



An innovative, fast method for landslide susceptibility mapping using GIS-based LSAT toolbox

Ali Polat¹

Received: 5 August 2020 / Accepted: 23 December 2020 / Published online: 6 March 2021
© The Author(s), under exclusive licence to Springer-Verlag GmbH, DE part of Springer Nature 2021

Abstract

In this study, landslide susceptibility maps (LSM) of the Akıncılar region were produced with the methods of frequency ratio (FR), information value (IV), logistic regression (LR), random forest (RF), and multi-layer perceptron (MLP) by using a new GIS-based toolbox (LSAT, Landslide Susceptibility Assessment Tool). LSAT was used to assess the landslide susceptibility of the Akıncılar region located 150 km northwest of Sivas city (Turkey). LSM was successfully constructed using five different methods for the study area. Area under the curve (AUC) values were calculated as 70.95%, 71.85%, 72.57%, 72.67%, 73.93% for prediction rate of FR, IV, LR, MLP and RF methods, respectively. Time-consuming processes are one of the significant problems of constructing LSM. LSAT can be used easily in this type of study and minimizes such problems. Data preparation processes, visualization of modeling results, and accuracy assessment of LSM could very quickly and automatically be done thanks to this tool.

Keywords Landslide susceptibility · Python scripting · Akıncılar · Weka · Geographic information system (GIS)

Abbreviations

ANN	Artificial neural network
AUC	Area under curve
DEM	Digital elevation model
DT	Decision tree
FIS	Fuzzy inference system
FR	Frequency ratio
GIS	Geographic information system
IV	Information value
LR	Logistic regression
LS factor	Slope Length and Steepness Index
LSAT	Landslide Susceptibility Assessment Tool
LSM	Landslide susceptibility map
M-AHP	Modified analytical hierarchy process
MLP	Multi-layer perceptron
NAF	North Anatolian Fault
NAFZ	North Anatolian Fault Zone
NDVI	Normalized Difference Vegetation Index
RF	Random forest
ROC	Receiver operating characteristic
SVM	Support vector machines

TRI	Topographic Ruggedness Index
TWI	Topographic Wetness Index
WofE	Weight of evidence

Introduction

Natural disasters cause the loss of many lives and property in the world. Some of the most occurring phenomena like earthquakes, floods, landslides, and avalanche are often observed in Turkey due to geological, geomorphological, and climatic characteristics. Landslides are the most devastating disasters after the earthquakes in Turkey (Gökçe et al. 2008). Therefore, it is crucial to predict landslide-prone areas. For this purpose, the landslide susceptibility map (LSM) is prepared. The main objective is to make predictions for the future using existing data with various methods (statistical or machine learning) (Dikshit et al. 2020).

Spatial prediction of landslides is the probability of potential instability of slopes related to a set of causal factors (Guzzetti et al. 2005). It can be carried out by analyzing the spatial relationship between past landslide events and a set of geo-environmental factors (Pham et al. 2016). It assumes that future landslides will occur under the same conditions as previous landslides (Ermini et al. 2005). Previous landslides are identified, and a landslide inventory map

✉ Ali Polat
ali.polat@afad.gov.tr

¹ Provincial Directorate of Disaster and Emergency Management of Turkey, 58050 Sivas, Turkey

is prepared (Guzzetti et al. 2012). Dependent and independent variables are determined. Dependent variables are landslide occurrences, and independent variables are determined by considering the geographical and geological conditions of the region. LSM is an outcome of the spatial prediction of landslides.

Geographic information system (GIS) based studies are used to produce LSM. Many methods have been applied for preparing LSM using GIS in recent years. These methods are expert opinion-based and data mining-based approaches (Song et al. 2012). The expert opinion-based approach is based on the perspective of experts in selecting variables and assigning weights to variables. This susceptibility modeling is one of the most suitable approaches for some regions with insufficient data in terms of spatial resolution and coverage (Ozer et al. 2020; Polat and Erik 2020). Data mining-based approach uses machine learning algorithms to determine factors leading to landslide occurrences and calculate weights of the factors during the learning of models. Out of these approaches, the data mining-based approach is more commonly used for creating LSM. Commonly used methods are logistic regression (LR) (Lee et al. 2007; Nefeslioglu et al. 2008; Xu et al. 2012; Devkota et al. 2013), decision tree (DT) (Nefeslioglu et al. 2010; Tien et al. 2012; Pradhan 2013; Hong et al. 2015), artificial neural network (ANN) (Das et al. 2013; Zare et al. 2013; Nourani et al. 2014; Dou et al. 2015) and support vector machines (SVM) (Tien et al. 2015; Hong et al. 2016; Chen et al. 2017). Frequency ratio (FR) and information value (IV) methods, which are bivariate statistical analysis methods, are frequently used in LSM studies (Pourghasemi et al. 2012a, b; Umar et al. 2014; Aghdam et al. 2016; Cui et al. 2016).

ArcGIS (ESRI 2011) is a widely used GIS software in many fields. It supports Python and Visual Basic scripting. Jimenez-Peralvarez et al. (2009), have used Model Builder in ArcGIS (ESRI) for landslide-susceptibility assessment. Zhu (2010) developed a risk evaluation model for Earthquake-induced landslide with the help of a Model Builder. Zhang et al. (2014) proposed a scripting model called GIScript, which presents a conceptual framework of directing access to geospatial data and parallel processing. Dobesova (2011) has discussed the relevance of the Python programming language for data processing in ArcGIS. Luo et al. (2011) have taken up the application of Python language and ArcGIS software into analyzing the Gorges Reservoir Area Remote Sensing data. Akgün et al. (2012) developed an easy-to-use program “MamLand” for the construction of LSM. This program was employed in Matlab based on expert opinion creates LSM using Mamdani fuzzy inference system. In Torizin (2012), an ArcGIS-toolbox ‘Landslide Susceptibility Assessment Tools’ (LSAT) was presented. This toolbox includes the FR and weight of evidence (WofE) as well as the multivariate LR. A single and standalone application

for constructing LSM was developed by Osna et al. (2014). Jebur et al. (2015) proposed a tool applied in the ArcMap. This tool constructs LSM using FR, WofE, and evidential belief function methods. A toolbar was created thanks to the software development kit available with ArcGIS v.10 by Palamakumbure et al. (2015). This tool includes data preparation, model optimizing, derivation of decision trees, calculating predictions, and validation tasks. Sezer et al. (2016) developed an expert-based LSM module for Netcad Architect Software. This module uses the methods of modified analytical hierarchy process (M-AHP) and Mamdani type fuzzy inference system (FIS). In Arca et al. (2018), a Python script, integrated within the ArcGIS, was developed to perform all the calculations and plot drawings of LSM. Narayanan and Sivakumar (2018) created a Python-based customized toolbox as an extension tool to ArcGIS that helps to understand the spatial subsurface events through the development of the seismic information system. Bragagnolo et al. (2019) created a tool written in the python language named r.landslide and was integrated within Qgis. Köse and Turk (2019) developed user interface programs creating LSM and calculating accuracy for methods of FR and WofE in the GIS environment.

In this study, landslide susceptibility maps related to the Akıncılar region were generated by using five methods (FR, IV, RF, MLP, LR) on the LSAT. The Weka tool (ver.3.8) (Eibe et al. 2016) was used to compare results (for LR method) and test for dataset produced by LSAT.

The main contributions of this paper are as follows:

- LSM of Akıncılar region is created automatically with high performance.
- A novel toolbox called LSAT is developed.
- LSM is done quickly and easily thanks to LSAT.
- Best model parameters are selected with the use of tuning scripts.

Materials

Study area

The study area is located 150 km northeast of Sivas city (Fig. 1). This region covers an area of approximately 319.886 km². Middle of the basin has an approximately flat area, but south margin and north margin have about 60° slope values. The elevation is low in the middle sections. Contrary, there are high mountainous areas in the north and south.

The North Anatolian Fault Zone (NAFZ) is the primary continental strike-slip fault system in Turkey. The study area is located on NAFZ. Due to this configuration of the region, landslides are observed frequently. The study area is also a

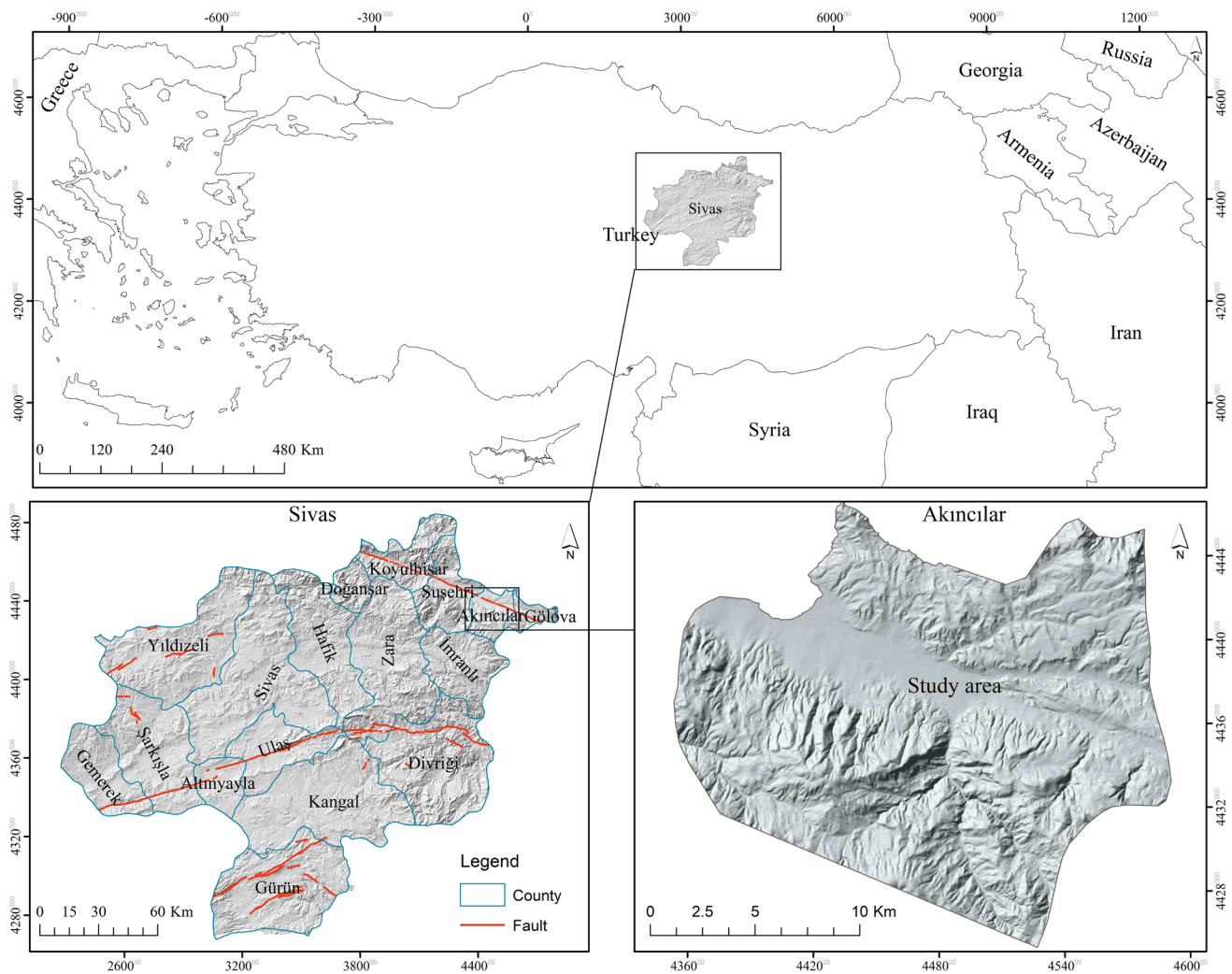


Fig. 1 Location map of the study area

part of the Suşehri pull-apart basin formed by NAFZ (Polat et al. 2014). NAF passes through the middle of the basin (Fig. 2). The 1939 Erzincan Earthquake ($M = 7.8$) occurred on the NAFZ and created a 360-km-long surface rupture. The recent landslide that happened in this fault zone is the Kuzulu landslide (Gökçeoğlu et al. 2005; Tatar et al. 2005; Ulusay et al. 2007; Polat and Gürsoy 2014). 15 people died due to occurred this landslide in the northwest of the Koyulhisar on 17 March 2005.

