Catena 145 (2016) 164–179

Contents lists available at ScienceDirect

Catena

journal homepage: www.elsevier.com/locate/catena

Comparison of a logistic regression and Naïve Bayes classifier in landslide susceptibility assessments: The influence of models complexity and training dataset size

Paraskevas Tsangaratos [a,*], Ioanna Ilia [b]

[a] Mining and Metallurgical Engineering, National Technical University of Athens, School of Mining and Metallurgical Engineering, Department of Geological Studies, Zografou Campus: Heroon Polytechniou 9, 15780 Zografou, Greece
[b] National Technical University of Athens, School of Mining and Metallurgical Engineering, Department of Geological Studies, Zografou Campus: Heroon Polytechniou 9, 15780 Zografou, Greece

ARTICLE INFO

Article history:
Received 28 August 2015
Received in revised form 14 April 2016
Accepted 3 June 2016
Available online xxxx

Keywords:
Landslide susceptibility
Logistic regression classifier
Naïve Bayes classifier
Geographical information system
Epirus
Greece

ABSTRACT

The main objective of the present study was to compare the performance of a classifier that implements the Logistic Regression and a classifier that employs a Naïve Bayes algorithm in landslide susceptibility assessments. The study provides an evaluation concerning the influence of model's complexity and the size of the training data, while it identifies the most accurate and reliable classifier.

The comparison of the two classifiers was based on the assessment of a database containing 116 sites located at the mountains of Epirus, Greece, where serious landslides events have been encountered. The sites are classified into two categories, non-landslide and landslide areas. The identification of those areas was established by analysing airborne imagery, extensive field investigation and the examination of previous research studies. The geo-environmental conditions in those locations where analyzed in regard with their susceptibility to slide. In particular, seven variables where analyzed: engineering geological units, slope angle, slope aspect, mean annual rainfall, distance from river network, distance from tectonic features and distance from road network.

Multicollinearity analysis and feature selection was implemented in order to estimate the conditional independence among the variables and to rank the variables according to their significance in estimating landslide susceptibility. By the above processes the construction of nine different datasets was accomplished. Further partition allowed creating subsets of training and validating data from the original 116 sites. Each dataset was characterized by the number of the variables used and the size of the training datasets.

The comparison and validation of the outcomes of each model was achieved using statistical evaluation measures, the receiving operating characteristic and the area under the success and predictive rate curves. The results indicated that model's complexity and the size of the training dataset influence the accuracy and the predictive power of the models concerning landslide susceptibility. In particular, the most accurate model with high predictive power was the eighth model (five variables and 92 training data), with the Naïve Bayes classifier having a slightly higher overall performance and accuracy than the Logistic Regression classifier, 87.50% and 82.61% on the validation datasets, respectively. The highest area under the curve was achieved by the Naïve Bayes classifier for both the training and validating datasets (0.875 and 0.806 respectively) while the Logistic Regression classifier achieved a lower AUC values for the training and validating datasets (0.844 and 0.711, respectively). When limited data are available it seems that more accurate and reliable results could be obtained by generative classifiers, like Naïve Bayes classifiers. Overall, landslide susceptibility assessments could serve as a useful tool for the local and national authorities, in order to evaluate strategies to prevent and mitigate the adverse impacts of landslide events.

1. Introduction

Landslides are natural, geological phenomena that involve a wide range of soil, debris or rock mass movements that may occur in offshore, coastal and inland areas, driven by the force of gravity and the aid of water. These movements are identified as the result of the progressive

* Corresponding author.
E-mail addresses: ptsag@metal.ntua.gr (P. Tsangaratos), gilia@metal.ntua.gr (I. Ilia).

http://dx.doi.org/10.1016/j.catena.2016.06.004
0341-8162/

or extreme evolution of natural events that are developed due to the action of geological, tectonic, geomorphological and climatic processes. In addition to these processes, it has been widely reported that many cases of landslides are developed as a result of the negative impact of human activities on the environment (Alexander, 1992; Hutchinson, 1995; Baioni, 2011; Alimohammadlou et al., 2013).

According to Varnes (1984), the risk against landslide phenomena, can be thought as the probability of occurrence of a potentially injurious event in a specified time period and in a given area. This definition contains two elements; space and time. The spatial variable specifies those areas that are susceptible to the development of failure at a given time, while the time variable specifies the time the event will occur in a given area.

In this context, the variables that determine the probability of occurrence of a landslide event can be divided into two categories (Dai et al., 2002): intrinsic variables that contribute to landslide susceptibility, such as geological and morphological characteristics of the earth's surface, geotechnical parameters, effects of vegetation cover, the hydrographic network, and external variables that are responsible for triggering landslides such as rainfall and seismic activity. If the external variables are not taken into account, the term susceptibility refers to the probability of the presence of a landslide event considering only the spatial dimension of the problem.

Korup and Stolle (2014) reports that the prediction through the use of various methods and techniques of the spatial distribution of landslides is by far the most investigated topic that aids land – use planning, decision making and overall landslide risk reduction strategies.

According to van Westen et al. (2006) a region could be considered to be prone to landslide phenomena when the geo – environmental conditions of the region share common features with a region where a failure has been manifested in the past. Thus, the susceptibility of the area could be defined by a set of geological, tectonic and hydrologic conditions, morphological characteristics, soil and vegetation features, land use and human practices.

In general, the analysis of landslide phenomenon is attempted through qualitative, semi – quantitative, quantitative methods and various modeling techniques (Fell et al., 2008; Rozos et al., 2008). The majority of the applied methods are based either on the experience and knowledge provided by experts or on statistical or probabilistic theories or even the use of deterministic models (Soeters and van Westen, 1996; Aleotti and Chowdhury, 1999; Castellanos Abella and van Westen, 2007; Fell et al., 2008).

Relatively recently, new techniques and methods where utilized as promising tools to evaluate the susceptibility and risk against landslides that come from the domain of Machine Learning and Data Mining (Flentje et al., 2007; Miner et al., 2010; Marjanovic et al., 2011; Pradhan, 2013; Tsangaratos et al., 2013; Korup and Stolle, 2014; Goetz et al., 2015). These methods are characterized by the ability of learning and discovering hidden and unknown patterns from large multi-thematic databases (Tsangaratos and Ilia, 2015).

Numerous papers can be found through the scientific literature that take advantage of their ability to sufficiently assess data, including: the logistic regression approach (Lee, 2004; Ayalew and Yamagishi, 2005; Lee and Sambath, 2006; Van Den Eeckhaut et al., 2006; Lee and Pradhan, 2007; Nefeslioglu et al., 2008; Das et al., 2010; Oh and Lee, 2010; Suzen and Kaya, 2011; Yalcin et al., 2011; Felicisimo et al., 2013; Pourghasemi et al., 2013a; Regmi et al., 2014; Hong et al., 2015), fuzzy logic method (Ercanoglu and Gokceoglu, 2002, 2004; Champatiray et al., 2007; Muthu et al., 2008; Pradhan et al., 2009, 2010a; Pradhan, 2010, 2011a, b; Akgun et al., 2012; Pourghasemi et al., 2012a; Tien Bui et al., 2012b; Feizizadeh et al., 2013, 2014; Zhu et al., 2014), artificial neural network method (Lee et al., 2003, 2004; Neaupane and Achet, 2004; Ermini et al., 2005; Ferentinou and Sakellariou, 2007; Caniani et al., 2008; Melchiorre et al., 2008; Nefeslioglu et al., 2008; Choi et al., 2010; Pradhan and Lee, 2010a, 2010b, 2010c; Pradhan et al., 2010a; Poudyal et al., 2010; Yilmaz, 2010; Tien Bui et al., 2012a; Zare et al., 2013; Alimohammadlou et al., 2014; Conforti et al., 2014; Tsangaratos and Benardos, 2014), Bayes theorem based on weights of evidence (Regmi et al., 2010a, 2010b; Kayastha et al., 2012; Pourghasemi et al., 2012b; Kouli et al., 2014; Ilia and Tsangaratos, 2015), neural–fuzzy method (Vahidnia et al., 2010; Pradhan et al., 2010b; Oh and Lee, 2011; Oh and Pradhan, 2011; Sezer et al., 2011; Sdao et al., 2013; Pradhan, 2013), support vector machines (Yao et al., 2008; Yilmaz, 2010; Xu et al., 2012; Ballabio and Sterlacchini, 2012; Tien Bui et al., 2012c; Pourghasemi et al., 2013b; Pradhan, 2013; Hong et al., 2015) and decision tree method (Saito et al., 2009; Yeon et al., 2010; Nefeslioglu et al., 2010; Tien Bui et al., 2012c; Pradhan, 2013; Tsangaratos and Ilia, 2015).

The present paper focuses on the quantitative methods that utilize statistical or probabilistic models and also Machine Learning and Data Mining methods, to assess the role of landslide – causative variables. Particular, the study addresses two methods, Logistic Regression (LR) and Naïve Bayes (NB) to develop appropriate classifiers in order to classify the research area into landslide or non-landslide zones. LR is a widely used statistical direct probability model, while NB is considered as a simple probabilistic model that is based on the Bayes' theorem.

