	2.259
STI	0.811	2.217
TRI	0.708	3.091
Lithology	0.907	2.736
Distance from streams	0.687	3.001
Distance from faults	0.794	2.241
Distance from roads	0.751	2.452
Land use type	0.716	1.984

readily realized as follows (Regmi et al. 2014; Jaafari et al. 2014):

$$FR = \frac{E/F}{M/L} \quad (1)$$

where E is the number of pixels with landslide for each conditioning factor; F, the number of total landslides in study area; M, the number of pixels in the class area of the conditioning factor; and L, the number of total pixels in the study area.

Weights of evidence model

WoE model is a data-driven method based on the Bayesian probability framework (Bonham-Carter et al. 1989). This model is suitable for LSM because its uncertainty is connected with landslide events and their associations with the complex landscape (Chung and Fabbri 1998; Gorsevski et al. 2003; Regmi et al. 2014; Jaafari et al. 2015a, b). This model is based on the determination of positive (W^+) and negative weights ($W^−$). The WoE calculates the weight for each class of conditioning factors (B) based on the presence or absence of the landslides (L) within the area as follows (Bonham-Carter et al. 1989):

$$W_i^+ = \ln \frac{P\{B|L\}}{P\{B|\bar{L}\}} \quad (2)$$

$$W_i^- = \ln \frac{P\{\bar{B}|L\}}{P\{\bar{B}|\bar{L}\}} \quad (3)$$

$$C = W^+ + W^- \quad (4)$$

$$S(C) = \sqrt{S^2(w^+) + S^2(w^-)} \quad (5)$$

$$S^2(w^+) = \frac{1}{N\{B \cap L\}} + \frac{1}{\{B \cap \bar{L}\}} \quad (6)$$

$$S^2(w^-) = \frac{1}{N\{\bar{B} \cap L\}} + \frac{1}{\{\bar{B} \cap \bar{L}\}} \quad (7)$$

$$W = \frac{C}{S(C)} \quad (8)$$

where W_i^+ is positive weight; W_i^-, negative weight; ln, natural log; P, conditional probability; B, the presence of a potential conditioning factor; $\bar{B}$, the absence of a potential conditioning factor; L, the presence of a landslide; $\bar{L}$, the absence of a landslide; C, weight contrast; $S^2 (w^+)$ and $S^2 (w^-)$, variances of positive and negative weights; W, studentized contrast (final weight); and $S (C)$, the standard deviation of the contrast.

Statistical index model

SI model is a bivariate statistical method proposed by van Westen (1997) for the purpose of LSM. This method is based on the following equation:

$$W_{SI} = \ln \left(\frac{F_{ij}}{F} \right) = \ln \left(\frac{L_{ij}/L_T}{P_{ij}/P_L} \right) \quad (9)$$

where W_{SI} is the weight given to a certain class i of factor j; F_{ij}, landslide density within class i of factor j; F, total landslide density within the entire map; L_{ij}, number of landslides in a certain class i of factor j; P_{ij}, number of pixels in a certain class i of factor j; L_T, total number of landslides in the entire map; and P_L, total pixels of the entire map.

A positive value of W_{SI} demonstrates the existence of a relationship between the presence of the factor class and landslide distribution, the stronger the relationship the higher the score. The W_{SI} is negative when the presence of

the factor class is not relevant in landslide development (Pourghasemi et al. 2013).

The weights calculated using the three models were then assigned to the classes of each conditioning factor to produce multi-class weighted maps for all factors, which were overlaid and numerically added according to the following equation in order to calculate the landslide susceptibility index (LSI) maps (Jaafari et al. 2015a):

$$ LSI = \sum_{i=1}^{13} W \tag{10} $$

where W corresponds to the multi-class weighted conditioning factors.

Performance validation and factor effect analysis

Validation is the most important step in a modeling effort, and without validation the model results lack scientific significance (Chung and Fabbri 2003; Jaafari et al. 2017b). The receiver operating characteristics (ROC) curves with the area under curve (AUC) approach are widely used criteria for evaluating the performance of the prediction models (Pourghasemi et al. 2013; Jebur et al. 2014; Jaafari et al. 2015a, b, 2017a, b; Hong et al. 2015; Pham et al. 2016; Nami et al. 2017). ROC curve is a binary classification metric created by plotting the false-positive rate and the true-positive rate for every possible binary classification of a dataset (Zweig and Campbell 1993). The ideal ROC curve passes through the point of (0, 1) with AUC = 1, indicating that there is no prediction error. An acceptable AUC value for a ROC curve should be higher than 0.5 (Yesilnacar 2005). Further, to investigate the reliability of the produced maps, we used landslide density and spatially agreed area analyses within different classes of each susceptibility map. Landslide density indicates the ratio of landslide pixels to the ratio of total pixels (Pham et al. 2017). The spatially agreed area, expressed in pixels, km², and as percentage of the total area, is calculated as the total area having the same landslide susceptibility zonation on two susceptibility maps (Bijukchhen et al. 2013; Kayastha et al. 2013). These analyses can reveal spatial differences between the maps and improve predictions in the agreed high susceptible zones.

In this research, we also performed an analysis of conditioning factors effect to explore the effect of each factor on the prediction results and to assess their uncertainties. To this end, we first took the method that achieved the highest ROC values and excluded each of the 13 factors in turn during the summation stage of Eq. 10. We next calculated the success and prediction rates for all cases and

compared these cases to the case in which all factors were included (Jaafari et al. 2017b).