According to the geological map prepared by the General Directory of Mineral Research and Exploration (Akbaş et al. 1991), a large part of the study area consists of Upper Miocene–Pliocene age conglomerate–sandstone–mudstone. Quaternary aged alluvium and Eocene aged andesite–basalt are the other most observed units in the region. Younger units are observed in the middle section of the study area, while older units are seen in the north and south edges (Fig. 2). As shown in Fig. 2, the great majority of landslides

are observed in units of conglomerate-sandstone and mudstones.

Installing required libraries

LSAT works on ArcGIS software with the Windows platform. Before using the LSAT, the required installations must be done. Python 2.7 is already installed with ArcGIS 10.4. The libraries must be compatible with the version of Python 2.7. Pip is recommended for easy installation. If a different version of python is installed, pip2.7 must be used. Scikit-learn (Pedregosa et al. 2011) is the main library for our tasks. Numpy (version of 1.8.2 or higher) and SciPy (version of 0.13.3 or higher) must be installed before installing Scikit-learn. Python 2.7 supports the versions of Scikit-learn 0.20 and earlier. In this study, the version of 0.20.2 was installed. Also, Numpy 1.15.4, Pandas 0.16.1, and Matplotlib 1.4.3

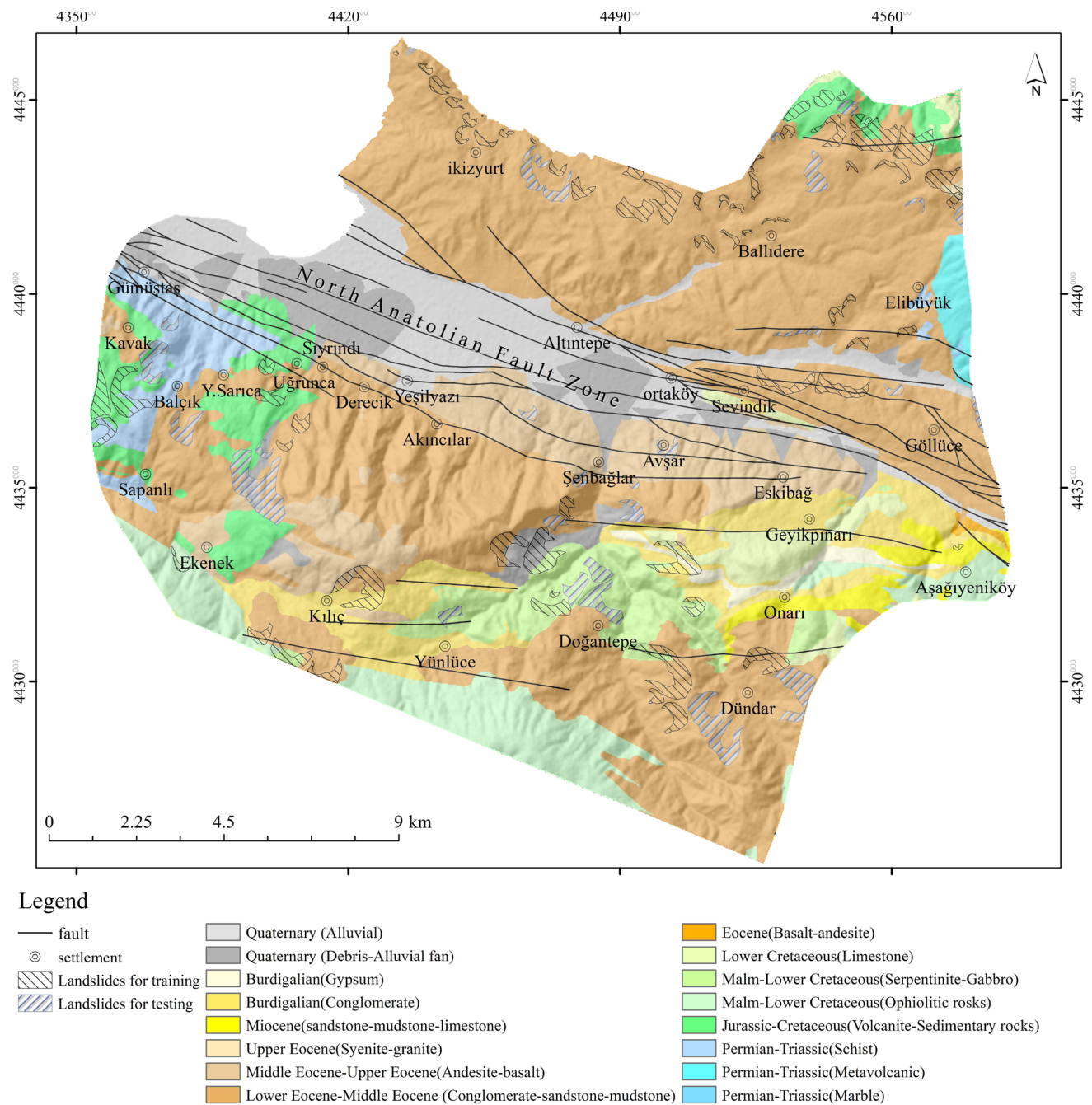


Fig. 2 Geological map of the study area and training landslides and test landslides

were installed. In addition to these, the C++ compiler must be installed for windows.

Preparing data

Area boundary data, controlling factors (parameters) data, and landslide inventory data are required as input parameters to create an LSM. All required parameter data were created and reclassified in GIS environment. Various sampling

methods are applied to create LSM (Dagdelenler et al. 2016). In this study, scarp of landslides was used as landslide data. These data were created manually in polygon type. A Python script was written to do the next steps. The name of the script is preparing data. Preparing data script processes the data and prepares it for analysis. Figure 3 shows the flow diagram of these processes. This script needs to landslide inventory data, area boundary data, and reclassified controlling factors data, as mentioned above. The outputs of

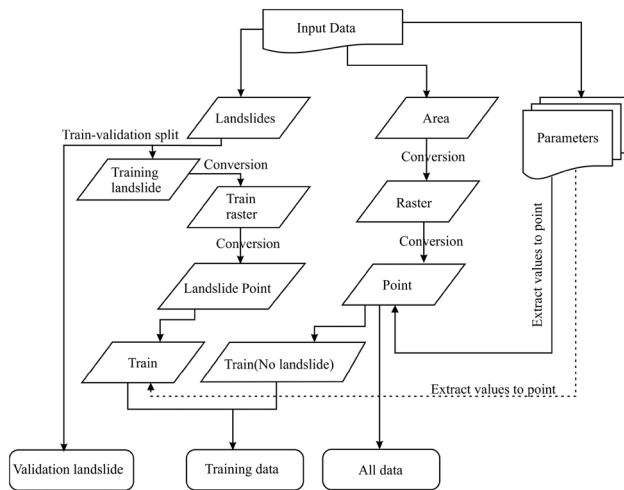


Fig. 3 Data preparation flow chart

the script are landslide validation data, landslide training data, and all parameter data. Training data and all data will be used in the analysis processes. Landslide validation data will be used for validating the performance of prediction.

The area, first input parameter, was used to clip input raster data. Also, the values of controlling factors were extracted through this parameter. Users may select any desired region for analysis. In this study, the Akincilar county boundary was used as area data. Area data was converted to point. This data consists of 511,828 pixels. All parameter pixel values were extracted and assigned to this point data.

Data preparation script splits landslide inventory data as training and test. A total of 100 landslide data were obtained in the study area. Random selection is made based on polygon-type landslide. There is no standard value for splitting data as a training set and test set. 70% of landslides data were selected as training, and 30% of landslides data were selected as validation randomly (Chen et al. 2017; Arabameri et al. 2019; Nohani et al. 2019). Users can also select the splitting ratio on the user interface.

All training pixel values were converted to point data. “1” value was assigned to landslide pixels, and “0” value assigned to the no-landslide pixels. Training landslide pixel count is 24,015 for this study area. It was preferred to use the same amount of no-landslide pixels. Therefore, 24,015 no-landslide pixels were selected randomly. So, 48,030 pixels were generated as training data. All parameter pixel values were extracted from this point data. Then, FR and IV values of all parameters were calculated. In this study, IV values of parameters were used as output data. This data can be used in two ways. External tools or our tool can be used to predict landslide probability. If an external tool or software was used, LSM could be constructed, and ROC values can

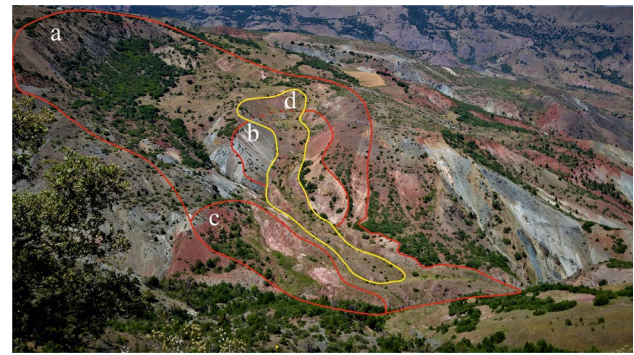


Fig. 4 Landslide image from study area: (a) main landslide body, (b), (c) and (d) secondary landslides

be calculated with the “Create LSM and Calculate ROC” script in our tool.