In more details, LR has been utilized in numerous landslide susceptibility assessments, providing accurate and reliable results in a rather simple manner. Based on its learning mechanism it is characterized as a discriminative model which estimates the probability for a given feature (x) and the label (y) directly from the training data by minimizing error (Ng and Jordan, 2001). On the other hand NB has been employed in rather fewer studies presenting respectively high accuracy. Based on its learning mechanism it is referred to as a generative model since for the given features (x) and the label (y) it estimates a joint probability from the training data (Ng and Jordan, 2001). The two techniques further differ in the adopted assumptions and also limitations of the models performance. The NB model assumes that all the features are conditionally independent, while LR splits feature space linearly, thus it works even if some of the variables are correlated (John and Langley, 1995; Montgomery et al., 2001). As for models limitations, the NB has been reported to work well even with less training data, as the estimates are based on the joint density function, while LR produces results that over fit the data, a condition when a model begins to memorize training data rather than learning to generalize from trend (Melchiorre et al., 2008; Tsangaratos and Benardos, 2014). This was our initial intension, to find for each model its limitations. Specifically, through the implementation of the developed classifiers, two objectives were achieved; the construction of a landslide susceptibility map for each approach and the comparison of their performance in regard with the complexity of the developed models and the size of the training data used.

Concerning the second objective of the study, several studies have compared LR and NB with other qualitative, semi – quantitative and quantitative methods to determine the optimum mathematical method to assess landslide susceptibility (Yesilnacar and Topal, 2005; Lee and Sambath, 2006; Miner et al., 2010; Pradhan and Lee, 2010a, 2010b; Yilmaz, 2010; Akgun, 2012; Ballabio and Sterlacchini, 2012; Tien Bui et al., 2012c; Bijukchhen et al., 2013; Felicisimo et al., 2013; Pourghasemi et al., 2013a; Shahabi et al., 2014). However, there are relative few studies that evaluate the performance of the applied classifiers in regard with the complexity of the models and the size of the training data sets (Brenning, 2005; Nefeslioglu et al., 2008; Pradhan and Lee, 2010b; Wang et al., 2013; Heckmann et al., 2014).

The study area covers the mountains of Central Tzoumerka, which are located at the administrative unit of Epirus Greece, where serious landslides events have been encountered. The computation process was carried out using Microsoft Visual Studio 2010 Professional (Halvorson, 2010) for implementing the NB algorithm, Weka 3.7.6 for feature selection process (Hall et al., 2009) and SPSS 16.0 (SPSS, 2007) for implementing multicollinearity analysis and LR, while ArcGIS 10.1

(ESRI, 2013) was used for compiling the data and producing the landslide susceptibility maps.

2. Material and methods

As already mentioned the main goal of the study was to compare the performance of LR and NB classifiers in regard with the complexity of the developed models and the size of the training data used. The complexity of a model is related with the number of variables used, while the size of the training data is crucial for the learning process. These two characteristics are somehow correlated and the selection of which provides optimum performance for the developed model.

According to Jain et al. (2000), as a rule of thumb, a minimum of $10 \cdot d \cdot C$ training samples is required for a d – dimensional classification problem of C classes. The higher the d – dimensions the higher the complexity of the model and larger volumes of training data are needed. The nature of the landslide phenomena often makes it difficult to obtain the required number of training data and thus the complexity of the model which depends on the number of variables used must be tuned to an optimal level by selecting the most appropriate variables (Chacon et al., 2006; Irigaray et al., 2007; Costanzo et al., 2012).

The developed methodology was separated into a five phase procedure; (a) constructing the inventory map, (b) the data pre-processing phase, (c) the phase of constructing the different models and training database, (d) the phase of implementing the two classifiers and constructing the landslide susceptibility map and (e) the validation and comparison of the two classifiers. Details of the developed methodology are included in the following paragraphs (Fig. 1).

2.1. Constructing the inventory database

The inventory database includes information about the location, features and abundance of landslide and non-landslides areas. The identification and acceptance of those areas was based on historical information concerning landslide incidence, the interpretation of aerial photos, the use of satellite imagery and extensive field observations. Landslides were mapped as polygons and transformed into a single point placed in the centroid of the area or multiple points when the surface of the area was greater than corresponding grid cell that was used in the analysis. In the case presented in this study the corresponding grid cell had a dimension of 20 × 20m yielding a landslide surface of 400 m^2. The size of the grid cell was considered being sufficient and representative of the pre-failure morphological settings of the landslide surface since the average size of the observed landslide was about 320 m^2.

2.2. Pre-processing data

The main purpose of the pre-processing phase was to prepare the data in order to be useful for further analysis. It involves four sub-processes; (a) the construction of a geodatabase environment, (b) the classification of the variables into classes of landslide susceptibility, (c) the multicollinearity analysis and (d) the feature selection process.

Concerning the first sub-process, besides the inventory database, the collection of all available data was necessary. Specifically, a GIS geodatabase was constructed including data concerning the morphological, geological, hydrogeological, tectonic, meteorological and anthropogenic settings of the wider area of research.

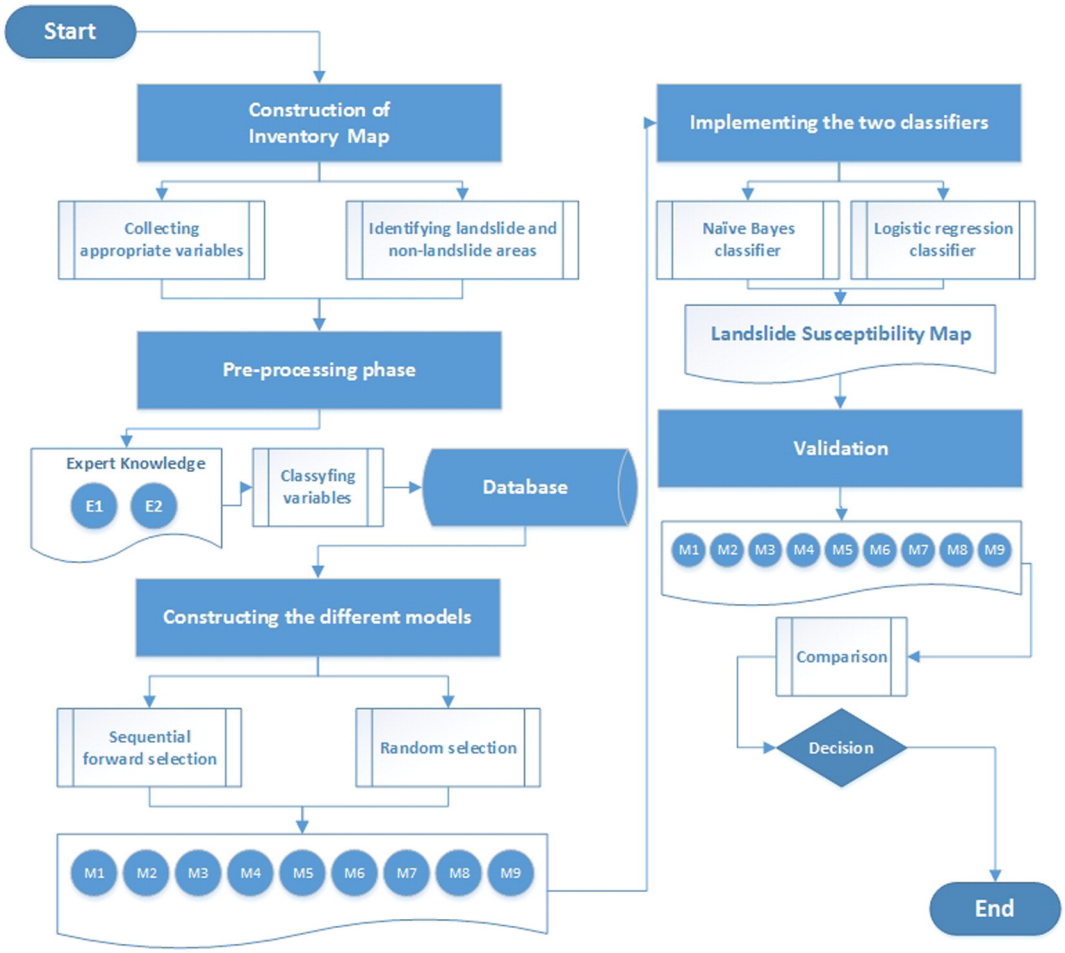

Fig. 1. Flowchart of the developed methodology.

In regard to the second sub-process, the classification of the variables was prepared according to expert knowledge, which follows the international experience and practice. That means that each variable, that may be represented both as a nominal or categorical, is reclassified and ranked according to their susceptibility to slide, assuming that the first class has a lower susceptibility index and that the last one has the highest susceptibility index.

The third and fourth sub-process is both necessary data mining procedures. Specifically, by implementing multicollinearity analysis an estimation of the correlation among the predictor features can be calculated (Dormann et al., 2013; Tien Bui et al., 2015a, 2015b). For this purpose, the proposed methodology uses the variance inflation factor (VIF) (Marquardt, 1970; Weisberg and Fox, 2010). Although no rules exist for interpretation of VIF, the most common rule of thumb is using 10 as a threshold for severe multicollinearity, while several authors apply a very strict threshold of 2, above which variables are considered multicollinear and are excluded from the model (O'brien, 2007; Van Den Eeckhaut et al., 2006, 2010; Guns and Vanacker, 2012).