Finally, for visual interpretation of the LSI maps, the data were classified into categorical susceptibility classes by examining different classifications methods, including quantile, natural breaks, standard deviation, equal interval, and geometrical interval (Ayalew and Yamagishi 2005). The comparison results indicated that the quantile method was able to produce better results than the other methods. Therefore, this method was chosen and the landslide susceptibility index maps were classified into four susceptibility classes that represent low, moderate, high, and very high susceptibility to landslide occurrence across the study area.

Results and discussion

Application of frequency ratio model

Table 4 shows the results of spatial relationship between landslide locations and landslide conditioning factors using the FR model. From this table, it is seen that a slope angle slope >35° has a higher FR value of 3.67, followed by 30–35° (1.17), whereas other slope classes have a very low value of FR. Given the lower shear stresses associated with low gradients areas, gentle slopes were frequently reported to have lower weight values of FR (Yalcin et al. 2011; Mohammady et al. 2012; Jaafari et al. 2014; Youssef et al. 2016). For slope aspect, the value of FR is higher for the areas facing the northwest, east, west, and southeast. The study conducted by Jaafari et al. (2014) in other regions of northern Iran supports the significance of these directions of slope angle on landslide occurrence. The relationship between altitude and landslide probability shows that the class of 0–1500 m has highest FR value (1.99), indicating higher landslide susceptibility at this range of elevation. This finding agrees with the field surveys since the landslides were commonly observed on low-elevation ranges of the study area. In the case of plan curvature, convex and concave areas have higher FR, with values of 1.59 and 1.37, respectively, whereas the flat class has the lowest FR value (0.56). Generally, the slope instability in the curvature areas is related to the high level of soil moisture. As the moisture content of the soil increases, soil stability generally decreases (Jaafari et al. 2014). According to the results of the FR model obtained for SPI, the highest FR value is related to the class of <200 (1.01). Similarly, for TWI, the areas with lowest value of TWI have higher FR values (1.44). In the case of the STI, landslide susceptibility is higher in areas where the STI is >15. In the case of TRI, the class of >12 shows the highest landslide susceptibility with a FR values of 5.44. Since the lithological

Table 4 Spatial relationship between each landslide conditioning factor and landslides extracted by using FR, WoE, and SI models

Factor	Class	Number of pixels in domain	Number of landslides	FR	SI	C	S^2 (w^+)	S^2 (w^-)	$S(C)$	W
Slope degree	0–15	157,373	6	0.36	−1.029	−1.201	0.167	0.015	0.426	−2.821
	15–20	106,389	1	0.09	−2.429	−2.582	1.000	0.014	1.007	−2.564
	20–25	115,404	5	0.41	−0.901	−1.013	0.200	0.014	0.463	−2.187
	25–30	106,678	4	0.35	−1.045	−1.157	0.250	0.014	0.514	−2.250
	30–35	87,897	11	1.17	0.000	0.185	0.091	0.016	0.327	0.567
	>35	120,099	47	3.67	0.000	2.119	0.021	0.037	0.242	8.772
Slope aspect	F	232	0	0.00	None	None	None	0.014	None	None
	N	88,623	9	0.95	−0.049	−0.056	0.111	0.015	0.356	−0.157
	NE	94,104	6	0.60	−0.514	−0.576	0.167	0.015	0.426	−1.352
	E	72,051	14	1.82	0.600	0.700	0.071	0.017	0.297	2.358
	SE	103,913	15	1.35	0.303	0.367	0.067	0.017	0.289	1.269
	S	112,382	2	0.17	−1.791	−1.940	0.500	0.014	0.717	−2.706
	SW	101,042	3	0.28	−1.279	−1.395	0.333	0.014	0.589	−2.366
	W	59,714	10	1.57	0.451	0.506	0.100	0.016	0.340	1.489
	NW	61,779	15	2.28	0.823	0.956	0.067	0.017	0.289	3.306
Altitude (m)	<1500	28,209	6	1.99	0.690	0.733	0.167	0.015	0.426	1.722
	1500–2000	153,463	28	1.71	0.537	0.762	0.036	0.022	0.240	3.181
	2000–2500	221,252	31	1.31	0.273	0.432	0.032	0.023	0.236	1.832
	2500–3000	207,018	8	0.362	−1.015	−1.255	0.125	0.015	0.374	−3.353
	>3000	83,898	1	0.112	0.000	−2.307	1.000	0.014	1.007	−2.291
Plan curvature (100.m)	Concave	130,212	19	1.37	0.313	0.402	0.053	0.018	0.266	1.512
	Flat	368,910	22	0.56	−0.581	−0.987	0.045	0.019	0.254	−3.882
	Convex	194,718	33	1.59	0.463	0.724	0.030	0.024	0.234	3.097
SPI	<200	685,024	74	1.01	0.013	3.374	0.014	0.007	0.144	23.414
	200–400	7751	0	0.00	None	None	None	0.005	None	None
	>400	1065	0	0.00	None	None	None	0.005	None	None
TWI	<6	339,110	52	1.44	0.363	0.905	0.019	0.045	0.254	3.559
	6–12	340,155	21	0.58	−0.547	−0.887	0.048	0.019	0.258	−3.439
	12–18	12,891	1	0.73	−0.318	−0.324	1.000	0.014	1.007	−0.321
	>18	1684	0	0.00	None	None	None	0.014	None	None
STI	<5	158,814	4	0.24	−1.444	−1.647	0.250	0.014	0.514	−3.203
	5–10	257,423	11	0.40	−0.915	−1.217	0.091	0.016	0.327	−3.723
	10–15	198,621	35	1.65	0.502	0.806	0.029	0.026	0.233	3.460
	>15	78,982	24	2.85	1.047	1.319	0.042	0.020	0.248	5.310
TRI	<4	171,699	5	0.27	−1.298	−1.513	0.200	0.014	0.463	−3.266
	4–8	301,998	11	0.34	−1.074	−1.485	0.091	0.016	0.327	−4.544
	8–12	166,760	27	1.52	0.417	0.597	0.037	0.021	0.241	2.470
	>12	53,383	31	5.44	1.695	2.158	0.032	0.023	0.236	9.157
Lithology	TR_e^2	127,963	42	3.08	1.124	1.759	0.024	0.031	0.235	7.495
	TR_e^1	2898	0	0.00	None	None	None	0.005	None	None
	Q^s	85,063	0	0.00	None	None	None	0.005	None	None
	P_r	7	0	0.00	None	None	None	0.005	None	None
	TR_b	3973	0	0.00	None	None	None	0.005	None	None
	Ku^2	5223	0	0.00	None	None	None	0.005	None	None
	J_s	468,700	32	0.64	−1.005	0.031	0.024	0.235	−4.284	−1.005
	Jl	13	0	0.00	None	None	None	0.005	None	None