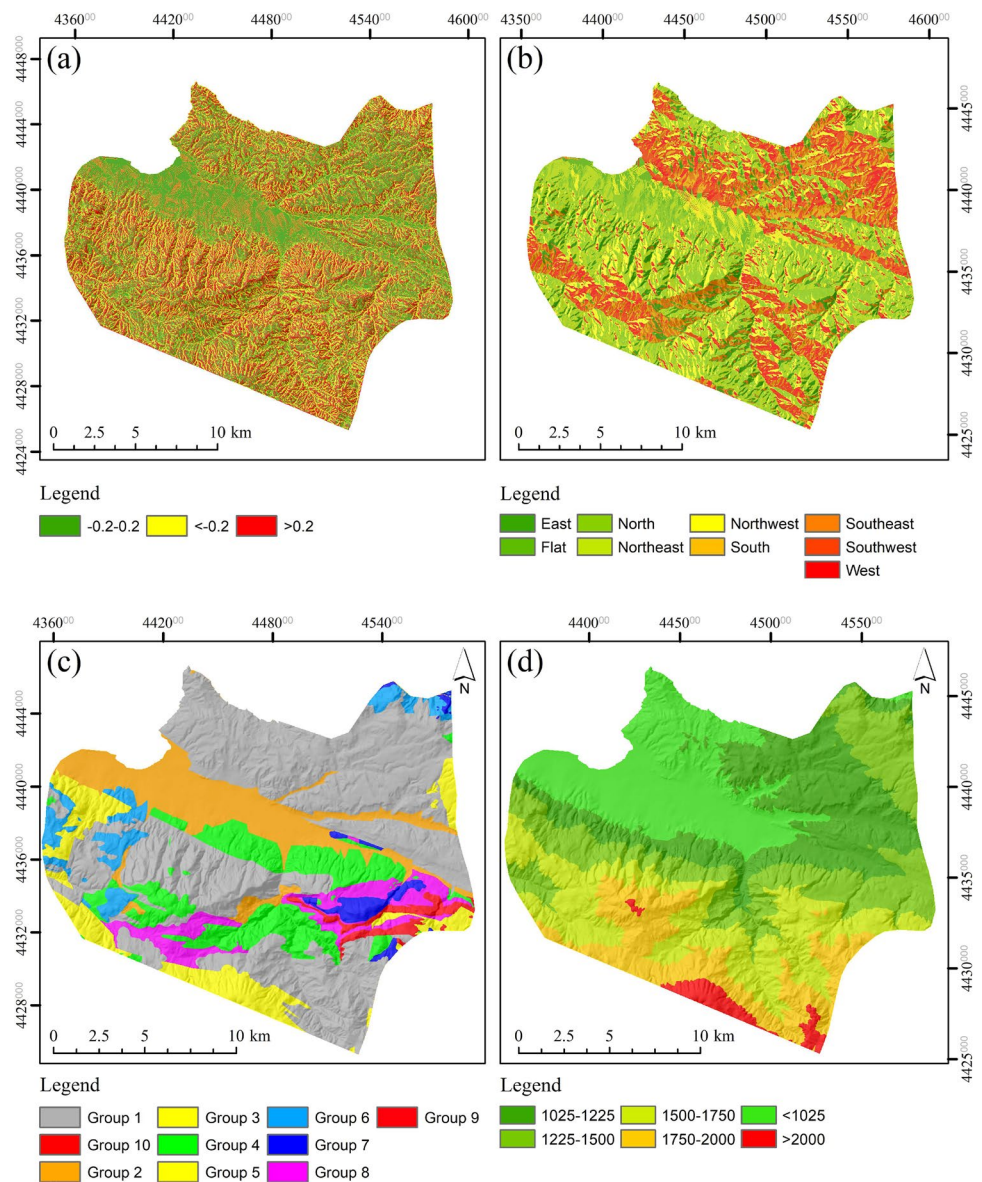
Using the accurate data is very important in this type of analysis. Therefore, detailed field studies were carried out for defining landslide locations. In the study area, mostly complex type landslides were observed. Large-scale landslides and small-scale secondary landslides developed within them were seen in the study area (Fig. 4). The movement started as a rotational slip in areas higher elevation and close to the ridge. Then, it turned into a flow type due to the high slope. Scarps of all landslides were used, and toes of the landslides were ignored in the analysis.

Another important criterion is to select landslide controlling factors. The most used controlling factors in the literature have been selected for the study area considering the geographic and geological location of the region (Hasekioğulları and Ercanoğlu 2012). Ten controlling factors were used in this study. These are elevation, aspect, curvature, slope, geology, distance to ridge, Normalized Difference Vegetation Index (NDVI), Topographic Ruggedness Index (TRI), Topographic Wetness Index (TWI), Slope Length and Steepness Index (ls factor) (Fig. 5). Distance to faults parameter was reclassified as <500 m, 500–1000 m, 1000–1500 m, 1500–2000 m, 2000–2500 m, 2500–3000 m, >3000 m. Analyses were made with these values. However, the performance decreased, and it caused the overfitting problem. So, this parameter was not used in analyses.

In this study, ArcGIS and Saga-GIS (Conrad et al. 2015) software were used for GIS studies. Firstly, 25 × 25 m resolution DEM data was created from 1/25,000 scaled topographical map in ArcGIS. Slope angle, aspect angle, curvature data were derived from DEM by ArcGIS and TWI, TRI, LS factor, ridge data were obtained from Saga-GIS. NDVI was calculated by using Landsat 8 satellite image in ArcGIS. The parameters used in the analyses are explained in detail below.

Curvature: The curvature is a parameter that defines the morphology of the topography, and the terrains are classified

Fig. 5 Landslide controlling factors: **a** curvature, **b** aspect, **c** lithology, **d** elevation, **e** Normalized Difference Vegetation Index (NDVI), **f** Slope Length and Steepness Index (ls factor), **g** slope, **h** distance to ridge, **i** Topographic Ruggedness Index (TRI), **j** Topographic Wetness Index (TWI)



as concave, convex, and flat. Curvature is a useful factor in the landslide occurrences. For example, flow-type landslides are more likely to occur in convex areas. Curve data were reclassified into three classes. It is assumed that the curve values lower than -0.2 are concave, curve values higher than 0.2 are convex, and curvature values between these values are flat (Magioni and Gruber 2003).

Slope aspect: The slope aspect describes the direction of the slope. The slope direction is related to the amount of precipitation and sun exposure. These two properties affect slope stability. In this study, this parameter was reclassified into nine classes. The FR of the flat area is minimum. The maximum FR is observed in the west direction.

Lithology: Lithology can provide valuable information for landslide prediction studies. Lithological properties are directly related to many properties of materials

such as strength, permeability, and hardness (Baeza and Corominas 2001). Therefore, lithological features of the area should be appropriately evaluated, and the convenient classification value should be selected. Geological units were simplified as ten groups in the study area (Table 1). Lithological units with similar characteristics were placed in the same groups. The final decision was made by studying the relationship between landslides and lithological units in the geological map.

Elevation: Elevation value is frequently used in this type of studies. The minimum height is 832 m, and the maximum height is 2230 m in the study area. These values were classified as six classes, as in Table 2. A linear relationship was observed between the elevation and possibility of landslide until 2000 m. The FR value is zero over 2000 m.

Fig. 5 (continued)

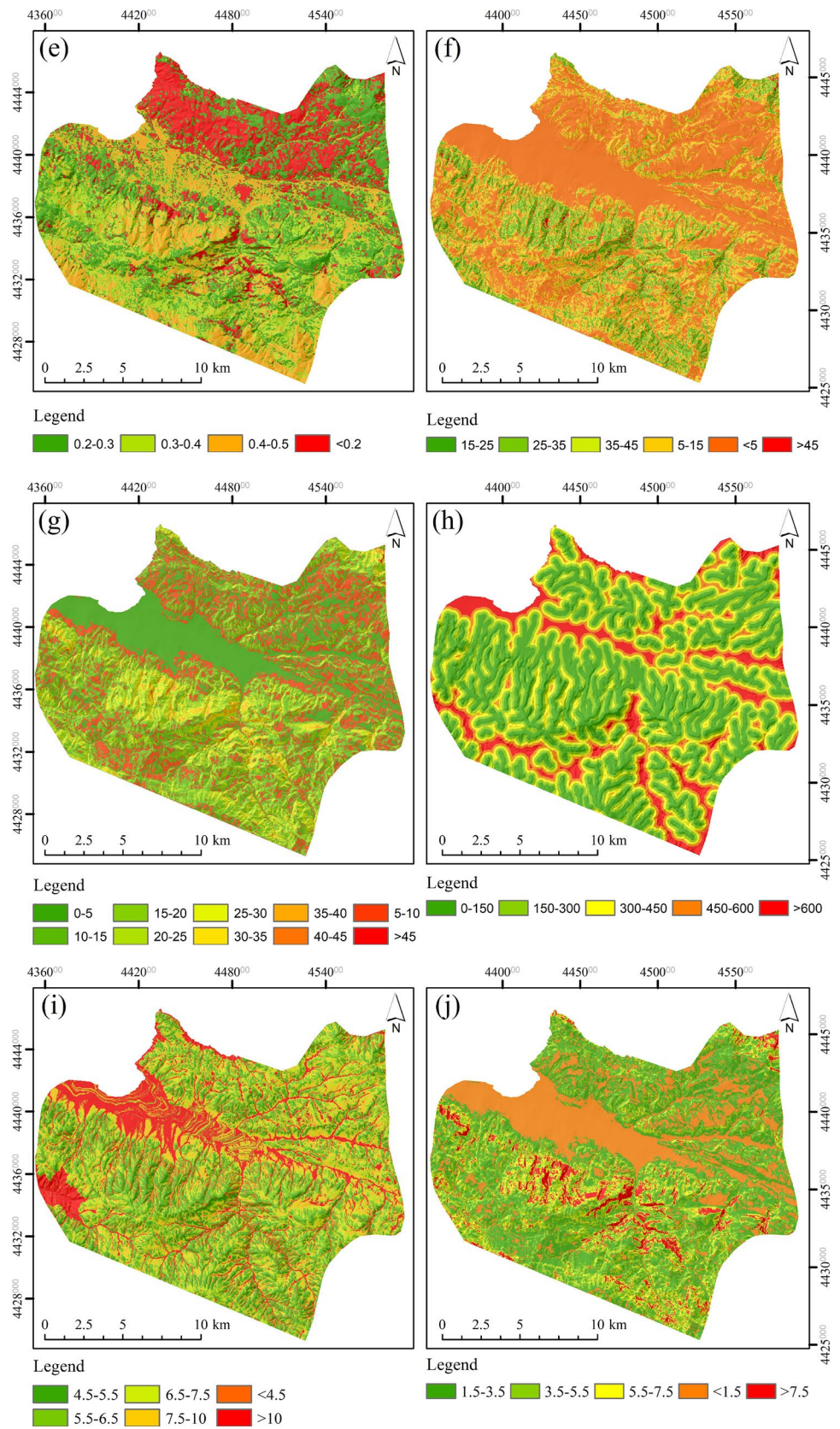


Table 1 Simplified geological units

Simplified groups	Units in geological map
Group 1	Conglomerate–sandstone–mudstone
Group 2	Debris–alluvial fan, alluvium
Group 3	Marble, metavolcanics, Schist
Group 4	Andesite–basalt, syenite–granite
Group 5	Ophiolitic rocks
Group 6	Volcanite–sedimentary rocks
Group 7	Limestone
Group 8	Conglomerate
Group 9	Sandstone–mudstone–limestone
Group 10	Gypsum

Normalized Difference Vegetation Index (NDVI): Plants have the effect of stopping or increasing movement on slopes. Stopping effect is observed, especially in shallow landslides, and the increasing effect is observed in a deep landslide (Eker and Aydın 2014). The Normalized Difference Vegetation Index (NDVI) is one of the frequently used environmental factors in the assessment of landslide susceptibility. The red band and the Near-Infrared band (NIR) are used for calculating NDVI (<http://www.esri.com>). Landsat 8 image acquired from USGS Earth Explorer was used for calculating NDVI. NDVI is calculated as follows:

$$NDVI = \frac{NIR - Red}{NIR + Red}, \quad (1)$$

where NDVI is Normalized Difference Vegetation Index, NIR is Near Infrared Band (band 5), and Red is the red band (band 4) of Landsat 8 image. NDVI data were classified into four classes, as shown in Table 2.