Besides the multicollinearity analysis the feature selection process was also implemented. According to Guyon and Elisseeff (2003), the objective of feature selection is three-fold: improving the prediction performance of the predictors, providing faster and more cost-effective predictors and providing a better understanding of the underlying process that generated the data. In the present study, an entropy-based feature evaluator, Gain Ratio along with a ranker search method has been used to rank features according to their evaluated gain ratio with respect to the class (Cooper and Herskovits, 1992; Dash and Liu, 1997). This process was necessary in order to create the different set of models that were based upon the complexity of the models, the number of variables used.

2.3. Constructing the different set of models

As proposed by the methodology three (3) different training and validating data sets were randomly produced from the total number of landslide and non-landslide areas. Specifically, by utilizing the subroutine subset wizard that is embodied in the Geostatistic toolbox (ArcMap 10.1), the first data set contained an initial number of data that equaled to approximately 80% of the total number of available data. The dataset where further partitioned into a training and validating data (80% for training and 20% for validating data) randomly utilized by the same subset wizard. The second data set contained the data from the first set plus a 10% additional data that had been divided into training and validating data as previous, to equal 90% of the available data. The third data set as previously contained the data from the second set plus the remaining 10% divided again into training and validating data. Each data set contained equal numbers of landslide and non-landslide areas.

2.4. Implementing the two classifiers

Two type of classifiers have been evaluated; a distinctive classifier that uses LR and a generative classifier that employs a NB algorithm. Paragraphs 2.4.1 and 2.4.2, describe in brief the theory and mathematical formula on which each classifier is based.

2.4.1. Naïve Bayes algorithm

Bayesian Classification is a process that estimates the probability of a new observation belonging to a predefined category, using a probability model defined according to the theory of Bayes (Cheeseman and Stutz, 1996). The technique assesses the prior probability of each category based on a large set of training data, that are described by a number of variables, and assumes that classification could be estimated be calculating the conditional probability density function and the posteriori probability (Soria et al., 2011). The posteriori probability could be calculated according to the equation:

$$p(C_j|X) = \frac{p\left(X|C_j\right) * p(C_j)}{p(X)} \quad (1)$$

where $p(C_j|X)$ is the probability the unknown observation X belongs to category C_j and is called posteriori probability, $p(X|C_j)$ is the probability, given category C_j, an unknown observation belongs to this category, $p(C_j)$ is the prior probability the unknown observation X to be observed in category C_j, and $p(X)$ is the prior probability of the unknown observation X the is the same for each category C_j.

In the case where all the variables that describe the training data are independently and each of them contributes equally to the problem of classification, a simple method for Bayesian classification known as Naive Bayes has been developed (Cestnik et al., 1987; Domingos and Pazzani, 1997; Soria et al., 2011). Because of the conditional independence assumptions the probability $p(X|C_j)$ could be calculated by the following equation (John and Langley, 1995):

$$p(X|C_j) = \prod_{i=1}^{k} p\left(\frac{x_i}{C_j}\right). \quad (2)$$

According to John and Langley (1995) within each class, the values of numeric attributes are normally distributed, in terms of its mean and standard deviation and that the conditional probability could be calculated by:

$$p\left(\frac{x_i}{C_j}\right) = \frac{1}{\sqrt{2\pi}std} e^{\frac{-(x_i-\mu)^2}{2std^2}} \quad (3)$$

where μ is then mean and δ is the standard deviation of x_i.

Usually, the learning algorithm searches for the most probable hypothesis through a set of candidate hypotheses. It searches the maximum posteriori hypothesis according to the following equation.

In the case of landslide susceptibility assessment, given an incidence that consists of k landslide related variables, y_j is the Boolean output of the analysis which expresses the prediction landslide or non-landslide areas. The prediction is made for the class with the largest posterior probability according to the following equation (Tien Bui et al., 2012c):

$$y_j = \text{argmax}P\left(y_j\right) \prod_{i=1}^{k} P\left(\frac{x_i}{y_j}\right) \quad (4)$$

where, j = {landslide, non-landslide}.

2.4.2. Logistic regression

LR is among those statistical methods that have been proven to be highly reliable when performing a landslide susceptibility assessment (Dai et al., 2002; Ayalew and Yamagishi, 2005; Yesilnacar and Topal, 2005; Gorsevski et al., 2006; Lee and Pradhan, 2007; Yilmaz, 2010; Akgun et al., 2012; Wang et al., 2013). The independent variables in this model are considered as predictors of the dependent variable and can be measured on a nominal, ordinal, interval or ratio scale, while the dependent variable is in a binary format. The relationship between the dependent variable and independent variables is nonlinear (Yesilnacar and Topal, 2005).

LR is thought as a special case of a generalized linear model; however, it is based on quite different assumptions concerning the relationship between the dependent and independent variables from those followed by linear regression models. The conditional distribution is a Bernoulli distribution rather than a Gaussian distribution, since the dependent variable has the form of a binary variable (presence or absence of landslides).

In logistic regression analysis the relationship between the occurrence and its dependency on several variables can be expressed by the following equation:

$$p = \frac{1}{1 + e^{-z}} \tag{5}$$

where p is the probability of a landslide occurrence. The probability can take values from 0 to 1 on an S-shaped curve and z is the linear combination of a set of landslide related variables. Logistic regression involves fitting an equation of the following form to the data:

$$z = b_0 + b_1x_1 + b_2x_2 + \ldots + b_nx_n \tag{6}$$

where b_0 is the intercept of the model, the bi (i = 0, 1, 2, ..., n) is the slope coefficients of the logistic regression model, and xi (i = 0, 1, 2, ..., n) are the independent variables. The linear model formed is then a logistic regression of presence or absence of landslides (present conditions) on the independent variables (pre-failure conditions).

2.5. The landslide susceptibility map

For the NB classifier the landslide susceptibility index equals the maximum likehood probability expressed by Eq. (7), while for the LR classifier it equals with the probability of occurrence that is calculated by Eq. (8).

$$LSI_{NB} = \frac{P(y_{landslide})\prod_{i=1}^{k} P\left(\frac{x_i}{y_{landslide}}\right)}{P(y_{landslide})\prod_{i=1}^{k} P\left(\frac{x_i}{y_{landslide}}\right) + P(y_{non\text{-}landslide})\prod_{i=1}^{k} P\left(\frac{x_i}{y_{non\text{-}landslide}}\right)} \tag{7}$$

$$LSI_{LR} = \frac{1}{1 + e^{-\left(b_0^{*} + b_1 * x_1 + \ldots + b_n * x_n\right)}} \tag{8}$$

In both classifiers, a value of landslide susceptibility index close to zero corresponds in non-landslide areas while a value close to one corresponds in landslide areas. However, in order to produce a landslide susceptibility map using classes of landslide susceptibility that corresponds to the degree of susceptibility, the natural break method for the determination of the class intervals was applied (Feizizadeh and Blaschke, 2013). Classes then were identified as follows: very high susceptibility (VHS), high susceptibility (HS), moderate susceptibility (MS), low susceptibility (LS) and very low susceptibility (VLS).

2.6. Validation and comparison

For the estimation of the performance of the classifiers the developed methodology utilized statistical evaluation criteria which are based on the true positive, false positive, true negative and false negative rates. A sample is classified as true positive when it is estimated to belong to the i^{th} class and it truly belongs, as false positive when it is estimated to belong to the ith class but it truly does not belong, as true negative when it is estimated not belonging to the ith class and it truly does not belong to it and final as false negative when it is estimated not belonging to the ith class but it truly does belongs to it (Tsangaratos and Benardos, 2014). Specifically, two criteria were calculated; the overall accuracy of the training data that could be thought as an index that express the successful power of the model and the overall accuracy on the validation data that could be thought as an index that express the predictive power of the model. In both cases, the accuracy was calculated as the ration of the true positive plus the true negative to the total number of data (Murakami and Mizuguchi, 2010).

In addition there was also an evaluation on the spatial product of the classifiers, the landslide susceptibility maps. An ideal landslide susceptibility map must satisfy two spatial effective rules (Can et al., 2005; Pradhan and Lee, 2010b): to have an increasing landslide density ratio when moving from low susceptible classes to high susceptible classes and the high susceptibility class to cover small extent areas.

In the present study the validation processes also included the comparison of the landslide susceptibility maps with the landslide locations using the receiver operating characteristic (ROC) curve analysis (Fawcett, 2006). Using the landslide grid cells in the training dataset, the success-rate results were obtained, while the validation dataset were used for the construction of the prediction-rate curves (Chung and Fabbri, 2003). The area under the ROC curve (AUC) has been used as a metric to access the overall quality of the predictive models by evaluating the models ability to anticipate correctly the occurrence or non-occurrence of predefined events (Hanley and McNeil, 1982; Negnevitsky, 2002; Fawcett, 2006). If AUC is close to 1, the outcomes of the analysis are excellent, while if the AUC is closer to 0.5, the less accurate the result of the analysis is.

3. Study area

The wider area of research is located on the west side of the central part of Pindos Sierra and range between the geographic regions of Epirus and Thessaly. On the east side, the area is defined by the Basin of Acheloos River, while on the west, north and south it is defined by the Basin of Arachthos River and its tributaries. The area comprises one of the historical centers of the Vlach culture in Pindos, with a cluster of significant historical villages Sirako, Killarites and Pramanda (Fig. 2). The main research area is approximately 222 km^2, clarified as the Kallaritikos watershed, a sub basin of the Greek water district Epirus (coded as 05), between longitudes 245,000 and 260,000 and latitudes 4,370,000 and 4,395,000, based on the coordinate system GGRS87/Greek Grid. The Kallaritikos watershed drains the northern region of Tzoumerka Mountains and the southern edge of Lakmos Mountain. The watershed is characterized by a dense hydrographic network that consists of a number of major rivers and tributaries with permanent or seasonal flow and several waterfalls. Most important rivers with significant flow are Kallaritikos River, Mellisourgiotiko River and Matsoukiotiko River (Fig. 2).