Table 4 continued

Factor	Class	Number of pixels in domain	Number of landslides	FR	SI	C	S^2 (w^+)	S^2 (w^-)	$S(C)$	W
Distance from streams (m)	0–100	173,583	20	1.08	0.104	0.050	0.019	0.262	0.399	0.104
	100–200	133,385	24	1.69	0.702	0.042	0.020	0.248	2.825	0.702
	200–300	109,237	14	1.20	0.222	0.071	0.017	0.297	0.748	0.222
	300–400	87,522	5	0.54	−0.689	0.200	0.014	0.463	−1.488	−0.689
	>400	190,113	11	0.54	−0.771	0.091	0.016	0.327	−2.359	−0.771
Distance from faults (m)	0–200	98,731	12	1.14	0.154	0.083	0.016	0.315	0.489	0.154
	200–400	97,920	5	0.48	−0.819	0.200	0.014	0.463	−1.768	−0.819
	400–600	90,185	11	1.14	0.156	0.091	0.016	0.327	0.477	0.156
	600–800	77,959	9	1.08	0.090	0.111	0.015	0.356	0.252	0.090
	>800	329,045	37	1.05	0.103	0.027	0.027	0.233	0.444	0.103
Distance from roads (m)	0–500	195,954	17	0.81	−0.277	0.059	0.018	0.276	−1.004	−0.277
	500–1000	131,438	20	1.43	0.460	0.050	0.019	0.262	1.759	0.460
	1000–1500	106,328	19	1.68	0.647	0.053	0.018	0.266	2.429	0.647
	1500–2000	80,810	6	0.70	−0.401	0.167	0.015	0.426	−0.943	−0.401
	2000–2500	61,273	1	0.15	−1.956	1.000	0.014	1.007	−1.943	−1.956
	>2500	118,037	11	0.87	−0.160	0.091	0.016	0.327	−0.491	−0.160
Land use type	Residential	8245	0	0.00	None	None	None	0.014	None	None
	Farmland	178,816	19	0.47	−0.761	−0.005	0.053	0.018	0.266	−0.019
	Forest	125,418	27	2.02	0.702	0.957	0.037	0.021	0.242	3.963
	Rangeland	381,361	28	1.47	0.384	−0.696	0.036	0.022	0.240	−2.902

units of TR_e^2 and J_s represent 18.44 and 67.55% of our study area, the highest FR values were found to be related to these units. In case of distance from streams, the landslide occurrence probability is higher in areas where the distance is 100–200 m, whereas the probability decreases at a distance of more than 300 m from a stream. For distance from faults, the distances of 0–200 and 400–600 m show higher correlation with the landslides. The relationship between the road networks of the study area, and landslide probability shows that the class of 1000–1500 m has highest FR value, followed by the class of 500–1000 m. Regarding the land use type, the FR value was higher for the forested areas (2.02). In our study area, forests are mainly scattered on steep terrains and unstable slopes. These results in line with results reported by Jaafari et al. (2014, 2015a, b) provide a counterexample for the widely held notion that vegetation coverage always contributes to decreasing landslides and increasing general slope stability. Huge hardwood trees in northern forests of Iran and our study area exposed to wind may transmit the forces into the slope and cause landslide. Weight of such trees surcharges the slope, increasing normal and downhill force components, and may cause failure during the monsoon (Ghimire 2011; Jaafari et al. 2014). The final result of FR model was a LSI map, in which the LSI values vary from 5.122 to 29.6.

Application of weights of evidence model

The WoE model was used to explore the spatial association between the conditioning factors and landslide distribution (Table 2). The studentized value of C and the value of W serve as a guide to the significance of spatial association and act as a measure of relative certainty of the posterior probability (Bonham-Carter et al. 1989). Given that higher values of W indicate a higher level of significance for a specific factor class (Kayastha et al. 2012; Jaafari et al. 2015a), from Table 4 it can be seen that this value is highest for slope degree >35, indicating a significant positive correlation with the landslides. Further, the results revealed that slope degrees of 0–30 were negatively correlated with the landslides, which can further be interpreted to mean that this range of slope degree has disfavored the occurrence of landslides events across our study area. In the case of slope aspect, most of the landslides occurred in northwest ($W = 3.306$) and east ($W = 2.358$) facings. North, northeast, south, and southwest aspects were negatively correlated with the landslides, which have disfavored the occurrence of landslides. For altitude, most of the landslides occurred in 1500–2000 m class with W value of 3.181. The negative values of W for the altitude >2500 m revealed that this range of altitude disfavored the occurrence of landslides in our study area. In the case of plan