Slope length and slope angle (LS-factor): LS-factor has a considerable effect on soil loss. The L-factor is the impact of slope length, and the S-factor is the effect of slope steepness. These two factors are significant for landslide occurrence. The LS-factor calculation was performed using the equation proposed by Boehner and Selige (2006) and implemented using the System for Automated Geoscientific Analyses (SAGA) software. LS data were classified into six classes. The values with the highest landslide density are between 5 and 15, and as shown in Table 2.

Slope angle: Slopes have a significant influence on landslide occurrence as they affect the flow of water and the soil (Jones et al. 1983). The slope angle was reclassified into ten classes. The maximum FR was obtained between 15°–20°.

Distance to ridge: It has been determined that landslides occur more frequently in areas close to the ridges. Therefore, the distance to the ridge parameter was used as a landslide controlling factor. The landslide scarps are also commonly observed in areas close to the ridge. Ridge data were

classified at intervals of 150 meters. As shown in Table 2, the maximum FR is between 150 and 300 m.

Topographic Wetness Index (TWI): TWI is a tool indicating areas accumulating water flow, often with the seasonally and permanently waterlogged ground. As such, it is beneficial to show the geomorphic complexity of a landslide terrain, including the pattern of topographic highs ('dry' areas) and lows ('wet' areas) (Rozycka et al. 2016). TWI is defined by the following equation:

$$TWI = \ln A / \tan \beta, \quad (2)$$

where A is the upslope contributing area, and $\tan \beta$ is the local slope (Moore et al. 1991; Sorensen et al. 2006; Gruber and Peckham 2009). TWI data were classified into six classes, as shown in Table 2.

Topographic Ruggedness Index (TRI): TRI is a measurement developed by Riley et al. (1999). TRI corresponds to average elevation differences between any point on a grid and its surrounding area (Rozycka et al. 2016). TRI is expressed as follow:

$$TRI = \left(\sum (Z_c - Z_i)^2 \right)^2, \quad (3)$$

where Z_c is the elevation of a central cell, and Z_i is the elevation of one of the eight neighbouring cells ($i = 1, 2, \dots, 8$). TRI data were classified into five classes, as shown in Table 2.

Methods

In this study, GIS, statistics, and machine learning methods were used. GIS-based studies were used for preparing data. Model results were processed in a GIS environment to display landslide susceptibility index (LSI). Weka tool and Python (Sklearn library) were used for creating a classification model and analysis. LSAT contains 10 Python scripts that were written by the author and integrated into the ArcGIS. The first script is used to prepare data for modeling. The second script is used to create LSM and to calculate the area under curve (AUC) values. The other scripts are used to construct LSM with the methods of FR, IV, LR, Random Forest (RF), and multi-layer perceptron (MLP). Data preparation script was tested to predict landslide probability in Weka software with the LR algorithm. Then classification results were processed with create LSM and calculate ROC script in GIS and susceptibility map was created.

Frequency ratio method

FR method proposed by Lee and Talib (2005) is based on density analyses. FR is the ratio of the area where landslides

Table 2 Frequency ratios and information values of landslide conditioning parameters

Factors	Classes	Landslide occurrence grids	Landslide occurrence grids (%)	Grids in domain	Grid (%)	Frequency ratio (FR)	Information value (IV)
Aspect							
1	Flat	156	0.74	20,836	4.07	0.18279	- 1.699404
2	North	4243	20.23	102,184	19.96	1.013767	0.013673
3	Northeast	3356	16.00	86,719	16.93	0.944835	- 0.056745
4	East	1889	9.01	45,440	8.87	1.014943	0.014833
5	Southeast	1396	6.66	29,629	5.79	1.150314	0.140035
6	South	1547	7.38	35,954	7.02	1.050488	0.049255
7	Southwest	2398	11.43	54,752	10.69	1.069294	0.066999
8	West	2504	11.94	54,677	10.68	1.118092	0.111624
9	Northwest	3485	16.62	81,879	15.99	1.03915	0.038403
Slope angle (°)							
1	< 5	591	2.82	110,469	21.57	0.130616	- 2.035496
2	5–10	3548	16.92	125,163	24.44	0.692079	- 0.368055
3	10–15	6810	32.47	112,694	22.01	1.475348	0.388894
4	15–20	4745	22.62	72,857	14.23	1.590058	0.463771
5	20–25	2634	12.56	43,691	8.53	1.471878	0.386539
6	25–30	1481	7.06	26,380	5.15	1.370656	0.315289
7	30–35	790	3.77	14,264	2.79	1.352178	0.301717
8	35–40	325	1.55	5444	1.06	1.457516	0.376734
9	40–45	48	0.23	946	0.18	1.238791	0.214136
10	>45	2	0.01	162	0.03	0.301414	- 1.199271
Elevation (m)							
1	< 1025	1405	6.70	111,723	21.82	0.307031	- 1.180808
2	1025–1225	4168	19.87	115,240	22.50	0.883024	- 0.124403
3	1225–1500	6303	30.05	108,912	21.27	1.412927	0.345663
4	1500–1750	5925	28.25	98,687	19.27	1.465806	0.382405
5	1750–2000	3173	15.13	65,898	12.87	1.175563	0.161747
6	> 2000	0	0.00	11,610	2.27	0	0
Curvature							
1	< -0.2	7587	36.17	152,236	29.73	1.216748	0.196182
2	- 0.2 to 0.2	6296	30.02	195,589	38.20	0.785902	- 0.240923
3	> 0.2	7091	33.81	164,245	32.07	1.054055	0.052645
Lithology							
1	Group 1	12,609	60.12	255,721	49.94	1.205312	0.186738
2	Group 2	903	4.31	70,560	13.78	0.312834	- 1.162082
3	Group 3	529	2.52	18,593	3.63	0.695491	- 0.363138
4	Group 4	2139	10.20	64,743	12.64	0.807613	- 0.213673
5	Group 5	212	1.01	27,747	5.42	0.186769	- 1.677882
6	Group 6	3088	14.72	23,665	4.62	3.189745	1.159941
7	Group 7	257	1.23	10,750	2.10	0.584399	- 0.53717
8	Group 8	1023	4.88	30,254	5.91	0.826567	- 0.190474
9	Group 9	27	0.13	7082	1.38	0.093195	- 2.37306
10	Group 10	181	0.86	3442	0.67	1.285443	0.251104
Distance to ridge (m)							
1	0–150	7137	34.03	182,767	35.69	0.953404	- 0.047716
2	150–300	7429	35.42	140,358	27.41	1.292267	0.256398
3	300–450	4163	19.85	95,120	18.58	1.068547	0.0663
4	450–600	1694	8.08	54,063	10.56	0.765019	- 0.267854

Table 2 (continued)

Factors	Classes	Landslide occurrence grids	Landslide occurrence grids (%)	Grids in domain	Grid (%)	Frequency ratio (FR)	Information value (IV)
5	> 600	551	2.63	39,775	7.77	0.338221	- 1.084056
NDVI							
1	< 0.2	3739	17.83	112,325	21.94	0.812081	- 0.208155
2	0.2–0.3	11,240	53.59	191,132	37.33	1.434675	0.360938
3	0.3–0.4	4426	21.10	121,141	23.66	0.891335	- 0.115035
4	> 0.4	1569	7.48	87,086	17.01	0.439537	- 0.822034
TRI							
1	< 1.5	864	4.12	124,912	24.39	0.168872	- 1.778614
2	1.5–3.5	8212	39.15	189,217	36.95	1.059588	0.05788
3	3.5–5.5	7039	33.56	111,193	21.71	1.545545	0.435376
4	5.5–7.5	3036	14.48	52,725	10.30	1.405831	0.340629
5	> 7.5	1823	8.69	34,023	6.64	1.308164	0.268624
TWI							
1	< 4.5	1071	5.11	23,890	4.67	1.094514	0.090311
2	4.5–5.5	5313	25.33	101,250	19.77	1.281129	0.247742
3	5.5–6.5	6426	30.64	125,299	24.47	1.252106	0.224827
4	6.5–7.5	4061	19.36	80,641	15.75	1.22949	0.2066
5	7.5–9	3527	16.82	93,133	18.19	0.924592	- 0.078403
6	> 9	576	2.75	87,857	17.16	0.160064	- 1.83218
LS factor							
1	< 5	5267	25.11	255,686	49.93	0.502926	- 0.687311
2	5–15	10,401	49.59	166,927	32.60	1.521236	0.419523
3	15–25	3206	15.29	52,982	10.35	1.477349	0.39025
4	25–35	1326	6.32	22,197	4.33	1.458469	0.377388
5	35–45	534	2.55	9200	1.80	1.417103	0.348615

occurred in the total study area and the ratio of the probabilities of a landslide occurrence to a non-occurrence for a given attribute (Bonham-Carter 1994). The proportion of landslide occurrence to non-occurrence (Regmi et al. 2014) was calculated for each parameter. The following equation defines the FR:

$$FR = \frac{N_{\text{pix}(si)}/N_{\text{pix}(Ni)}}{\sum N_{\text{pix}(si)}/\sum N_{\text{pix}(Ni)}}. \quad (4)$$

$N_{\text{pix}(si)}$: The number of pixels containing landslide in class (i).

$N_{\text{pix}(Ni)}$: Total number of pixels having class (i) in the whole area.

$\sum N_{\text{pix}(si)}$: Total number of pixels containing landslide.

$\sum N_{\text{pix}(Ni)}$: Total number of pixels in the whole area.

In this method, it is stated that FR values greater than 1 are relatively more effective in landslide occurrence, and FR values less than 1 have less effect in landslide occurrence

(Lee and Talib 2005). The FR method was implemented with the FR script in a GIS environment. The obtained FR values are assigned to parameters and subgroups to construct LSM in the GIS environment. The FR value was calculated, as shown in Table 2. This script requires data created with the Data Preparation script.

Information value method

The IV method, a statistical approach, was used for calculating the weight of parameters like FR. Yin and Yan (1988) suggested this method. The weight value is defined as the natural logarithm of the ratio of the landslide density in the relevant parameters and subgroups to the landslide density in the total area (Van Westen et al. 1997). The following equation defines the weight:

$$W_i = \ln \frac{\frac{N_{\text{pix}}Si}{N_{\text{pix}}Ni}}{\sum \frac{N_{\text{pix}}Si}{N_{\text{pix}}Ni}}. \quad (5)$$

W_j :	weight of parameter
$N_{pix}Si$:	number of pixels of the landslide within class i
$N_{pix}Ni$:	number of pixels of class i
$\sum N_{pix}Si$:	number of pixels of the landslide within the all study area
$\sum N_{pix}Ni$:	number of pixels of the all study area.

Negative values of weight show lower density, and the positive values of weight show a higher density of landslide relatively (Van Westen 1997).

Any GIS software can determine the pixel number of parameter and their distribution in landslide pixels. Data preparation script was used to calculate weight with the IV method. Then IV script was used to create LSM and calculate AUC value. The AUC value of the IV method is higher than the FR method. So, the weights of the IV method were used as the input parameter for the LR, RF, and MLP algorithms.