Concerning the geomorphological settings, the relief of the research area is mainly shaped because of the geological structure, the recent tectonic activity and the weathering and erosion mechanisms. It is characterized as mountainous with strong relief, massive rocky limestone ridges, high peaks, canyons with sloping slopes and extensive subalpine plateaus. In particular, the highest observed altitude is 2354 m, while the lowest point is at 325 m above sea level. Areas with slopes greater than 46° cover approximately 2.85% of the total area, while areas with slope angle less than 15° cover about 11.92%. The majority of the area, about 39.83%, is characterized by slope angle that range between 16° and 30° followed by areas with slope angle ranging between 31° and 45°. Catalyst in shaping the terrain of the study area was the existence of two major rivers in the region, Arachthos and Acheloos. More specifically, the main watercourse of Arachthos reaches 105 km in length, with the main direction being NNE - SSW. Respectively, Acheloos is along the main river bed with a length of about 220 km. The basin is of equally extent with the corresponding Arachthos basin, covered mainly by limestone formations occupying the whole area of the eastern side of Tzoumerka Mountains.

The study area is covered mostly by extensive and dense thickets, forests and local woodlands. In particular, 37.31% of the area comprises natural grassland, 20.11% transitional woodland/scrub, 15.09% sparsely vegetated areas, 9.23% agriculture and natural vegetation, 6.33% coniferous forest and 5.66% sclerophyllous vegetation.

Concerning the geology of the wider region, it consists of formations that are part of the Ionian tectonic zone, mainly constituted by Upper Eocene – Lower Miocene sedimentary sequences, as well as part of the Olonos - Pindos tectonic zone, where Upper Cretaceous – Eocene

sedimentary sequences outcrop (Brunn, 1956; Aubouin, 1959). The Ionian tectonic zone is characterized by the large successive anticlines and synclines that overthusts to the west, heading mainly NW - SE. Generally, the formations of the geological structure of Ionian Zone is separated into three main stratigraphic units; at the base there are the evaporites with gypsum and ternary breccia, followed by limestone formations and clastic flysch series. The stratigraphic sequence is completed with the Neogene and Quaternary formations. Correspondingly, Olono - Pindos zone is considered as a deep groove between the Pelagonian and Gavrovrou zones, showing large variations in sedimentation. Geological formations that are present and cover the wider area are siliceous limestone, chert and radiolarites alternating with platy limestones, calcareous Cretaceous limestone and pelagic Upper Cretaceous limestone with flint.

More specifically, concerning the lithological formations that cover the two tectonic zones in the research area, they are listed as follows (Fig. 3): For Ionian zone: (a) Flysch formation that corresponds to the transition between the flysch of Ionian zone and the flysch of Gavrorou zone, separated into flysch with materials from mantle erosion, flysch with predominant sandstone phases and flysch with predominant siltstone phases (fi), (b) massive, brecciated limestone with intercalations of limestone with lenses of silica (e–k) and (c) brecciated limestone with intercalations of pelagic thin bedded limestone (J_{k8i}). For Olono – Pindos zone: (a) flysch formations that consists of alternating siltstones and sandstones with less frequent participation of conglomerates and intermediate lithological types (fo), (b) pelagic limestone with lenses of silica and intercalations micro-brecciated limestone (k_8), (c) micro-brecciated limestone with silica layers and marly and shale formations (k_{3-4}), (d) chert formations with intercalations of pelagic limestone (J-k_1), (e) grey limestone with intercalations of marl formations (t_{s-k}) and (f) post alpine formations which are separated into alluvial deposits, scree and river terraces (al) and glacial deposits (mg_1, mg_2) at the foot of the mountains. According to the Köppen climate classification system (Aguado and Burt, 2012), the wider area is characterized as Mediterranean type (Csa) with heavy winters and cool summers. During the winter, the temperature reaches low levels, rainfall and snowfall is abundant with a notable prolonged snow cover. The cloud coverage is high, while frosts occur from October to May. Summer is cool, with several local rainstorms.

The rainy season is from October to May accounting to almost 90% of the total amount of annual rainfall with approximately reaches 47 to 150 mm/month. December appears the rainiest month (153.2 mm) followed by November (140.3 mm), while the driest month appears to be August (13.9 mm) followed by July (17.7 mm). The annual average temperature is 13.67 °C with the highest and lowest average temperature being 19.17 °C and 8.22 °C respectively (Fig. 4). The climate data were obtained from the University of East Anglia Climate Research Unit (CRU) and referred to a period over 100 years between 1901 and 2008 (Jones and Harris, 2008).

Regarding the landslide phenomena encountered in the area, they are mainly caused due to the physical conditions (weathering and fracturing) and the general geotechnical behavior of the geological formation covering the area. In most cases, the main triggering factors were the combined action of intense rainfall events and anthropogenic

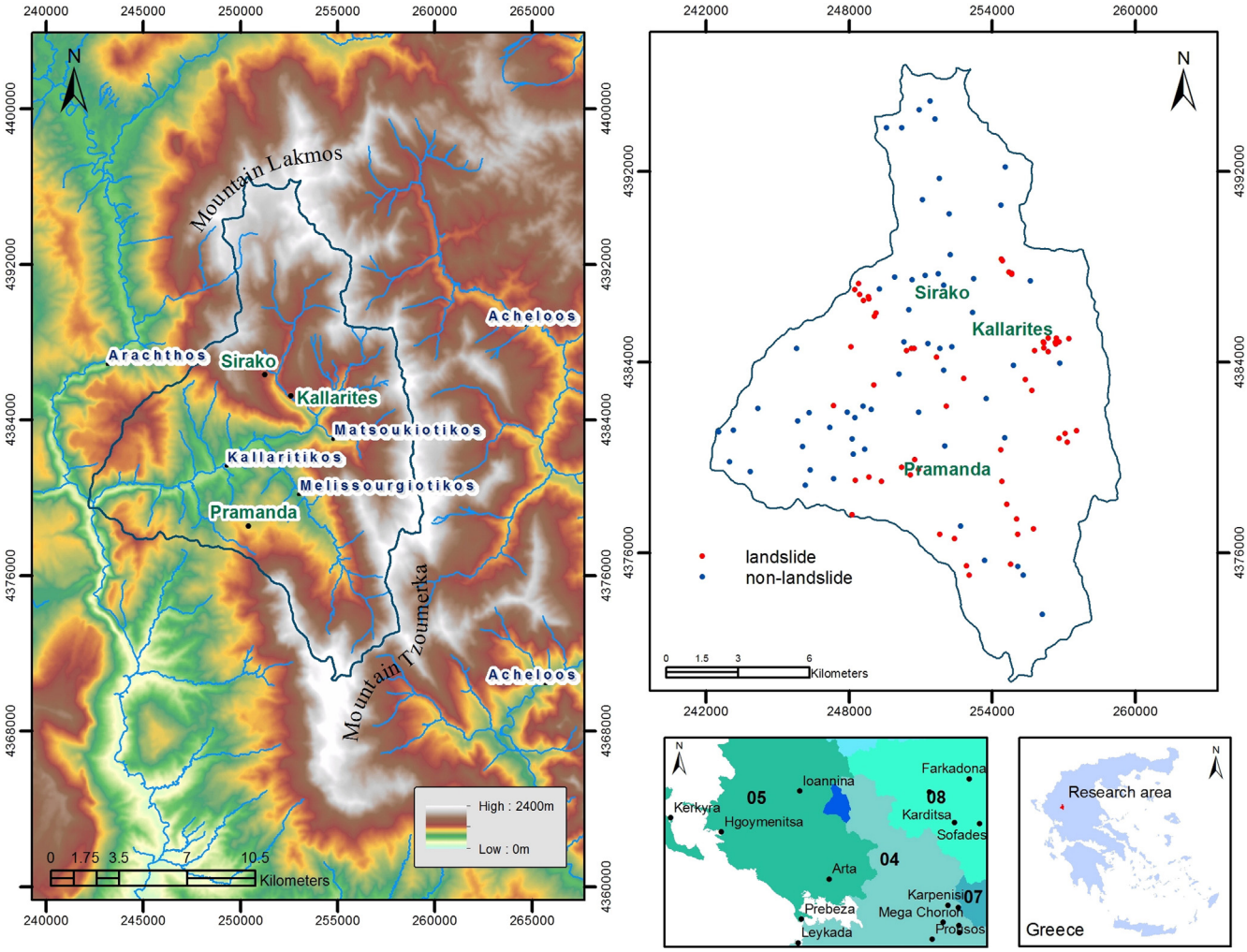

Fig. 2. The study area and the spatial distribution of landslide and non-landslide area.