Table 5 AUC values for the landslide index maps produced by using FR model

Excluded factor	AUC (%)	
	Success rate	Prediction rate
Slope degree	78.21	77.82
Slope aspect	82.08	80.77
Altitude	80.83	80.06
Plan curvature	82.02	80.83
SPI	81.42	81.08
TWI	78.93	78.02
STI	79.51	78.73
TRI	78.21	76.96
Lithology	80.88	80.6
Distance from streams	83.06	81.91
Distance from faults	**83.76**	**82.17**
Distance from roads	81.35	80.18
Land use type	79.37	78.14
All factors used	81.51	80.44

The highest values are shown in bold

curvature, convex class has W value of 3.097. So, most of the landslides occurred in this class. In the case of the SPI, the W is highest (23.414) for <200 class. The relation between TWI and landslide probability showed that <6 class has highest value of W (3.559). For STI, the class >15 has most W value (3.460). In the case of TRI, the W is highest in >12 class (9.157) and lowest in <4 class (−3.266). In the case of lithology, the W is highest in TR_e^2

unit (7.495) since 42 cases out of 74 landslides used in training dataset were captured by this unit. In the case of distance from streams, higher W values of 0.702 and 0.222 were found for distances between 100–200 and 200–300 m, respectively. Assessment of distance from faults showed that distance of 400–600 m is more suitable for landslide occurrence with W value of 0.156. Investigation of distance from roads showed that distances of 1000–1500 have highest correlation with landslide occurrence. In the case of land use type, higher W value was seen for forest areas (3.963). The final result of WoE model was a LSI map, in which the LSI values vary from 70.2403 to 260.709.

Application of statistical index model

Similarly, LSI was calculated using SI model. The landslide distribution for each class of landslide conditioning factors was used to calculate the SI values (Table 2). As mentioned earlier, the larger the value, the stronger the relationship between landslide occurrence and the given factor's class. The final result of SI model was a LSI map, in which the LSI values vary from −13.6014 to 7.93953 (Table 5).

Landslide susceptibility maps

Landslide susceptibility levels across the study area that ranged from low to very high are shown in Figs. 6, 7, and 8. In these three maps, high and very high susceptibility

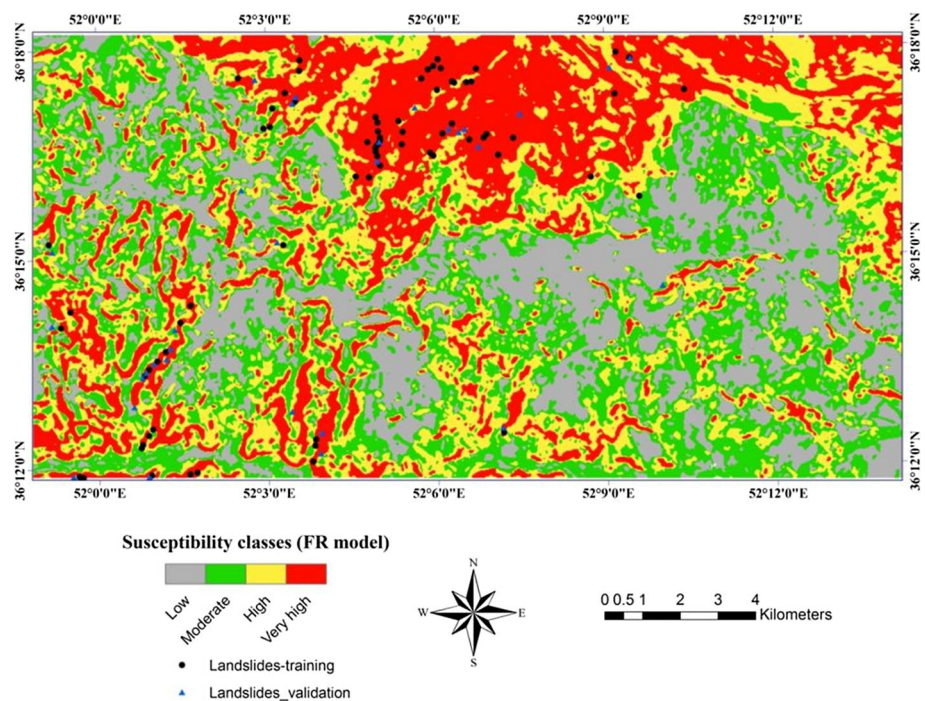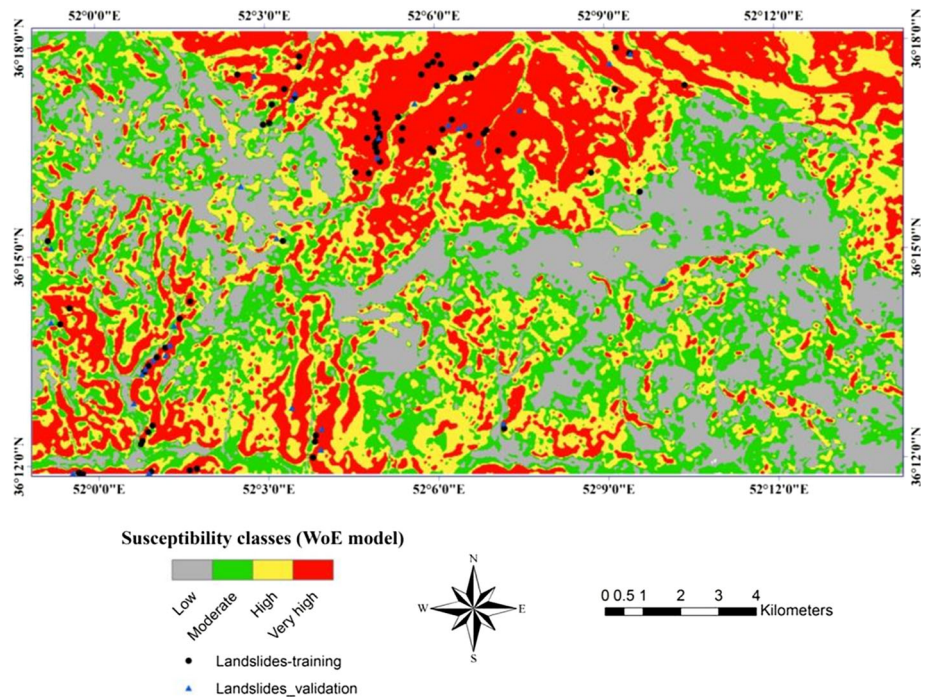**Fig. 6** Landslide susceptibility map produced by frequency ratio model