Logistic regression method

LR is a method frequently used in the literature in recent years (Chen et al. 2018, 2019; Benchelha et al. 2019; Sevgen et al. 2019; Sahin et al. 2020). The dependent variable is categorical. Landslide pixels are represented by 1, and non-landslide pixels are represented by 0. Parameters are independent variables. LR is a method used to determine the cause-effect relationship with independent variables. The most important feature of the LR method is that it provides a direct calculation of the estimated probability values of the dependent variables, regardless of the assumptions made in factor analysis or discriminant analysis. LR formula is given as follows:

$$Pr = \frac{e^u}{1 + e^u}, \quad (6)$$

where Pr is the probability of a landslide occurrence and u is the independent variable.

$\beta_1, \beta_2, \beta_3$ are corresponding coefficients to each of the respective contributing factors.

$$u = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_3. \quad (7)$$

LR script and Weka software were used for constructing the susceptibility map by LR. Weka was used because of the need for external software to test the scripts of data preparation and create LSM and calculate ROC.

Data preparation script saves the weight values as CSV format. This file is easily used in Weka. After the required data is imported, tenfold cross-validation was applied for the logistic regression model in Weka. Coefficient values were calculated as 0.002, 0.0012, 0.0106, 0.0281, 0.0035, 0.0098,

0.0083, 0.0117, 0.0028, 0.0094 for Aspect, Curve, Elevation, Lithology, Ls-factor, NDVI, distance to ridge, Slope, TRI and TWI, respectively. Intercept value was calculated as -6.5421 for this method by Weka. After this process, the model was evaluated on all data set. Class distribution values were used for prediction. Distributions of class-1 (probability of landslide occurrence) values were exported to CSV format, and the coordinate of a pixel was added. This file was used as an input parameter of create LSM and calculate ROC script. Both scripts worked successfully. Success rate and predict rate were calculated as 75.40 and 72.57. These results obtained with the Weka software are the same as the results of the LR script in the LSAT tool.

Tuning LR method

Parameters of C , max_iter , and $solver$, are used as tuning parameters in tuning LR script. C is inverse of regularization strength.

$$C = 1/\lambda. \quad (8)$$

Lambda (λ) controls the trade-off between allowing the model to increase its complexity as much as it wants with trying to keep it simple. For example, if λ is very low or 0, the model will have enough power to increase its complexity (overfit) by assigning high values to the weights for each parameter. If, on the other hand, the value of λ is increased, the model will tend to underfit, as the model will become too simple.

Scikit-learn offers several techniques for training LR, called solvers. These are Newton-cg (Newton's method), lbfgs (Limited-memory Broyden-Fletcher-Goldfarb-Shanno Algorithm), liblinear (a Library for Large Linear Classification—Fan et al. 2008), SAG (Stochastic Average Gradient—Schmidt et al. 2017), SAGA (a variant of SAG—Defazio et al. 2014).

The last parameter is max_iter , the maximum number of iterations taken for the solvers to converge. These three parameters can be tuned individually, or the best parameters can be selected automatically. GridsearchCV method was used to select the best three parameters for the LR algorithm. GridsearchCV is the method that finds the best parameter values using the defined parameter values for an estimator. This method uses cross-validation. Fivefold cross-validation was used in the tuning script. LSM was created successfully with these parameter values of the GridsearchCV method. Tuning parameters and their values are shown in Table 3.

Random forest method

RF is one of the machine learning algorithms developed by Breiman (2001). It is an ensemble machine learning

Table 3 Tuning parameters of LR

Tuning parameters	Selected values
$C = [0.0001, 0.0005, 0.001, 0.005, 0.01, 0.05, 0.1, 0.5, 1, 10, 50, 200, 500]$	50
$max_iter = [10, 100, 200, 300, 500]$	100
$solver = [“newtoncg”, “lbfgs”, “liblinear”, “sag”, “saga”]$	lbfgs

algorithm used the method of Bootstrap Aggregation or bagging. RF creates multiple decision trees and combines them to achieve a more accurate and stable prediction. The samples in each tree are randomly selected. RF uses the majority voting method for the result. It can be used for both classification and regression problems. RF has the capability of robust and accurate performance on complex and big datasets with little need for fine-tuning. Another significant feature of RF is that it is more resistant to over-fitting. RF method has been frequently used recent LSM studies (Taalab et al. 2018; Kim et al. 2018; Dou et al. 2019; Park and Kim 2019).

Tuning RF algorithm

Machine learning algorithms need to be tune of parameters for more accurate predictions. The complexity and size of the trees should be controlled by setting hyperparameter values to reduce memory consumption.

The tuning script includes parameters of $n_estimators$, max_depth , $min_samples_splits$, and $min_samples_leaf$ for tuning. Also, the user can use the RandomizedSearchCV (RSCV) method to choose the best value of the parameters. It is recommended to use the RSCV method first.

RSCV implements a fit and a score method. It also implements predict, predict_proba, decision_function, transform, and inverse_transform if they are implemented in the estimator used. RSCV uses a cross-validation method for searching. The tuning parameters are shown in Table 4.

$N_estimators$ is the number of trees in the forest. Mostly, the higher the number of trees, the better the learning rate. Max_depth is the depth of each tree in the forest. The greater the depth, the more information is obtained

about data. $Min_samples_split$ is the minimum number of samples required to split. $Min_samples_leaf$ is the minimum number of samples required to be at a leaf node (Ben Fraj 2017).

Also, each parameter can be tuned individually. All tuning results could be saved as a png file for visualization.

RandomizedSearchCV results were used as parameter values, and LSM was created. But LSM was not reliable, and the difference between the success rate and prediction rate was too much. Therefore, parameters were tuned individually. Integer and float values were tested for $min_samples_splits$ and $min_samples_leaf$. It was observed that integer values were better than float. After tuning process, $n_estimators=2000$, $max_depth=4$, $min_samples_splits=2$, $criterion=entropy$ and $min_samples_leaf=100$ values are selected as the best values of parameters.

Multi-layer perceptron method

MLP is a supervised learning method used in many areas. It is also a preferred method in LSM studies due to its computational simplicity, finite parameterization, and stability. The simple perceptron model has only the input and output layer. The input layer sends incoming data to the hidden layer(s). The incoming information is transferred to the next layer. The number of hidden layers varies according to the problem, at least one, and adjusted according to the need. The number of neurons in the layer is also determined by the problem. The output layer determines the output of the network by processing the data from the previous layers. Any mathematical function can be used as an activation function in the model. The learning method of the system consists of two stages in general. The first part is a forward calculation (feedforward). The second part is a backward calculation (backpropagation).

MLP script uses an MLP classifier algorithm that trains using backpropagation. Landslide factors are the input feature, and landslide probability is the output layer. One hidden layer or more hidden layer can be used. Using multiple layers often does not increase accuracy (Rumelhart et al. 1986; Lippman 1987).

Table 4 Tuning parameters of RF

Tuning parameters	Selected values
$n_estimators = [1, 2, 4, 8, 16, 32, 64, 100, 200]$	8
$max_depths =$ a list from 1 to 32	18
$min_samples_splits = [0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 50]$	2
$min_samples_leaf = [0.1, 0.2, 0.3, 0.4, 0.5, 10, 50, 100, 250, 500, 1000]$	50

Tuning MLP algorithm

MLP classifier has many parameters (see scikit-learn.org). In this study, parameters of `hidden_layer_size`, `activation`, `solver`, `alpha`, `learning_rate`, `learning_rate_init`, `max_iter`, and `momentum` were used as tuning parameters. The methods of RSCV was used for the best parameter selection. After assessment of tuning results with Randomized-SearchCV method, the values of parameters were selected as shown in Table 5.

Tol parameter is used to stop training when the loss or score is not improving. The default value of tol is $1e - 4$. The values of $1e - 5$ and $1e - 6$ were tested with 20,000 iterations, and MLP results show overfitting with $1e - 6$ value. MLP script was executed with the above values of parameters, and LSM was created successfully.

Results and discussion

There are many studies in literature related to the assessment of landslide susceptibility, as described in the introduction. These studies will continue in the future to create a more accurate result. One of the most critical processes in these studies is the preparation of data. It takes a long time and requires attention. Pixel values of the raw data need to be extracted in this stage. This process creates a large data set.

Moreover, if the number of parameters increases, the data size increases too. Evaluation of prepared data could be done with different algorithms in the analysis phase. Machine learning methods are mostly used to predict landslide areas. The results of the analysis need to be visualized in the GIS environment. The other process is the validation of LSM. This step is necessary to test the accuracy of the analyses.

In this study, a novel tool was developed to produce LSM. LSAT automatically prepares data, analyses, calculates the success and prediction rate, and creates result

maps. LSM of the study area was prepared by using LSAT with five different models. All LSM were reclassified into five classes, such as very low, low, moderate, high, and very high using the natural break method (Fig. 6). ROC analysis was used to assess the accuracy of the models. ROC Curves plot the true positive rate (sensitivity) against the false positive rate (1-specificity). AUC values are generally evaluated between 0.5 and 1. Yeşilnacar and Topal (2005) mention the relationship between the prediction accuracy and the AUC value that can be classified as follows: 0.5–0.6 (poor), 0.6–0.7 (average), 0.7–0.8 (good), 0.8–0.9 (very good), and 0.9–1 (excellent).

ROC calculation script accepts susceptibility map data (in raster format), train data (polygon type), and test data (polygon type) as the input parameters. The outputs are the ROC curve and AUC values of test and train data. Landslide validation data was used to calculate the prediction rate. Landslide training data was used to calculate the success rate of LSM. For this method, the LSM was reclassified into 100 equal zones in GIS environment. Train and validation landslide pixel counts were defined for every class. Cumulative pixel values of zones were used to calculate the true positive rate (TPR) and false positive rate (FPR). The TPR formula is as follows:

$$TPR = \frac{TP}{TP + FN}, \tag{9}$$

where TPR is true positive rate (sensitivity), TP is true-positive, and FN is false-negative. The false positive rate (FPR) is equal to 1-specificity. FPR formula is as follows:

$$FPR = \frac{FP}{FP + TN}, \tag{10}$$

where FPR is false positive rate, FP is false-positive, and TN is true-negative. Then the trapezoidal formula was used to calculate the AUC (Bernard and Liengme 2002). This formula is as follows:

Table 5 Tuning parameters of MLP

Tuning parameters	Selected values
<i>Hidden_layer_sizes</i> : One or more hidden layers can be used. If one layer is used, this value must be neuron number as an integer type. Otherwise, it must be a list, including neuron numbers	100
<i>Solver</i> : lbfgs, sgd, and adam can be used as a solver function	sgd
<i>Activation</i> : identity, logistic, tanh, relu, can be used as activation function	tanh
<i>Alpha</i> : regularization parameter	0.00001
<i>Learning_rate</i> : it can be used as constant, invscaling, and adaptive	Constant
<i>Learning_rate_init</i> : It is used to update the weights	0.0001
<i>max_iter</i> : Maximum number of iterations for convergence	2000
<i>Momentum</i> : It is used for gradient descent update	0.7