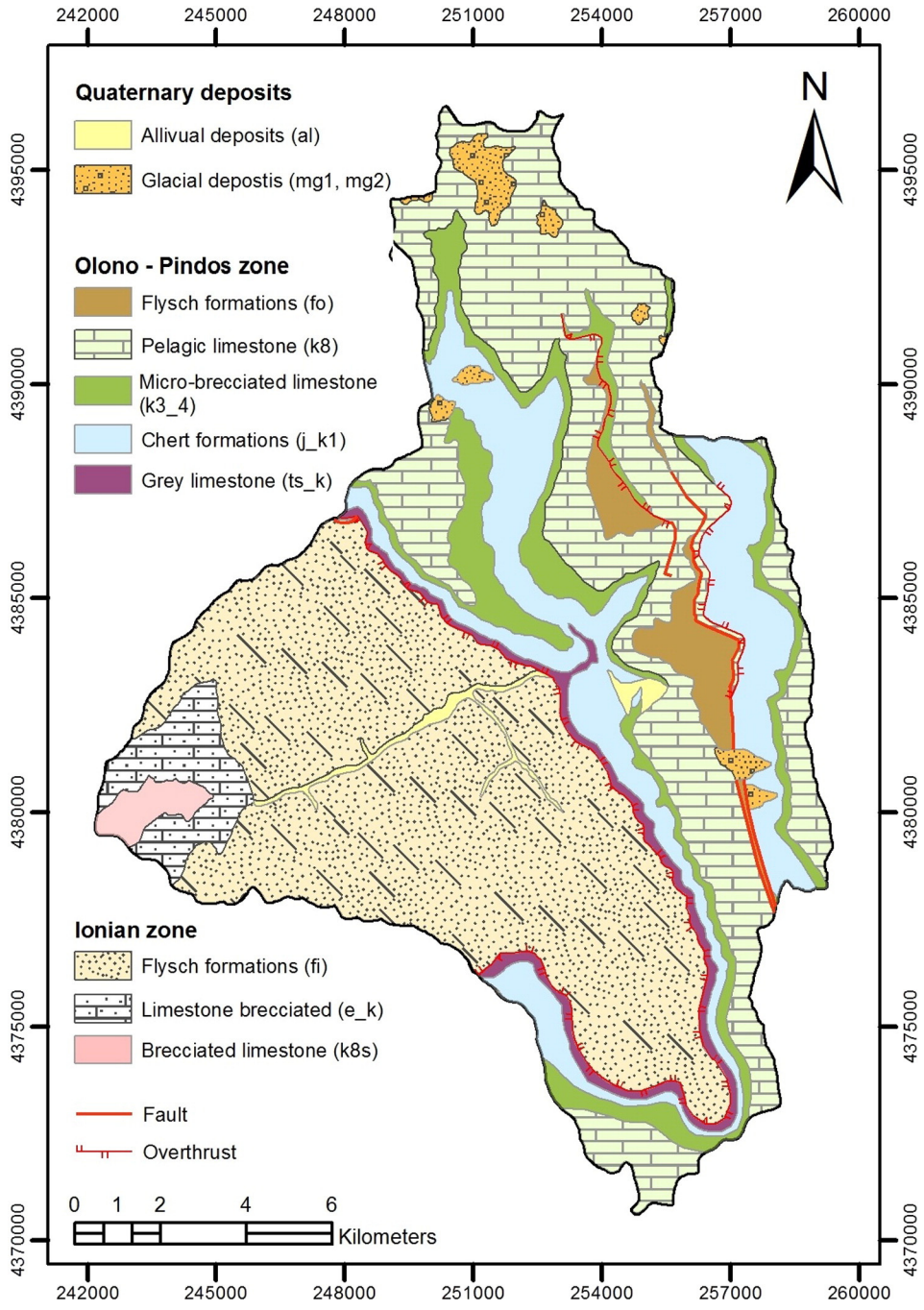

Fig. 3. The geological formation of research area.

activities. The majority of the reported landslide events were located along the road network and within the residential complexes, classified according to the Varnes Classification System (Varnes, 1984), as rotational and translational slides and also rockfalls. In particular, rotational and translational slides are usually developed at the weathering mantle and the upper fragmentation zone of flysch formations. On the other hand rockfalls are mainly reported on slopes with intense inclination that are covered by chert formations. The maximum size of the landslide surface exceeded 6500 m^2, while the average size was about 320 m^2.

4. Results

According to the methodology the first step is to construct the inventory database which contained information about the landslide and non-landslide areas. Regarding the landslide areas, they are mainly located along the road network and within the residential complexes. They are classified according to the Varnes Classification System (Varnes, 1984), as rotational and translational slides and also rockfalls. In particular, rotational and translational slides are usually developed at the weathering mantle and the upper fragmentation zone of flysch formations. On the other hand rockfalls are mainly reported on slopes with intense inclination that are covered by chert formations. The non-landslide areas were identified by using airborne imagery and extensive field investigation. A total of 116 sites, 58 landslide and 58 non-landslide areas was recorded the spatial distribution of which is shown in Fig. 2.

The second step is to construct a geodatabase that contains landslide related variables and to classify those variables into classes of landslide susceptibility. The geo-environmental conditions of the research area that are used for the assessment of landslide susceptibility are described

by the following seven variables: engineering geological units, slope angle, slope aspect, distance from tectonic features, distance from river network, mean annual rainfall and distance from road network.

Concerning the first variable, the geological formations that cover the research area (Aubouin, 1959) were grouped into four categories based on their engineering geological behavior, the spatial distribution of failures identified in the region, but also the experience and knowledge that has been recorded in related studies (Koukis et al., 2005; Sabatakakis et al., 2013). Specifically, the following were found and classified: (a) quaternary loose, fine grained deposits that consist mainly of cobbles, pebbles, grits and sands with low proportions of fines, such as clayey silts and sandy silts; (b) limestones formations, that are characterized as Pelagic, thin to medium – bedded, often micro - brecciated with nodules or lenticular silica layers and local thin intercalations of shales; (c) flysch formations with alternating siltstones and sandstones and frequent participation of conglomerates and intermediate lithological types, and (d) chert formations with limestone interbeds (Fig. 5a). The fault density maps was also constructed based on the geological map and was classified into three zones of influence: a. <250 m, b. 251–500 m and c. >501 m (Fig. 5b).

A digital elevation model (DEM) with a spatial resolution of 20 × 20 m was generated from national topographic maps in scale 1:50.000. Based on the DEM data, slope angle, slope aspect and distance from the river network were constructed. Specifically, four classes for slope angle have been identified and classified: a. 0°–15°, b. 16°–30°, c. 31°–45° and d. slopes greater than 46° (Fig. 5c). In accordance to the previous, four classes for slope aspect have been identified and classified: a. 225°–270°, b. 45°–90°, c. 90°–135°, 270°–315°, and d. 315°–45°, 135°–225° (Fig. 5d). Concerning the river network density map, it was formed using the DEM data and further classified into three zones of influence: a. <100 m, b. 101–300 m, and c. >301 m (Fig. 5e).

The mean annual rainfall was not evaluated as an external variable but rather identified as an intrinsic variable. The mean annual rainfall data have been used in several landslide susceptibility studies (Feizizadeh et al., 2013, 2014; Tien Bui et al., 2015a, 2015b). In most cases the underlying assumption of utilizing the rainfall as an intrinsic variable is that it could serve as a variable that represents the climatological conditions of an area, which in turn play an important role in slope instability. Lacking detail rainfall data concerning rainfall intensity values, the usage of mean annual rainfall could provide the necessary information about the response of geological formations on the likelihood of sliding to different climate conditions.

The mean annual rainfall map was constructed from the annual average of rainfall data provided by the Institute of Meteorology and Hydrology for the period 1950 to 2000 using the Inverse Distance Weighted method. For the study area, three zones of mean annual