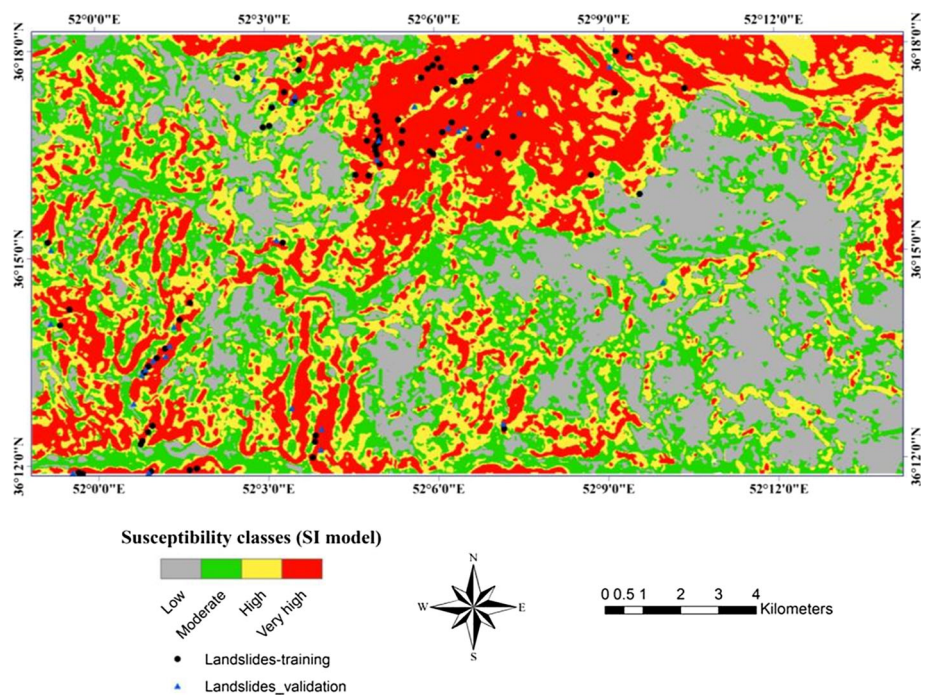**Fig. 7** Landslide susceptibility map produced by weights of evidence model

classes cover the northern and southwestern parts of the study area, whereas low and moderate susceptibility classes cover central parts. Further, the results showed that the high and very high susceptibility classes cover approximately 50% of the study area (Fig. 9). To investigate the reliability of the produced maps, we used landslide density analysis and spatially agreed area approach. The results showed that the value of landslide density varied among

the classes and ranged from 0.04 to 2.88 (Table 6). In each map, the highest value is for very high susceptibility class which is followed by high class, moderate class, and low class, respectively.

The results of the spatially agreed area revealed that the landslide susceptibility map produced using the FR method has 79.15 and 70.80% agreed area with the maps produced by the WoE and SI methods, respectively (Table 7),

Fig. 8 Landslide susceptibility map produced by statistical index model
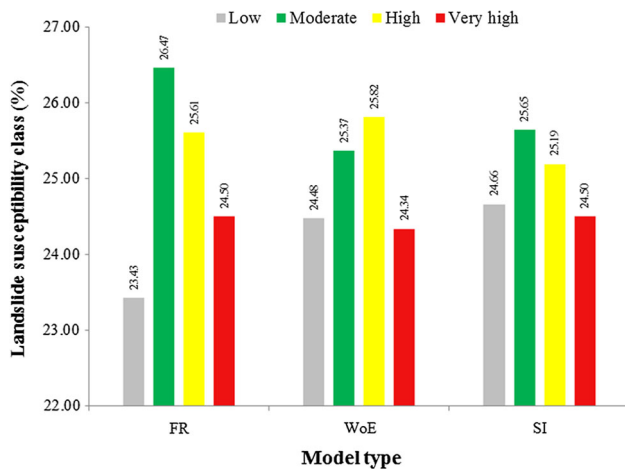
Fig. 9 Landslide susceptibility classes delimited by the five ANFIS models

whereas the WoE and SI susceptibility maps have 76.56% the same content. These results indicate an average 24.5% mismatch in these three maps that require the planner to pay special attention for the construction works using these maps.