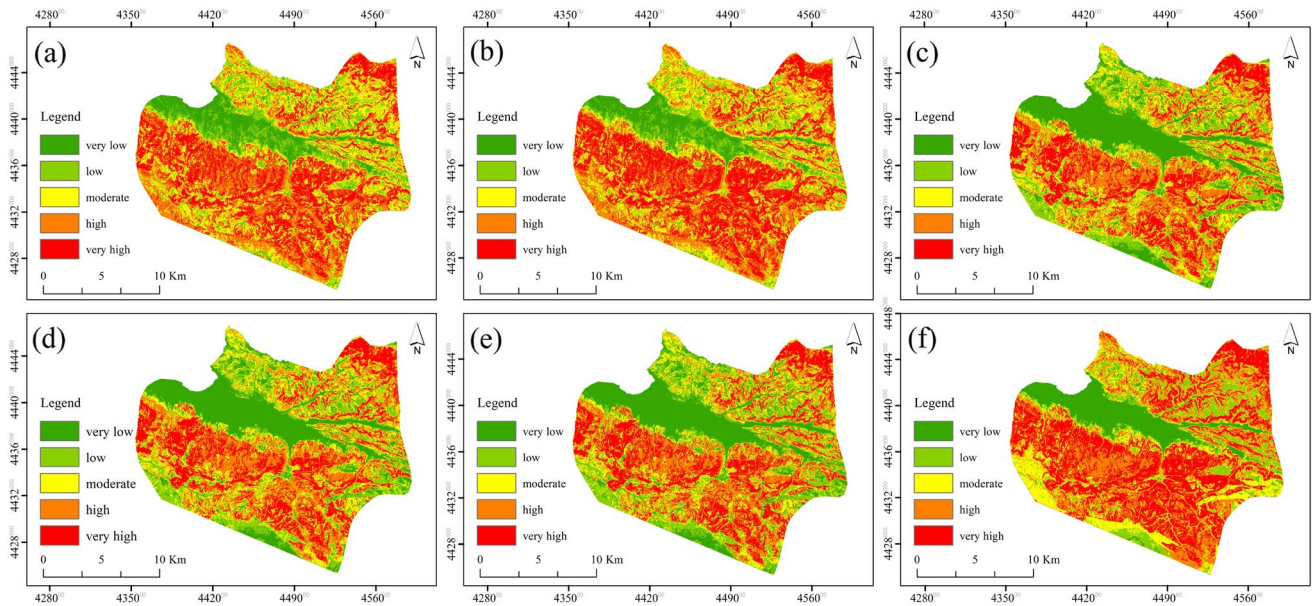


Fig. 6 LSM of study area constructed with the methods of **a** FR, **b** IV, **c** LR, **d** LR (Weka), **e** MLP and **f** RF

$$\begin{aligned}
 AUC = & \left(\left(\frac{y_1 - y_2}{2} \right) \times (x_1 - x_2) \right) \\
 & + \left(\left(\frac{y_2 - y_3}{2} \right) \times (x_2 - x_3) \right) \\
 & + \dots \left(\left(\frac{y_{n-1} - y_n}{2} \right) \times (x_{n-1} - x_n) \right),
 \end{aligned}
 \tag{11}$$

where y_0 is sensitivity values (TPR), and x is 1-specificity (FPR) values.

The prediction and the success rates of the models are shown in Fig. 7. AUC values were calculated as 70.95 %, 71.85 %, 72.57 %, 72.67 %, 73.93% for prediction rate of FR, IV, LR, MLP and RF methods, respectively. AUC values also were calculated as 72.71 %, 72.94 %, 75.40 %, 75.39 %, 77.39 % for success rate of FR, IV, LR, MLP and RF methods, respectively. The success and prediction capability of the RF algorithm are better than the other algorithms. The model derived using FR has the lowest success and prediction capability (Fig. 7).

Elevation, aspect, curvature, slope, geology, distance to ridge, NDVI, TRI, TWI, ls-factor were used as landslide controlling factors. Fault data were not used in this study. When the fault distances data were used, it was observed that the value of the success rate increased, and the value of the prediction rate decreased. Different training and test data sets experimented; different class ranges were tested, but the problem could not be solved. This issue could be about the accumulation of faults in a flat area in the middle of the basin. The parameter of distance to road was not used because it did not show a proper distribution for the

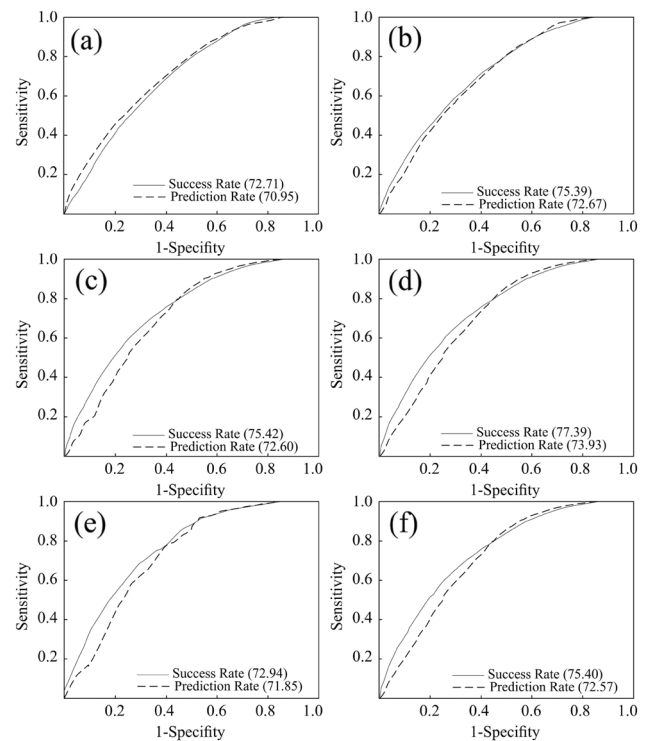


Fig. 7 ROC curves of models **a** FR, **b** MLP, **c** LR (Weka), **d** RF, **e** IV, **f** LR

whole area. LS factor parameter was preferred instead of the proximity to streams.

The main purpose of this study is to make the landslide susceptibility assessment quickly and accurately with a GIS

Table 6 Elapsed time for scripts (m:minutes, s:seconds)

Scripts	Elapsed time
Data preparation	14 m, 15 s
FR	26.56 s
IV	15.67 s
LR	58.76 s
RF	53.47 s
MLP	10 m, 24 s
Create LSM and calculate ROC (Weka)	2 m, 7 s

tool. The other tools designed for this purpose use statistical methods such as FR, WofE. Most of them are prepared with a model builder. Our tool uses both statistical and machine learning methods. In this study, only five methods were used. Different methods can also be added to this tool if desired.

Interfaces of all scripts, analysis processes, and processing times are given at <https://github.com/apolat2018/LSAT>. Also, the elapsed time for scripts is given in Table 6. The elapsed time varies according to the size of the study area, the number of used parameters, and the computer speed. In this study, ten raster data (controlling factors) covering an area of 319.886 km² were used. One raster data corresponds to 511,828 pixels. As seen in Table 6, the most time-consuming processes are MLP script, data preparation script, and create LSM and calculate ROC script, respectively. Other scripts give results in seconds.

Conclusions

The main objective of this study is to prepare a tool for creating LSM automatically. In literature, the LSM tools were created by bivariate statistical analysis methods. The tool developed in this study also includes machine learning methods. LSAT includes five methods (FR, IV, LR, RF, MLP). Also, data preparation, create LSM and calculate ROC, and tuning (LR, RF, and MLP) scripts were created by Python. Data preparation script makes the data ready for analysis using the weights of FR and IV. Create LSM and calculate ROC script visualizes the LSM and calculates the AUC value if external software is used for analysis. Weka software was used to create a data set with the LR method for this purpose. Tuning scripts are used to select the best values of parameters with the methods of GridSearchCv and RandomizedSearchCV. Tuning scripts were supported with graphics.

In conclusion, these scripts have greatly accelerated and simplified the data preparation and analysis processes. The best values of model parameters can be selected easily to tuning scripts with machine learning methods. LSAT is useful for automatically creating LSM with five different

methods, tuning of the LR, RF, MLP models, and comparing models for any area. LSAT was created with the Python language. It has very comprehensive machine learning libraries. LSAT was developed by selecting methods that are frequently used in LSM studies. In future work, different methods can be used, such as deep learning algorithms, and LSAT can be developed. Moreover, LSM can be created with various python libraries without using a GIS environment.

Acknowledgements The author wants to thank the Provincial Directorate of Disaster and Emergency Management for all supports. The author is also grateful to anonymous reviewers who reviewed this paper and, in that line, improved the manuscript significantly.

Code availability LSAT includes ten python script. These are: 1 Preparing data. 2 Create LSM and calculate ROC. 3 Frequency ratio. 4 Information value. 5 Logistic regression. 6 Tune lr. 7 Random forest. 8 Tune rf. 9 mlp. 10 tune mlp. A tool file also was created for ArcGIS software (Landslide Susceptibility Assessment Tool). The tool file and all Python scripts are available at: <https://github.com/apolat2018/LSAT>. Scripts by Dr. Ali POLAT (2018). ali.polat@afad.gov.tr. Provincial Directorate of Disaster and Emergency Management of Turkey, 58040 Sivas-Turkey.

Declarations

Conflict of interest The authors declare that they do not have any commercial or associative interests that represent conflicts of interest in connection with the paper they submitted.

References

- Aghdam IN, Varzandeh MHM, Pradhan B (2016) Landslide susceptibility mapping using an ensemble statistical index (Wi) and adaptive neuro-fuzzy inference system (ANFIS) model at Alborz Mountains (Iran). *Environ Earth Sci* 75:553. <https://doi.org/10.1007/s12665-015-5233-6>
- Akbaş B, Akdeniz N, Aksay A, Altun İ, Balcı V, Bilginer E, Bilgiç T, Duru M, Ercan T, Gedik İ, Günay Y, Güven İH, Hakyemez HY, Konak N, Papak İ, Pehlivan Ş, Sevin M, Şenel M, Tarhan N, Turhan N, Türkecan A, Ulu Ü, Uğuz MF, Yurtsever A (1991) Turkey geology map general directorate of mineral research and exploration publications. Ankara Turkey (**In Turkish**)
- Akgün A, Sezer EA, Nefeslioğlu HA, Pradhan B (2012) An easy to use MATLAB program (MamLand) for the assessment of landslide susceptibility using Mamdani fuzzy algorithm. *Comput Geosci* 38(1):23–34. <https://doi.org/10.1016/j.cageo.2011.04.012>
- Arabameri A, Pradhan B, Rezaei K, Lee CW (2019) Assessment of landslide susceptibility using statistical- and artificial intelligence-based FR-RF integrated model and multiresolution DEMs. *Remote Sens* 11:999. <https://doi.org/10.3390/rs11090999>
- Arca D, Kutoğlu HŞ, Becek K (2018) Landslide susceptibility mapping in an area of underground mining using the multicriteria decision analysis method. *Environ Monit Assess* 190:725. <https://doi.org/10.1007/s10661-018-7085-5>
- Baeza C, Corominas J (2001) Assessment of shallow landslides susceptibility by means of multivariate statistical techniques. *Earth Surf Process Landf* 26:1251–1263. <https://doi.org/10.1002/esp.263>