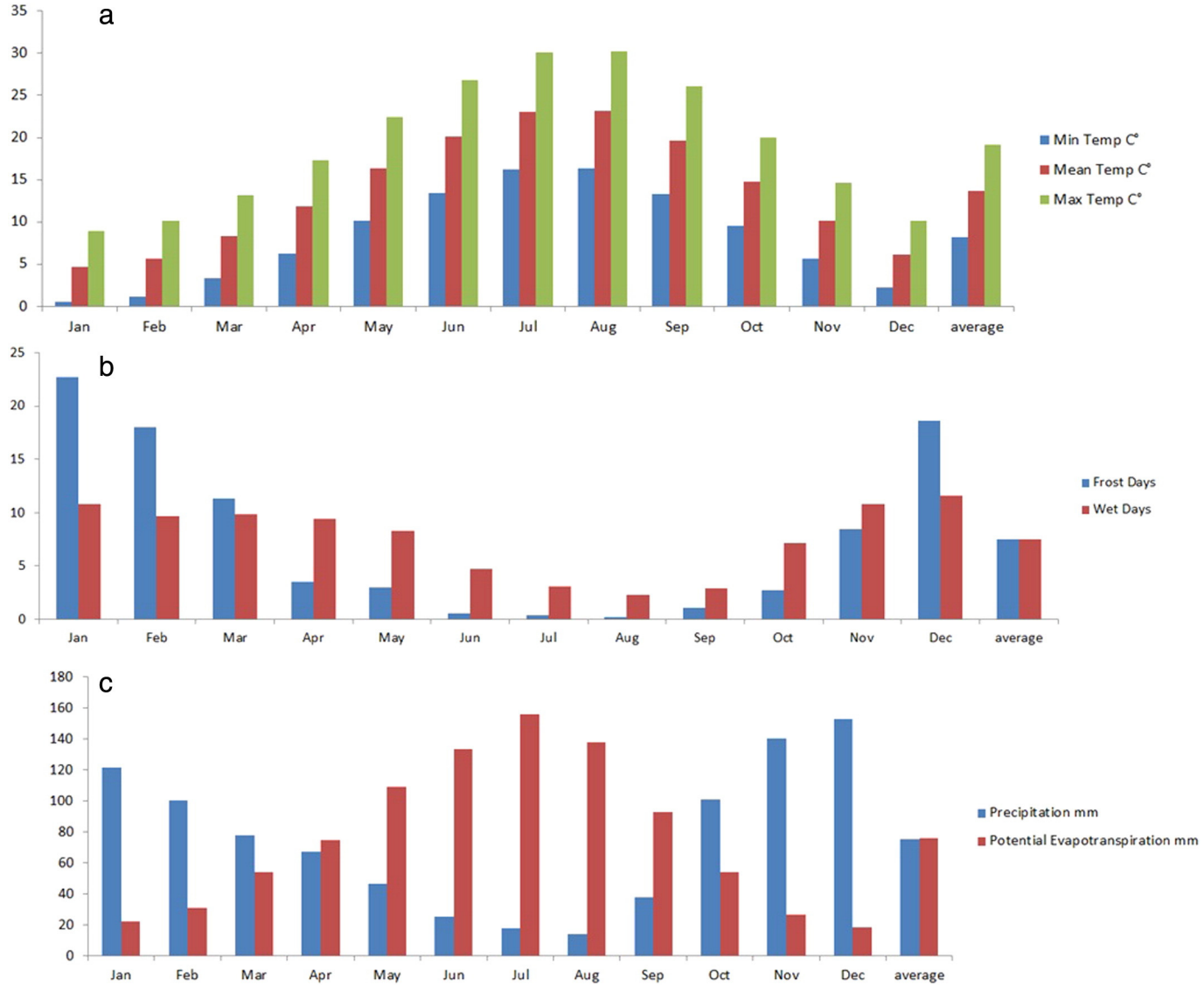

Fig. 4. The climate data; a. temperature data, b. frost and wet days, c. precipitation and potential evapotranspitation.

rainfall were formed: a. <1350 mm, b. 1351–1650 mm and c. >1651 mm (Fig. 5f).

Finally, the distance from the road network was constructed based on the national topographic maps and classified into three zones of influence, characterizing the distance of landslide incidence from the road network: a. <100 m, b. 101–300 m, and c. >301 m (Fig. 5g).

The analysis of VIFs yields values between 1.075 and 1.410, values that stand below the threshold of 2 resulting in using all the predictor variables for further analysis (Table 1). By implementing FSP the available landslide related features where ranked according to their evaluated gain ratio as follow; (1) distance from road network, (2) distance from tectonic features, (3) distance from river network, (4) mean annual rainfall, (5) slope angle, (6) slope aspect and (7) engineering geological units, respectively (Table 1).

Based on the results of the FSP and the multicollinearity analysis, three different set of models were produced (Table 2). The first set of models (models 1, 4 and 7) used the total number of variables (distance from road network, distance from tectonic features, distance from river network, mean annual rainfall, slope angle, slope aspect and engineering geological units). The second set of models (models 2, 5 and 8) used the five (5) most significant variables (distance from road network, distance from tectonic features, distance from river network, mean annual rainfall and slope angle), while the last set of models (models 3, 6 and 9) used the three (3) most significant variables (distance from road network, distance from tectonic features and distance from river network).

Following the suggested by the methodology procedure, a set of three different training and validation datasets were constructed. According to the rule of thumb about the size of training samples, a minimum of 60 training samples is required for models 3, 6 and 9; a minimum of 100 training samples is required for models 2, 5 and 8 and a minimum of 140 training samples is required for models 1, 4 and 7. This means that only models 3, 6, 8 and 9 satisfies the rule of minimum training samples. Table 2 illustrates the characteristics of each model.

Table 3 provides the accuracy of the training and validation datasets for each model. From the outcomes of the experiment it's induced that in logistic regression analysis, for the first set of models (1, 4 and 7), as the training samples increased the training accuracy of the models decreased, however their predictive power increased, as can be shown from the accuracy index estimated from the test dataset. For the second set of models (2, 5 and 8), as the training samples increased, the accuracy of the models for the training and testing datasets also increased. For the third set of models (3, 6 and 9), as the training samples increased, the training accuracy remained relatively the same, while the testing accuracy increased.

In logistic regression analysis, when using a small number of training data, models 1, 2 and 3, as the complexity of the models decreased so did the accuracy of the training and testing datasets. When more data are used (80 data, models 4, 5 and 6), and the complexity of the models decreased, the accuracy of the training still decreases, however the accuracy of the testing dataset increased.

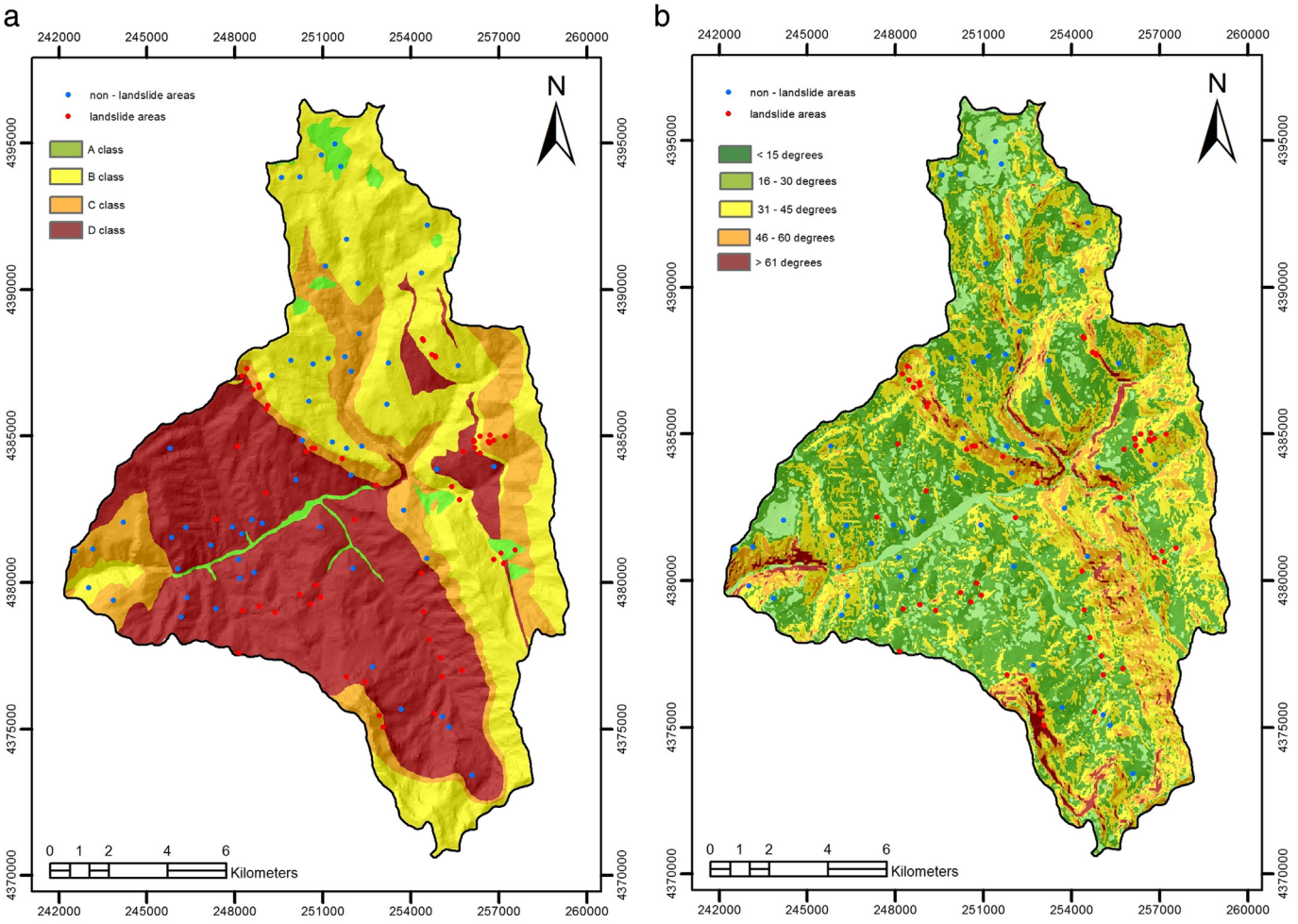

Fig. 5. The landslide related variables; a. engineering geological units, b. slope angle, c. slope aspect, d. mean annual rainfall, e. distance to tectonic features, f. distance to river network, g. distance to road network.

When even more training data are used (92 data, models 7, 8 and 9), the reduction of models complexity has the same effect on the training accuracy, while the predictive accuracy tops when 5 variables are used. The highest accuracy is achieved for model 8 (92 data and 5 variables), 82.61%.

Concerning the implementation of the NB approach for the first set of models (1, 4 and 7), as the training samples increased the training accuracy of the models increased, and so did their predictive power. For the second set of models (2, 5 and 8), as the training samples increased, the accuracy of the models for the training and testing datasets also increased. For the third set of models (3, 6 and 9), as the training samples increased, the training and testing accuracy remained relatively the same.