Validation and factor importance

The ability of FR, WoE, and SI model in LSM was examined using the ROC-AUC method with success and prediction rate curves. The success rate was produced by comparing the three susceptibility maps with the training dataset (Fig. 10). The result demonstrated that FR had the highest success rate value of 81.51%. It was followed by SI (81.27%) and WoE (79.43%). Although the success rate shows how well a model could fit to the training dataset, the prediction ability of the model cannot be measured by success rate because it is measured by landslides that have

already been utilized for building the model (Bui et al. 2012). In this context, the prediction rate can be used to evaluate the prediction ability of the model. The prediction rates were measured by comparing the landslide susceptibility maps with the testing dataset (Fig. 11). They explain how well the models and conditioning factors predict the landslide. The prediction rate curves showed that the prediction ability of the models is highest for FR (80.44%), followed by SI (79.55%) and WoE (77.94%).

From these results, it is seen that the models employed in this study showed reasonably good accuracy in predicting the landslide susceptibility of the study area. The map produced by FR model exhibited the best result for the purpose of LSM in the study area. When the results of this study are compared with those of reported by other authors it can be stated that the prediction ability of the models are within a similar range. For instance, Mohammady et al. (2012) indicated that the FR model has a success rate of 80.13% and a prediction rate of 75.16%, whereas the WoE model has a success rate of 74.6% and a prediction rate of 69.98%. Regmi et al. (2014) achieved success rates of 76.8, 75.6, and 75.5%, and prediction rates of 75.4, 74.9, and 74.6%, for FR, WoE, and SI, respectively. In a recent paper, Youssef et al. (2016) reported that the FR model has a success rate of 81.3% and a prediction rate of 95%, whereas the WoE model has a success rate of 81.5% and a prediction rate of 95.2%.

Given the highest predictive capability of the FR model compared to the others, this model was used to perform the sensitivity analysis (Table 5). The analysis showed that some factors (i.e., distance from faults and streams, and SPI) were possible source of bias as the AUC values increased when they were omitted from the modeling process. Similarly, excluding slope aspect, plan curvature, and lithology slightly improved the prediction rate

Table 6 Landslide density within the landslide susceptibility maps

Map	Susceptibility class	Number of pixels in class	Number of landslides in class	% of pixels in class	% of landslides in class	Landslides density
FR	Low	161,697	1	23.43	0.95	0.04
	Moderate	182,678	7	26.47	6.67	0.25
	High	176,766	25	25.61	23.81	0.93
	Very high	169,106	72	24.50	68.57	2.80
WoE	Low	168,952	4	24.48	3.81	0.16
	Moderate	175,095	4	25.37	3.81	0.15
	High	178,199	25	25.82	23.81	0.92
	Very high	168,001	72	24.34	68.57	2.82
SI	Low	170,239	2	24.66	1.90	0.08
	Moderate	177,020	7	25.65	6.67	0.26
	High	173,881	22	25.19	20.95	0.83
	Very high	169,107	74	24.50	70.48	2.88

Table 7 Agreed area between FR method, and WoE and SI methods

Susceptibility zones based on the FR method	Susceptibility zones based on the WoE and SI methods	WoE method			SI method		
		Pixel	Percentage	Area (km^2)	Pixel	Percentage	Area (km^2)
Low	Low	3,018,536	89.53	3018.54	2,857,734	95.42	2857.73
	Moderate	352,986	10.47	352.99	132,455	4.43	132.45
	High	0	0	0	4289	0.14	4.29
	Very high	0	0	0	340	0.01	0.34
Moderate	Low	96,511	7.89	96.511	507,944	37.24	507.94
	Moderate	873,828	71.48	873.83	744,403	54.55	744.403
	High	251,610	20.58	251.61	85,560	6.22	58.560
	Very high	657	0.05	657	25,897	1.99	25.89
High	Low	3112	0.64	3.112	5844	0.71	5.84
	Moderate	66,327	13.6	66.33	339,315	41.47	339.31
	High	293,270	60.12	293.27	298,826	36.52	298.83
	Very high	125,050	25.64	125.05	174,196	21.30	174.20
Very high	Low	0	0	0	0	0	0
	Moderate	17,819	0.54	17.82	6433	0.19	6.43
	High	131,862	3.97	131.86	99,084	3.08	99.08
	Very high	3,171,132	95.49	3171.13	3,120,380	96.73	3120.38
Average agreed area		1,839,166	79.15	1839.2	1,755,336	70.80	1755.33

compared to the full FR model. Conversely, the exclusion of slope degree, TWI, STI, TRI, and land use type decreased the AUC values, presumably because they have the most information that was not present in the other factors.

Although we performed a multi-collinearity analysis before model building, the results of the sensitivity analysis revealed that a single multi-collinearity analysis may fail to fully protect a landslide modeling process from including slightly useful conditioning factors. Therefore, an integrated framework of model building and sensitivity analysis is desirable to identify those landslide conditioning factors that either introduce null usefulness to the model performance or decrease the reliability of the produced susceptibility maps (Jaafari et al. 2017a).
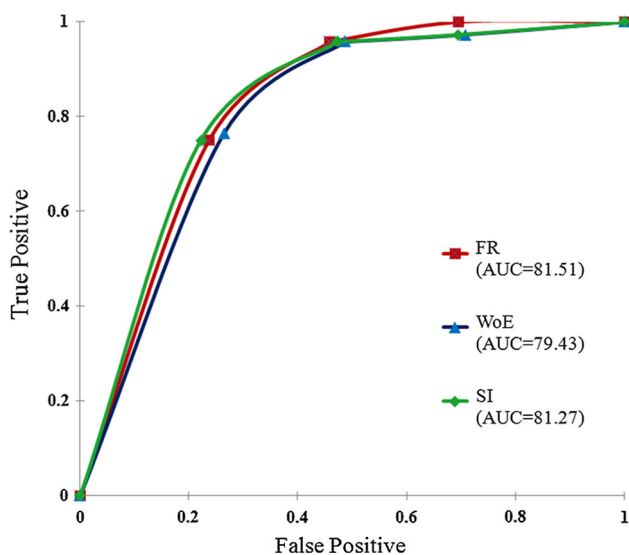