- Ben Fraj M (2017) In depth: parameter tuning for random forest. <https://medium.com/all-things-ai/in-depth-parameter-tuning-for-random-forest-d67bb7e920d>. Accessed June 2018
- Benchelha S, Aoudjehane HC, Hakdaoui M, Hamdouni RE, Mansouri H, Benchelha T, Layelmam M, Alaoui M (2019) Landslide susceptibility mapping in the Municipality of Oudka, Northern Morocco: a comparison between logistic regression and artificial neural networks models. *Int Arch Photogramm Remote Sens Spat Inf Sci* 4:41–49. <https://doi.org/10.5194/isprs-archives-XLII-4-W12-41-2019>
- Bernard V, Liengme (2002) A guide to Microsoft excel 2002 for scientists and engineers, 3rd edn. Butterworth-Heinemann Publishers, Oxford
- Boehner J, Selige T (2006) Spatial prediction of soil attributes using terrain analysis and climate regionalisation. In: Boehner J, McCloy KR, Strobl J (eds) SAGA—Analysis and modelling applications, vol 115. Geographische Abhandlungen, Goettinger, pp 13–27
- Bonham-Carter GF (1994) Geographic information systems for geoscientists. *Model GIS Pergamon* 13:398
- Bragagnolo L, da Silva RV, Grzybowski JMV (2019) Landslide susceptibility mapping with r Landslide: a free open-source GIS-integrated tool based on artificial neural networks. *Environ Model Softw* 123:104565. <https://doi.org/10.1016/j.envsoft.2019.104565>
- Breiman L (2001) Random forests. *Mach Learn* 45:5–32. <https://doi.org/10.1023/A:1010933404324>
- Chen W, Xie X, Wang J, Pradhan B, Hong H, Bui DT, Duan Z, Ma J (2017) A comparative study of logistic model tree, random forest, and classification and regression tree models for spatial prediction of landslide susceptibility. *Catena* 151:147–160. <https://doi.org/10.1016/j.catena.2016.11.032>
- Chen W, Shahabi H, Zhang S, Khosravi K, Shirzadi A, Chapi K, Pham BT, Zhang T, Zhang L, Chai H et al (2018) Landslide susceptibility modeling based on GIS and novel bagging-based kernel logistic regression. *Appl Sci* 8:2540. <https://doi.org/10.3390/app8122540>
- Chen W, Zhao X, Shahabi H, Shirzadi A, Khosravi K, Chai H, Zhang S, Zhang L, Ma J, Chen Y et al (2019) Spatial prediction of landslide susceptibility by combining evidential belief function, logistic regression and logistic model tree. *Geocarto Int* 34(11):1177–1201. <https://doi.org/10.1080/10106049.2019.1588393>
- Conrad O, Bechtel B, Bock M, Dietrich H, Fischer E, Gerlitz L, Wehberg J, Wichmann V, Böhner J (2015) System for automated geoscientific analyses (SAGA) v. 2.1.4. *Geosci Model Dev* 8:1991–2007. <https://doi.org/10.5194/gmd-8-1991-2015>
- Cui K, Lu D, Li W (2016) Comparison of landslide susceptibility mapping based on statistical index, certainty factors, weights of evidence and evidential belief function models. *Geocarto Int* 32(9):935–955. <https://doi.org/10.1080/10106049.2016.1195886>
- Dagdelenler G, Nefeslioglu HA, Gökçeoğlu C (2016) Modification of seed cell sampling strategy for landslide susceptibility mapping: an application from the Eastern part of the Gallipoli Peninsula (Canakkale, Turkey). *Bull Eng Geol Environ* 75:575–590. <https://doi.org/10.1007/s10064-015-0759-0>
- Das HO, Sonmez H, Gökçeoğlu C, Nefeslioglu HA (2013) Influence of seismic acceleration on landslide susceptibility maps: a case study from NE Turkey (the Kelkit Valley). *Landslides* 10:433–454. <https://doi.org/10.1007/s10346-012-0342-8>
- Defazio A, Bach F, Lacoste-Julien S (2014) SAGA: a fast incremental gradient method with support for non-strongly convex composite objectives. [arXiv:1407.0202v3](https://arxiv.org/abs/1407.0202v3)
- Devkota KC, Regmi AD, Pourghasemi HR, Yoshida K, Pradhan B, Ryu IC, Dhital MR, Althuwaynee OF (2013) Landslide susceptibility mapping using certainty factor, index of entropy and logistic regression models in GIS and their comparison at Mugling–Narayanghat road section in Nepal Himalaya. *Nat Hazards* 65:135–165. <https://doi.org/10.1007/s11069-012-0347-6>
- Dikshit A, Pradhan B, Alamri AM (2020) Pathways and challenges of the application of artificial intelligence to geohazards modelling. *Gondwana Research*. In Press, Corrected Proof, Available online 17 September 2020. <https://doi.org/10.1016/j.gr.2020.08.007>
- Dobesova Z (2011) Programming language Python for data processing. In: The international conference on electrical and control engineering (ICECE), IEEE 2011
- Dou J, Yamagishi H, Pourghasemi HR, Yunus AP, Song X, Xu Y, Zhu Z (2015) An integrated artificial neural network model for the landslide susceptibility assessment of Osado Island, Japan. *Nat Hazards* 78:1749–1776. <https://doi.org/10.1007/s11069-015-1799-2>
- Dou J, Yunus AP, Bui DT, Merghadi A, Sahana M, Zhu Z, Chen CW, Khosravi K, Yang Y, Pham PT (2019) Assessment of advanced random forest and decision tree algorithms for modeling rain-fall-induced landslide susceptibility in the Izu-Oshima Volcanic Island, Japan. *Sci Total Environ* 662:332–346. <https://doi.org/10.1016/j.scitotenv.2019.01.221>
- Eibe F, Hall MA, Witten IH (2016) The WEKA workbench. Online Appendix for “Data mining: practical machine learning tools and techniques”, Morgan Kaufmann, Fourth Edition
- Eker R, Aydın A (2014) Ormanların heyelan oluşumu üzerindeki etkileri (The effects of forests on landslides). *Turkish Journal of Forestry*. 15:84–93. <https://doi.org/10.18182/tjf.31067>. (In Turkish)
- Ermimi L, Catani F, Casagli N (2005) Artificial neural networks applied to landslide susceptibility assessment. *Geomorphology* 66(1–4):327–343. <https://doi.org/10.1016/j.geomorph.2004.09.025>
- ESRI (2011) ArcGIS Desktop: Release 10. Environmental Systems Research Institute, Redlands
- Fan RE, Chang KW, Hsieh CJ, Wang XR, Lin CJ (2008) LIBLINEAR: a library for large linear classification. *J Mach Learn Res* 9(2008):1871–1874
- Gökçe O, Özden Ş, Demir A (2008) Türkiye’de afetlerin mekansal ve istatistiksel dağılımı afet bilgileri envanteri. Afet İşleri Genel Müdürlüğü Afet Etüt ve Hasar Tespit Daire Başkanlığı, Ankara (In Turkish)
- Gökçeoğlu C, Sonmez H, Nefeslioglu HA, Duman TY, Can T (2005) The 17 March 2005 Kuzulu landslide (Sivas, Turkey) and landslide-susceptibility map of its near vicinity. *Eng Geol* 81(1):65–83. <https://doi.org/10.1016/j.enggeo.2005.07.011>
- Gruber S, Peckham S (2009) Land surface parameters and objects in hydrology. In: Hengl T, Reuter HI (eds) *Geomorphometry: concepts, software, applications*. Elsevier, Amsterdam, pp 171–194. [https://doi.org/10.1016/S0166-2481\(08\)00007-X](https://doi.org/10.1016/S0166-2481(08)00007-X)
- Guzzetti F, Reichenbach P, Ardizzone F, Cardinali M, Galli M (2005) Probabilistic landslide hazard assessment at the basin scale. *Geomorphology* 72:272–299. <https://doi.org/10.1016/j.geomorph.2005.06.002>
- Guzzetti F, Mondini AC, Cardinali M, Fiorucci F, Santangelo M, Chang KT (2012) Landslide inventory maps: new tools for an old problem. *Earth Sci Rev* 112(1–2):42–66. <https://doi.org/10.1016/j.earscirev.2012.02.001>
- Hasekioğulları GD, Ercanoğlu M (2012) A new approach to use AHP in landslide susceptibility mapping: a case study at Yenice (Karabuk, NW Turkey). *Nat Hazards* 63:1157–1179. <https://doi.org/10.1007/s11069-012-0218-1>
- Hong HY, Pradhan B, Xu C, Tien Bui D (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. <https://doi.org/10.1016/j.catena.2015.05.019>