In NB approach, when using a small number of training data, models 1, 2 and 3, as the complexity of the models decreased so did the accuracy of the training, however the predictive power of the models increased. When more data are used (80 data, models 4, 5 and 6), and the complexity of the models decreased, the same pattern is identified, however with having higher values.

When additional training data are used (92 data, models 7, 8 and 9), the reduction of models complexity has the same effect on the training accuracy, while the predictive accuracy tops when 5 variables are used. The highest accuracy is also achieved for model 8 (92 data and 5 variables), 87.50%.

For both classifiers model 8 provides the best prediction, with NB achieving an 87.50% accuracy and LR 82.61%.

The constructed map corresponds to the degree of susceptibility using classes of landslide susceptibility according to the natural break method for the determination of the class intervals (Feizizadeh and Blaschke, 2013). Classes identified are described as follows: very high susceptibility (VHS), high susceptibility (HS), moderate susceptibility (MS), low susceptibility (LS) and very low susceptibility (VLS) (Fig. 6a, b).

When applying LR analysis the percentage of landslides located within the zones high and very high susceptibility is estimated to be 74.60%, while for NB the percentage is 77.56% (Fig. 7). In addition the area the percentage of the areas within the zones high and very high susceptibility is estimated to be 28.37% and 20.41%, respectively.

The results of the implementation of the two classifiers on model 8 were also validated using the training and validation dataset through the use of the ROC graphs and the success and prediction rate curves, which are summarized by the calculation of AUC values. Fig. 8a illustrates the results of the validation analysis indicating that both classifiers have good prediction capabilities. In particular, the highest AUC value was achieved by the NB classifier for both the training and validating datasets, 0.875 and 0.806 respectively. The LR classifier achieved a lower AUC values for the training and validating datasets, 0.844 and 0.711 respectively.

The next step was to estimate how well the two models had classified the research area according to the landslide susceptibility classes and the cumulative percentage of the observed landslide occurrence. The validation process was performed by comparing the produced landslide susceptibility map with the actual landslide locations using the success rate and the prediction rate methods. Fig. 8b the success and prediction rate curve of the two models.

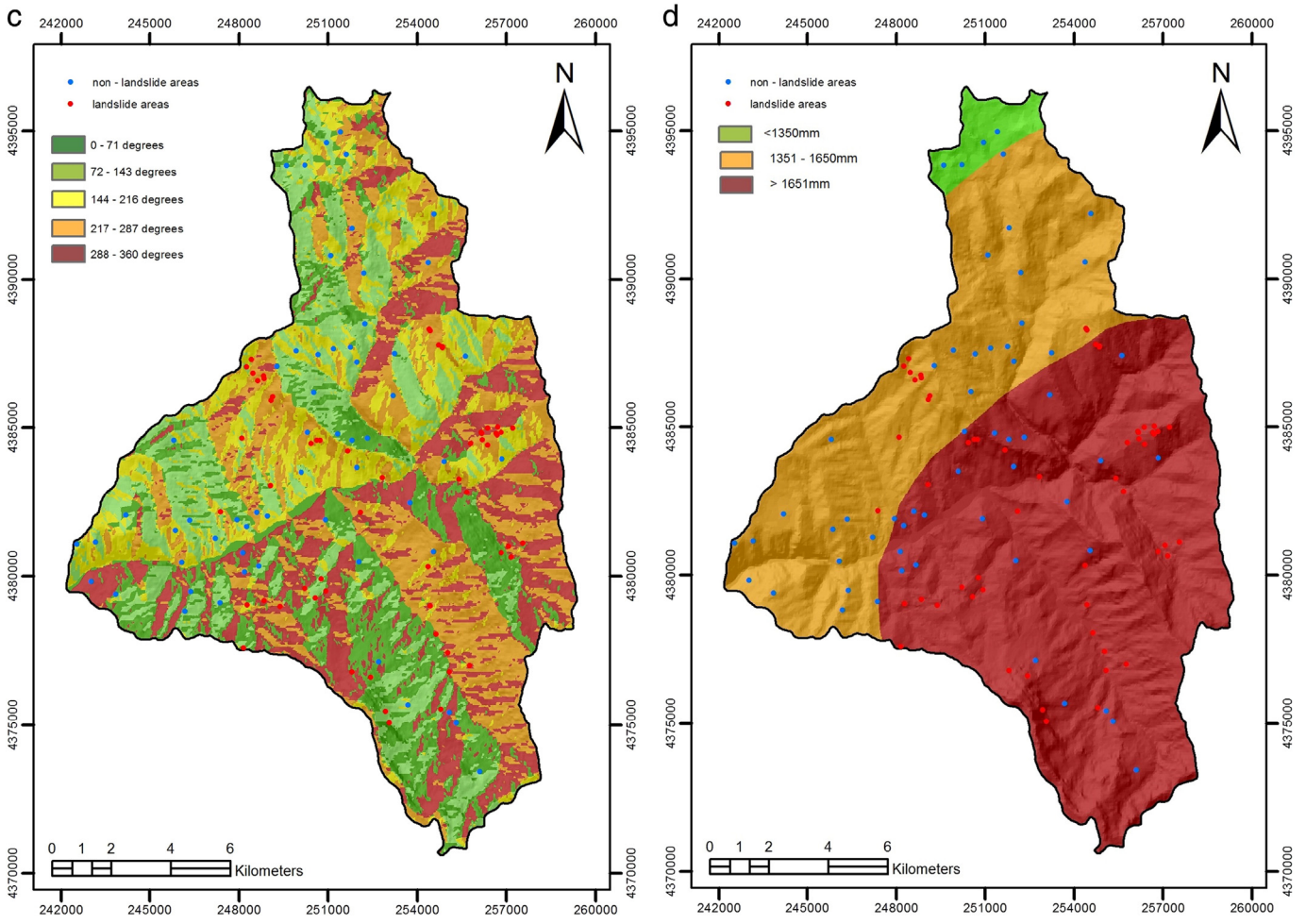

Fig. 5 (*continued*).

The calculated value for each of the two models for the area under the curve (AUC) showed similar results. The efficiency of the NB classifier was the highest among the models (AUC = 0.811), followed by the LR classifier (AUC = 0.771). Also, the predictive power of the NB classifier was the highest among the two models (AUC = 0.946), followed by the LR classifier (AUC = 0.786).

5. Discussion

The concept of landslide susceptibility is associated with the understanding of the influence of the physical factors that represent the cause and not the reason for the manifestation of landslides (Varnes, 1984; Soeters and van Westen, 1996; Guzzetti et al., 1999, 2005; Van Den Eeckhaut et al., 2006). The most common physical factors that are used are geomorphology, geology, tectonic, hydrology, soil and vegetation features, land use and human practices. During the last two decades, many research papers were published having as main objective the assessment of landslide susceptibility using different statistical and probabilistic methods and techniques that utilized GIS technology. However, only few had studied the influence of models complexity and the size of the training dataset. In this context, the present study evaluates the performance of two types of classifiers; a distinctive classifier that implements LR and a generative classifier that employs a NB algorithm. Although both NB and LR classifiers are considered as linear classifiers, they differ in the way they classify. LR classifiers make a prediction using a direct functional form, whereas NB classifiers figures out how the data was generated given the results.

The two objectives of the study was; a) to predict trends and patterns in response to the evolution of landslide processes and produce on that bases a landslide susceptibility map, and b) to investigate the behavior of the two classifiers when changes are made on the number of landslide variables used and the size of training datasets.

From the pre-processing phase, the multicollinearity analysis revealed no multicollinearity among the seven conditioning variables, which had values that stand below the threshold of 2. Another interesting outcome during the pre-processing phase was the ascertainment that although lithology normally plays a significant role in the estimation of landslide susceptibility, in our study the FSP indicated that engineering geological unit is the least informative feature. This trend can be explained by the fact that the spatial distribution of landslide and non – landslide records is similar among the classes of the engineering geological units. However, the rotational and translational slides are connected with the anisotropic geotechnical behavior of the flysch formations, which in turn are largely influenced by the heterogeneous structure, the degree of looseness, weathering and fragmentation, the orientation of the discontinuities surfaces and the intense morphological relief.