Fig. 10 Success rate curves for the susceptibility maps produced in this study
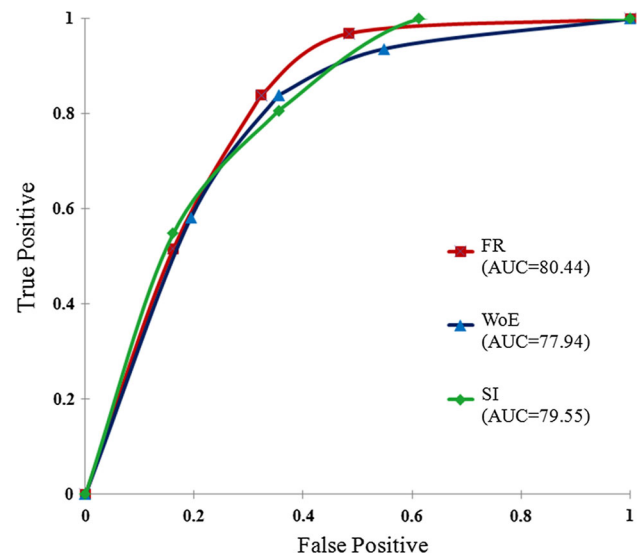

Fig. 11 Prediction rate curves for the susceptibility maps produced in this study

Conclusion

The field of LSM is one of the most popular areas of research. Various methods have been examined for this field of research by numerous researchers. In this study we applied widely accepted models, i.e., FR, WoE, and SI, for the purpose of production of a reliable map of landslide susceptibility. Using these models with the integration of a GIS provides a relatively flexible and easy-to-use framework to spatial prediction of landslides.

The validation results showed that the susceptibility map produced by the FR model has the highest prediction accuracy (80.44%), followed by the SI model (79.55%) and the WoE model (77.94%). Success rate curve also gives similar result, with FR model the highest AUC value (81.51%), followed by the SI model (81.27%) and the WoE model (79.43%). Further, the sensitivity analysis using the FR model revealed that the modeling process was sensitive to input conditioning factors as the exclusion of each of these factors changed the model performance. Overall, this comparative study showed that three models used have performed reasonably well in predicting the landslide susceptibility. These results can indeed greatly help planners and policy makers to adopt appropriate land use planning policies to guide the future developments of infrastructures in the area.

References

Anderson MG, Holcombe E (2013) Community-based landslide risk reduction: managing disasters in small steps. World Bank Publications, Washington

Ardizzone F, Cardinali M, Galli M, Guzzetti F, Reichenbach P (2007) Identification and mapping of recent rainfall-induced landslides using elevation data collected by airborne LiDAR. Nat Hazard Earth Syst Sci 7:637–650

Ayalew L, Yamagishi H (2005) The application of GIS-based logistic regression for landslide susceptibility mapping in the Kakuda-Yahiko Mountains, Central Japan. Geomorphology 65(1–2):15–31

Bijukchhen SM, Kayastha P, Dhital MR (2013) A comparative evaluation of heuristic and bivariate statistical modelling for landslide susceptibility mappings in Ghurmi–Dhad Khola, east Nepal. Arab J Geosci 6(8):2727–2743

Bonham-Carter GF, Agterberg FP, Wright DF (1989) Weights of evidence modelling: a new approach to mapping mineral potential. Stat Appl Earth Sci 89(9):171–183

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

Catani F, Lagomarsino D, Segoni S, Tofani V (2013) Landslide susceptibility estimation by random forests technique: sensitivity and scaling issues. Nat Hazards Earth Syst Sci 13(11):2815–2831

Chung CJF, Fabbri AG (1998) Three Bayesian prediction models for landslide hazard. In: Buccianti A, Nardi G, Potenza R (eds) Proceedings of international association for mathematical geology 1998 annual meeting (IAMG'98), Ischia, Italy, October 1998, pp 204–211

Chung CJF, Fabbri AG (2003) Validation of spatial prediction models for landslide hazard mapping. Nat Hazards 30:451–472

Cruden DM (1991) A simple definition of a landslide. Bull Int Assoc Eng Geol 43:27–29

Dai FC, Lee CF, Xu ZW (2001) Assessment of landslide susceptibility on the natural terrain of Lantau Island, Hong Kong. Environ Geol 40(3):381–391

Darvishzadeh A (2004) Geology of Iran. Amirkabir Publisher, Tehran **(In Persian)**

Ghimire M (2011) Landslide occurrence and its relation with terrain factors in the Siwalik Hills, Nepal: case study of susceptibility assessment in three basins. Nat Hazards 56(1):299–320

Glade T, Crozier MJ (2005) A review of scale dependency in landslide hazard and risk analysis. Landslide hazard and risk. Wiley, Chichester, pp 75–138

Gorsevski PV, Jankowski P (2010) An optimized solution of multi-criteria evaluation analysis of landslide susceptibility using fuzzy sets and Kalman filter. Comput Geosci 36(8):1005–1020

Gorsevski PV, Gessler PE, Jankowski P (2003) Integrating a fuzzy k-means classification and a Bayesian approach for spatial prediction of landslide hazard. J Geogr Syst 5(3):223–251

Gorsevski PV, Jankowski P, Gessler PE (2005) Spatial prediction of landslide hazard using fuzzy k-means and Dempster–Shafer theory. Trans GIS 9(4):455–474