- Hong H, Naghibi SA, Pourghasemi HR, Pradhan B (2016) GIS-based landslide spatial modeling in Ganzhou City, China. *Arab J Geosci* 9(2):112. <https://doi.org/10.1007/s12517-015-2094-y>
- Jebur MN, Pradhan B, Shafri HZM, Yusoff ZM, Tehrani MS (2015) An integrated user-friendly ArcMAP tool for bivariate statistical modelling in geoscience applications. *Geosci Model Dev* 8:881–891. <https://doi.org/10.5194/gmd-8-881-2015>
- Jimenez-Peralvarez JD, Irigaray C, El Hamdouni R, Chacon J (2009) Building models for automatic landslide-susceptibility analysis, mapping and validation in ArcGIS. *Nat Hazards* 50(3):571–590. <https://doi.org/10.1007/s11069-008-9305-8>
- Jones D, Brunnsden D, Goudie A (1983) A preliminary geomorphological assessment of part of the Karakoram highway. *Q J Eng Geol Hydrogeol* 16:331–355. <https://doi.org/10.1144/GSL.QJEG.1983.016.04.10>
- Kim J, Lee S, Jung H, Lee S (2018) Landslide susceptibility mapping using random forest and boosted tree models in Pyeong-Chang, Korea. *Geocarto Int* 33(9):1000–1015. <https://doi.org/10.1080/10106049.2017.1323964>
- Köse DD, Turk T (2019) GIS-based fully automatic landslide susceptibility analysis by weight-of-evidence and frequency ratio methods. *Phys Geogr* 40(5):481–501. <https://doi.org/10.1080/02723646.2018.1559583>
- Lee S, Talib JA (2005) Probabilistic landslide susceptibility and factor effect analysis. *Environ Geol* 47:982–990. <https://doi.org/10.1007/s00254-005-1228-z>
- Lee S, Ryu J, Kim S (2007) Landslide susceptibility analysis and its verification using likelihood ratio, logistic regression, and artificial neural network models: case study of Youngin Korea. *Landslides* 4(4):327–338. <https://doi.org/10.1007/s10346-007-0088-x>
- Lippmann RP (1987) An introduction to computing with neural nets. *IEEE ASSP Mag* 4(2):4–22. <https://doi.org/10.1109/MASSP.1987.1165576>
- Luo Y, Su B, Yuan J, Li H, Zhang Q (2011) GIS Techniques for Watershed Delineation of SWAT Model in Plain Polders. 2011 3rd International Conference on Environmental Science and Information Application Technology (ESIAT 2011):2050–2057. <https://doi.org/10.1016/j.proenv.2011.09.321>
- Maggioni M, Gruber U (2003) The influence of topographic parameters on avalanche release dimension and frequency. *Cold Reg Sci Technol* 37:407–419. [https://doi.org/10.1016/S0165-232X\(03\)00080-6](https://doi.org/10.1016/S0165-232X(03)00080-6)
- Moore ID, Grayson RB, Ladson AR (1991) Digital terrain modeling—a review of hydrological geomorphological and biological applications. *Hydrol Process* 5:3–30. <https://doi.org/10.1002/hyp.3360050103>
- Narayanan S, Sivakumar R (2018) Development of ArcPy based customized tool in GIS for seismic information system. *Int J Pure Appl Math* 118(22):377–382
- Nefeslioglu HA, Duman TY, Durmaz S (2008) Landslide susceptibility mapping for a part of tectonic Kelkit Valley (Eastern Black Sea region of Turkey). *Geomorphology* 94(3–4):401–418. <https://doi.org/10.1016/j.geomorph.2006.10.036>
- Nefeslioglu HA, Sezer E, Gökçeoğlu C, Bozkir AS, Duman TY (2010) Assessment of landslide susceptibility by decision trees in the metropolitan area of Istanbul, Turkey. *Math Probl Eng* 2010:15. <https://doi.org/10.1155/2010/901095>
- Nohani E, Moharrami M, Sharafi S, Khosravi K, Pradhan B, Pham BT, Lee S, Melesse AM (2019) Landslide susceptibility mapping using different GIS-based bivariate models. *Water* 11(7):1402. <https://doi.org/10.3390/w11071402>
- Nourani V, Pradhan B, Ghaffari H, Sharifi SS (2014) Landslide susceptibility mapping at Zonouz Plain, Iran using genetic programming and comparison with frequency ratio, logistic regression, and artificial neural network models. *Nat Hazards* 71:523–547. <https://doi.org/10.1007/s11069-013-0932-3>
- Osna T, Sezer EA, Akgun A (2014) GeoFIS: an integrated tool for the assessment of landslide susceptibility. *Comput Geosci* 66:20–30. <https://doi.org/10.1016/j.cageo.2013.12.016>
- Ozer BC, Mutlu B, Nefeslioglu HA, Sezer EA, Rouai M, Dekayir A, Gokceoglu C (2020) On the use of hierarchical fuzzy inference systems (HFIS) in expert-based landslide susceptibility mapping: the central part of the Rif Mountains (Morocco). *Bull Eng Geol Environ* 79:551–568. <https://doi.org/10.1007/s10064-019-01548-5>
- Palamakumbure D, Stirling D, Flentje P, Chowdhury R (2015) ArcGIS v.10 landslide susceptibility data mining add-in tool integrating data mining and GIS techniques to model landslide susceptibility. In: Lollino G, Giordian D, Crosta G, Corominas J, Azzam R, Wasowski J, Sciarra N (eds) *Engineering geology for society and territory*, vol 2. Springer, New York, pp 1191–1194. https://doi.org/10.1007/978-3-319-09057-3_208
- Park S, Kim J (2019) Landslide susceptibility mapping based on random forest and boosted regression tree models, and a comparison of their performance. *Appl Sci* 9:942. <https://doi.org/10.3390/app9050942>
- Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J (2011) Scikit-learn: machine learning in Python. *J Mach Learn Res* 12:2825–2830
- Pham BT, Bui DT, Prakash I, Dholakia MB (2016) Rotation forest fuzzy rule-based classifier ensemble for spatial prediction of landslides using GIS. *Nat Hazards* 83:97–127. <https://doi.org/10.1007/s11069-016-2304-2>
- Polat A, Erik D (2020) Debris flow susceptibility and propagation assessment in West Koyulhisar, Turkey. *J Mt Sci* 17:2611–2623. <https://doi.org/10.1007/s11629-020-6261-6>
- Polat A, Gürsoy H (2014) Sayısal yükselti modeli (SYM) verileri yardımıyla 17 Mart 2005 Kuzul (Koyulhisar, Sivas) heyelanının hacim hesabı. ATAG 18. Muğla, Sıtkı Koçman Üniversitesi. (**IN Turkish**)
- Polat A, Tatar O, Gürsoy H, Yalçiner CÇ, Büyüksaraç A (2014) Two-phased evolution of the Suşehri Basin on the North Anatolian Fault Zone Turkey. *Geodinamica Acta* 25(3–4):132–145. <https://doi.org/10.1080/09853111.2013.861997>
- Pourghasemi HR, Mohammady M, Pradhan B (2012a) Landslide susceptibility mapping using index of entropy and conditional probability models in GIS: Safarood Basin, Iran. *Catena* 97:71–84. <https://doi.org/10.1016/j.catena.2012.05.005>
- Pourghasemi HR, Pradhan B, Gökçeoğlu C, Moezzi KD (2012b) Landslide susceptibility mapping using a spatial multi criteria evaluation model at Haraz watershed, Iran. In: Pradhan B, Buchroithner M (eds) *Terrigenous mass movements*. Springer, Berlin, pp 23–49. https://doi.org/10.1007/978-3-642-25495-6_2
- 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. <https://doi.org/10.1016/j.cageo.2012.08.023>
- Regmi NR, Giardino JR, McDonald EV, Vitek JD (2014) A comparison of logistic regression-based models of susceptibility to landslides in western Colorado, USA. *Landslides* 11(2):247–262. <https://doi.org/10.1007/s10346-012-0380-2>
- Riley SJ, DeGloria SD, Elliot R (1999) Index that quantifies topographic heterogeneity. *Intermountain J Sci* 5(1–4):23–27
- Rozycka M, Migoń P, Michniewicz A (2016) Topographic Wetness Index and Terrain Ruggedness Index in geomorphic characterisation of landslide terrains, on examples from the Sudetes, SW Poland. *Zeitschrift für Geomorphol Suppl Issues* 61(2):61–80. https://doi.org/10.1127/zfg_suppl/2016/0328
- Rumelhart DE, Hinton GE, Williams RJ (1986) Learning internal representations by error propagation. In: Rumelhart DE, McClelland JL (eds) *Parallel distributed processing*, vol 1. Springer, New York, pp 318–362

- Sahin EK, Colkesen I, Kavzoglu T (2020) A comparative assessment of canonical correlation forest, random forest, rotation forest and logistic regression methods for landslide susceptibility mapping. *Geocarto Int* 35(4):341–363. <https://doi.org/10.1080/10106049.2018.1516248>
- Schmidt M, Roux NL, Bach F (2017) Minimizing finite sums with the stochastic average gradient. *Math Program* 162:83–112. <https://doi.org/10.1007/s10107-016-1030-6>
- Sevgen E, Kocaman S, Nefeslioglu HA, Gökçeoğlu C (2019) A novel performance assessment approach using photogrammetric techniques for landslide susceptibility mapping with logistic regression, ANN and random forest. *Sensors* 19(18):3940. <https://doi.org/10.3390/s19183940>
- Sezer EA, Nefeslioglu HA, Osna T (2016) An expert-based landslide susceptibility mapping (LSM) module developed for Netcad Architect Software. *Comput Geosci* 98:26–37. <https://doi.org/10.1016/j.cageo.2016.10.001>
- Song Y, Gong J, Gao S, Wang D, Cui T, Li Y, Wei B (2012) Susceptibility assessment of earthquake-induced landslides using Bayesian network: a case study in Beichuan China. *Comput Geosci* 42:189–199. <https://doi.org/10.1016/j.cageo.2011.09.011>
- Sorensen R, Zinko U, Seibert J (2006) On the calculation of the topographic wetness index: evaluation of different methods based on field observations. *Hydrol Earth Syst Sci* 10:101–112. <https://doi.org/10.5194/hess-10-101-2006>
- Taalab K, Cheng T, Zhang Y (2018) Mapping landslide susceptibility and types using random forest. *Big Earth Data* 2(2):159–178. <https://doi.org/10.1080/20964471.2018.1472392>
- Tatar O, Gursoy H, Kocbulut F, Mesci BL (2005) Active fault zones and landslides: the 17 March 2005 Kuzulu (Koyulhisar) landslide. *Cumhuriyet Bilim Teknik Dergisi* 941:5–6
- Tien Bui D, Pradhan B, Lofman O, Revhaug I (2012) Landslide susceptibility assessment in Vietnam using support vector machines, decision tree, and Naive Bayes models. *Math Probl Eng* 2012:26. <https://doi.org/10.1155/2012/974638>
- Tien Bui D, Tuan TA, Klempe H, Pradhan B, Revhaug I (2015) Spatial prediction models for shallow landslide hazards: a comparative assessment of the efficacy of support vector machines, artificial neural networks, kernel logistic regression, and logistic model tree. *Landslides* 13:361–378. <https://doi.org/10.1007/s10346-015-0557-6>
- Torizin J (2012) Landslide susceptibility assessment tools for ArcGIS 10 and their application. In: *Proceedings of 34th IGC, Brisbane, Australia*, pp 5–10
- Ulusay R, Aydan Ö, Kılıç R (2007) Geotechnical assessment of the 2005 Kuzulu landslide (Turkey). *Eng Geol* 89(1–2):112–128. <https://doi.org/10.1016/j.enggeo.2006.09.020>
- Umar Z, Pradhan B, Ahmad A, Jebur MN, Tehrany MS (2014) Earthquake induced landslide susceptibility mapping using an integrated ensemble frequency ratio and logistic regression models in West Sumatera province, Indonesia. *Catena* 118:124–135. <https://doi.org/10.1016/j.catena.2014.02.005>
- Van Westen CJ, Rengers N, Terlien MTJ, Soeters R (1997) Prediction of the occurrence of slope instability phenomena through GIS-based hazard zonation. *Geol Rundsch* 86:404–414. <https://doi.org/10.1007/s005310050149>
- Xu C, Xu XW, Dai FC, 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. <https://doi.org/10.1016/j.cageo.2012.01.002>
- Yeşilnacar E, Topal T (2005) Landslide susceptibility mapping: a comparison of logistic regression and neural networks methods in a medium scale study Hendek region (Turkey). *Eng Geol* 79:251–266. <https://doi.org/10.1016/j.enggeo.2005.02.002>
- Yin KL, Yan TZ (1988) Statistical prediction model for slope instability of metamorphosed rocks. In: *Proceedings of 5th international symposium on landslides, Lausanne, Switzerland vol 2*, pp 1269–1272
- Zare M, Pourghasemi HR, Vafakhah M, Pradhan B (2013) Landslide susceptibility mapping at Vaz Watershed (Iran) using an artificial neural network model: a comparison between multilayer perceptron (MLP) and radial basic function (RBF) algorithms. *Arab J Geosci* 6:2873–2888. <https://doi.org/10.1007/s12517-012-0610-x>
- Zhang X, Song W, Liu L (2014) An implementation approach to store GIS spatial data on NoSQL database. In: *The 22nd International conference on geoinformatics, Kaohsiung*, pp 1–5. <https://doi.org/10.1109/GEOINFORMATICS.2014.6950846>
- Zhu Z (2010) An earthquake-induced landslide risk assessment model using the model builder of ArcGIS. In: *The 18th International conference on geoinformatics, Beijing*, pp 1–4. <https://doi.org/10.1109/GEOINFORMATICS.2010.5567701>

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