Gorsevski PV, Gessler PE, Boll J, Elliot WJ, Foltz RB (2006) Spatially and temporally distributed modeling of landslide susceptibility. Geomorphology 80(3–4):178–198

Guillard C, Zezere J (2012) Landslide susceptibility assessment and validation in the framework of municipal planning in Portugal: the case of Loures municipality. Environ Manag 50(4):721–735

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):181–216

Hair JF, Black WC, Babin BJ, Anderson RE, Tatham RL (2006) Multivariate data analysis, vol 6. Pearson Prentice Hall, Upper Saddle River, NJ

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

Jaafari A, Najafi A, Pourghasemi HR, Rezaeian J, Sattarian A (2014) GIS-based frequency ratio and index of entropy models for landslide susceptibility assessment in the Caspian forest, northern Iran. Int J Environ Sci Technol 11(4):909–926

Jaafari A, Najafi A, Rezaeian J, Sattarian A (2015a) Modeling erosion and sediment delivery from unpaved roads in the north mountainous forest of Iran. GEM Int J Geomath 6(2):343–356

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

Jaafari A, Mafi Gholami D, Zenner EK (2017a) A Bayesian modeling of wildfire probability in the Zagros Mountains, Iran. Ecol Inform 39:32–44

Jaafari A, Rezaeian J, Omrani MSO (2017b) Spatial prediction of slope failures in support of forestry operations safety. Croat J For Eng 38(1):107–118

Jebur MN, Pradhan B, Tehrany MS (2014) Optimization of landslide conditioning factors using very high-resolution airborne laser

scanning (LiDAR) data at catchment scale. Remote Sens Environ 152:150–165

Kayastha P, Dhital MR, De Smedt F (2012) Landslide susceptibility mapping using the weight of evidence method in the Tinau watershed, Nepal. Nat Hazards 63(2):479–498

Kayastha P, Dhital MR, De Smedt F (2013) Evaluation of the consistency of landslide susceptibility mapping: a case study from the Kankai watershed in east Nepal. Landslides 10(6):785–799

Lee S, Min K (2001) Statistical analysis of landslide susceptibility at Yongin, Korea. Environ Geol 40:1095–1113

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

Mohammady M, Pourghasemi HR, Pradhan B (2012) Landslide susceptibility mapping at Golestan Province Iran: a comparison between frequency ratio, Dempster–Shafer, and weights-of evidence models. J Asian Earth Sci 61:221–236

Montgomery DR, Dietrich WE (1994) A physically based model for the topographic control on shallow landsliding. Water Resour Res 30(4):1153–1171

Nami MH, Jaafari A, Fallah M, Nabiuni S (2017) Spatial prediction of wildfire probability in the Hyrcanian ecoregion using evidential belief function model and GIS. Int J Environ Sci Technol. doi:10.1007/s13762-017-1371-6

Nefeslioglu HA, Gokceoglu C, Sonmez H (2008) An assessment on the use of logistic regression and artificial neural networks with different sampling strategies for the preparation of landslide susceptibility maps. Eng Geol 97(3):171–191

Oh HJ, Kim YS, Choi JK, Park E, Lee S (2011) GIS mapping of regional probabilistic groundwater potential in the area of Pohang City. Korea J Hydrol 399(3):158–172

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

Pham BT, Pradhan B, Bui DT, Prakash I, Dholakia MB (2016) A comparative study of different machine learning methods for landslide susceptibility assessment: a case study of Uttarakhand area (India). Environ Model Soft 84:240–250

Pham BT, Bui DT, Prakash I, Dholakia MB (2017) Hybrid integration of multilayer perceptron neural networks and machine learning ensembles for landslide susceptibility assessment at Himalayan area (India) using GIS. CATENA 149:52–63

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

Pourghasemi HR, Moradi HR, Aghda SF (2013) Landslide susceptibility mapping by binary logistic regression, analytical hierarchy process, and statistical index models and assessment of their performances. Nat Hazards 69(1):749–779

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. Comput Geosci 51:350–365

Regmi AD, Devkota KC, Yoshida K, Pradhan B, Pourghasemi HR, Kumamoto T, Akgun A (2014) Application of frequency ratio, statistical index, and weights-of-evidence models and their comparison in landslide susceptibility mapping in Central Nepal Himalaya. Arab J Geosci 7(2):725–742

van Westen C (1997) Statistical landslide hazard analysis. ILWIS 2.1 for Windows application guide. ITC Publication, Enschede, pp 73–84

Wu W, Sidle RC (1995) A distributed slope stability model for steep forested basins. Water Resour Res 31(8):2097–2110

Xu C, Xu X, Dai F, Saraf AK (2012) Comparison of different models for susceptibility mapping of earthquake triggered landslides related with the 2008 Wenchuan earthquake in China. Comput Geosci 46:317–329

Yalcin A, Reis S, Aydinoglu AC, Yomralioglu T (2011) A GIS-based comparative study of frequency ratio, analytical hierarchy process, bivariate statistics and logistics regression methods for landslide susceptibility mapping in Trabzon, NE Turkey. CATENA 85(3):274–287

Yesilnacar EK (2005) The application of computational intelligence to landslide susceptibility mapping in Turkey. Ph.D. thesis, Department of Geomatics the University of Melbourne, 423 pp

Youssef AM, Pourghasemi HR, El-Haddad BA, Dhahry BK (2016) Landslide susceptibility maps using different probabilistic and bivariate statistical models and comparison of their performance at Wadi Itwad Basin, Asir Region, Saudi Arabia. Bull Eng Geol Environ 75(1):63–87

Zweig MH, Campbell G (1993) Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. Clin Chem 39(4):561–577