On the other hand, the most informative variable is the distance from the road network followed by the variable distance from tectonic features and distance from river network. According to Mancini et al. (2010), the road network can be characterized as an activating agent and/or a susceptibility factor as it is an area of intense anthropogenic intervention. Extensive excavations, external loads, the removal of vegetation and the incorrect design and construction can result in the activation of mechanisms that favor landslide manifestation. The second and third in rank informative feature, distance from tectonic features

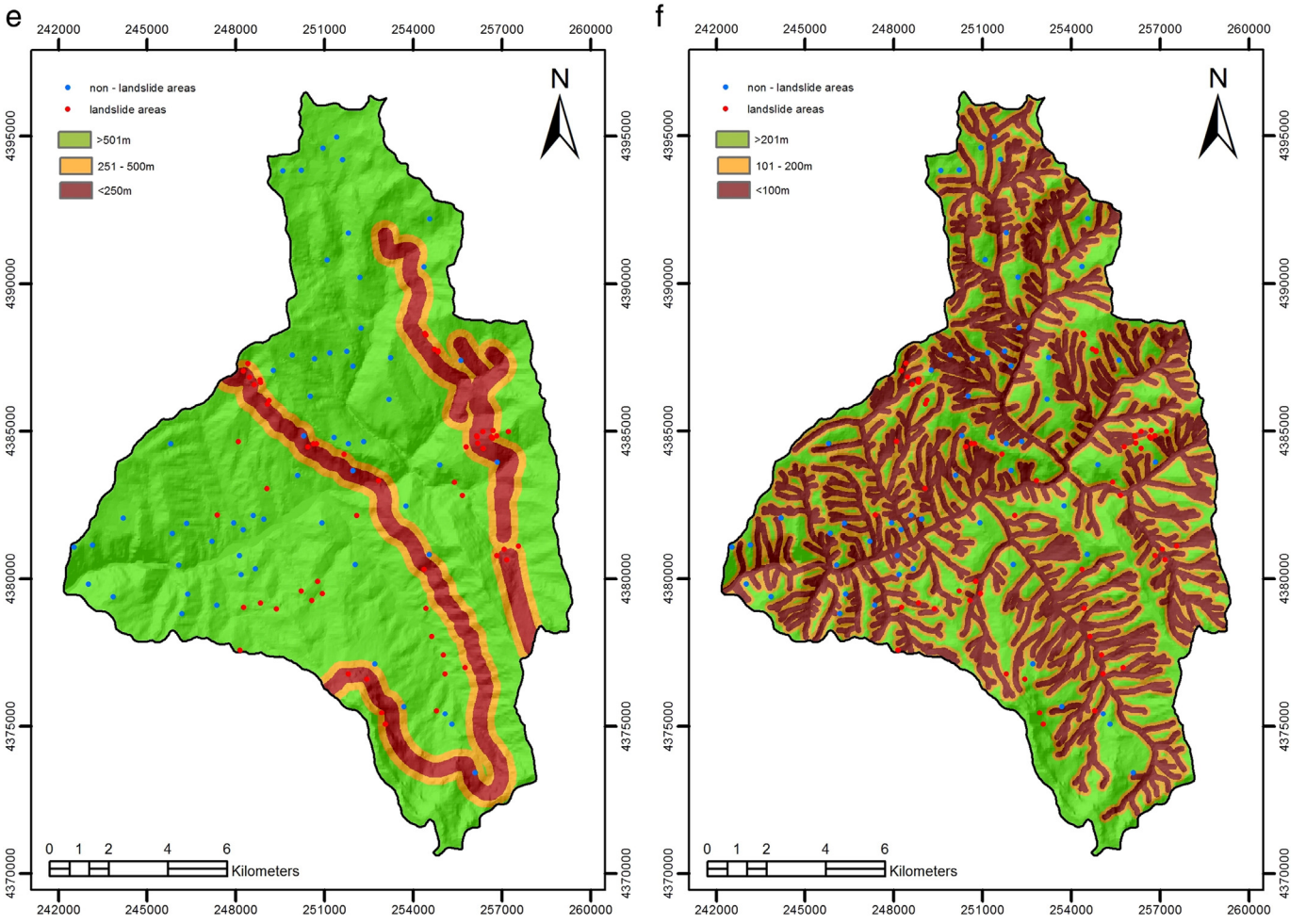

Fig. 5 (*continued*).

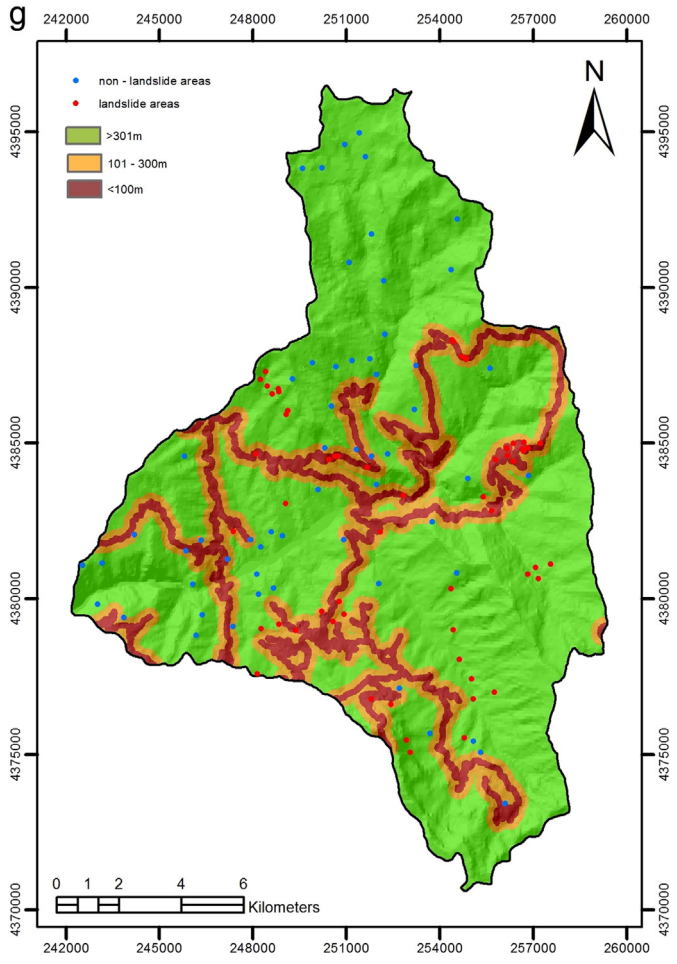

Fig. 5 (*continued*).

and distance from river network, reveals the dominant role of fault zones and dense river network in determining the stability of natural slopes.

From the interpretation of the results, a sound evaluation can be formulated; when applying logistic regression to models with relative small number of variables (models 3, 6 and 9) the increase of the size of the training data had a small influence in the training accuracy; however it had a positive contribution on the predictive power of the models. This pattern is not observed when applying Naïve Bayes algorithm. The influence of the size of the training data had an insignificant effect on the training and predictive accuracy. However, if one compares the performance of the two approaches, the Naïve Bayes approach provides a more accurate result. In the case of model 3, (74 training data and 3 variables) the predictive accuracy of the Naïve Bayes algorithm is almost double in value (85.71%) than the value of the logistic regression (42.85%). The behavior of the models that perform logistic regression implies over fitting of data. The models memorize training data rather than learning to generalize from trend. As already stated, according to the rule of thumb about the size of training samples, a minimum of approximately 60 training samples is required for models 3, 6 and 9. In the study for models 3, 6 and 9, 74, 82 and 92 training samples were used, respectively. Thus the estimation provided was sound and valuable.

Both classifiers have a similar behavior when applied to medium complexity models (models 2, 5 and 8) in which the increase of the size of the training data had a positive contribution on both the training and the predictive power of the models. In this situation the NB classifier gives the best results.

According to the rule of thumb about the size of training samples, a minimum of approximately 100 training samples is required for models 2, 5 and 8. Model 8, with 92 training samples is close enough to the required number thus the estimation provided could be used with great confidence.

In general, it seems that for LR classifiers when moving from smaller data sets and low complexity to larger data sets and high model complexity, training AUC values increase. On the other hand the test AUC values have a different behavior. They increase as the training data set increase, however they decrease as the model becomes more complex. The poor predictive performance of the models could be explained by the fact that it can exaggerate minor fluctuations in the data.

The NB classifier converges quicker than LR classifier, meaning that it can use less training data and still have a higher predictive power. However, the NB classifier can't learn interactions between the conditioning features and its performance depends on the absence of multicollinearity. On the other hand, the outcome of the analysis performed by LR classifier is not influenced by the assumption of conditional independence among the features and in addition provides a more sound probabilistic interpretation.

Table 1
Collinearity statistics and FSP results.

Variables	VIF values	Ranking FSP
Engineering geological units (EGU)	1.208	7
Slope angle (SL1)	1.284	5
Slope aspect (SL2)	1.075	6
Mean annual rainfall (RAIN)	1.287	4
Distance from river network (RIVER)	1.102	3
Distance from tectonic features (TECTONIC)	1.410	2
Distance from road network (ROAD)	1.153	1

Table 2
Models characteristics.

Models	Number of training data	Min number of training data	Parameters
Model 1	74	140	EGU, SL1, SL2, RAIN, RIVER, TECTONIC, ROAD
Model 2	74	100	SL1, RAIN, RIVER, TECTONIC, ROAD
Model 3	74	60	RIVER, TECTONIC, ROAD
Model 4	82	140	EGU, SL1, SL2, RAIN, RIVER, TECTONIC, ROAD
Model 5	82	100	SL1, RAIN, RIVER, TECTONIC, ROAD
Model 6	82	60	RIVER, TECTONIC, ROAD
Model 7	92	140	EGU, SL1, SL2, RAIN, RIVER, TECTONIC, ROAD
Model 8	92	100	SL1, RAIN, RIVER, TECTONIC, ROAD
Model 9	92	60	RIVER, TECTONIC, ROAD

Table 3
Accuracy of LR and NB classifiers.

Models	Accuracy train (%) LR	Accuracy train (%) NB	Accuracy validation (%) LR	Accuracy validation (%) NB
Model 1	92.65	82.35	57.14	57.14
Model 2	77.94	73.53	57.14	77.57
Model 3	76.47	76.47	42.85	85.71
Model 4	90.00	81.25	62.50	63.63
Model 5	82.50	80.50	75.00	78.28
Model 6	77.50	78.75	75.00	87.50
Model 7	86.96	84.78	63.63	64.29
Model 8	82.61	80.63	82.61	87.50
Model 9	77.17	75.00	77.17	86.37

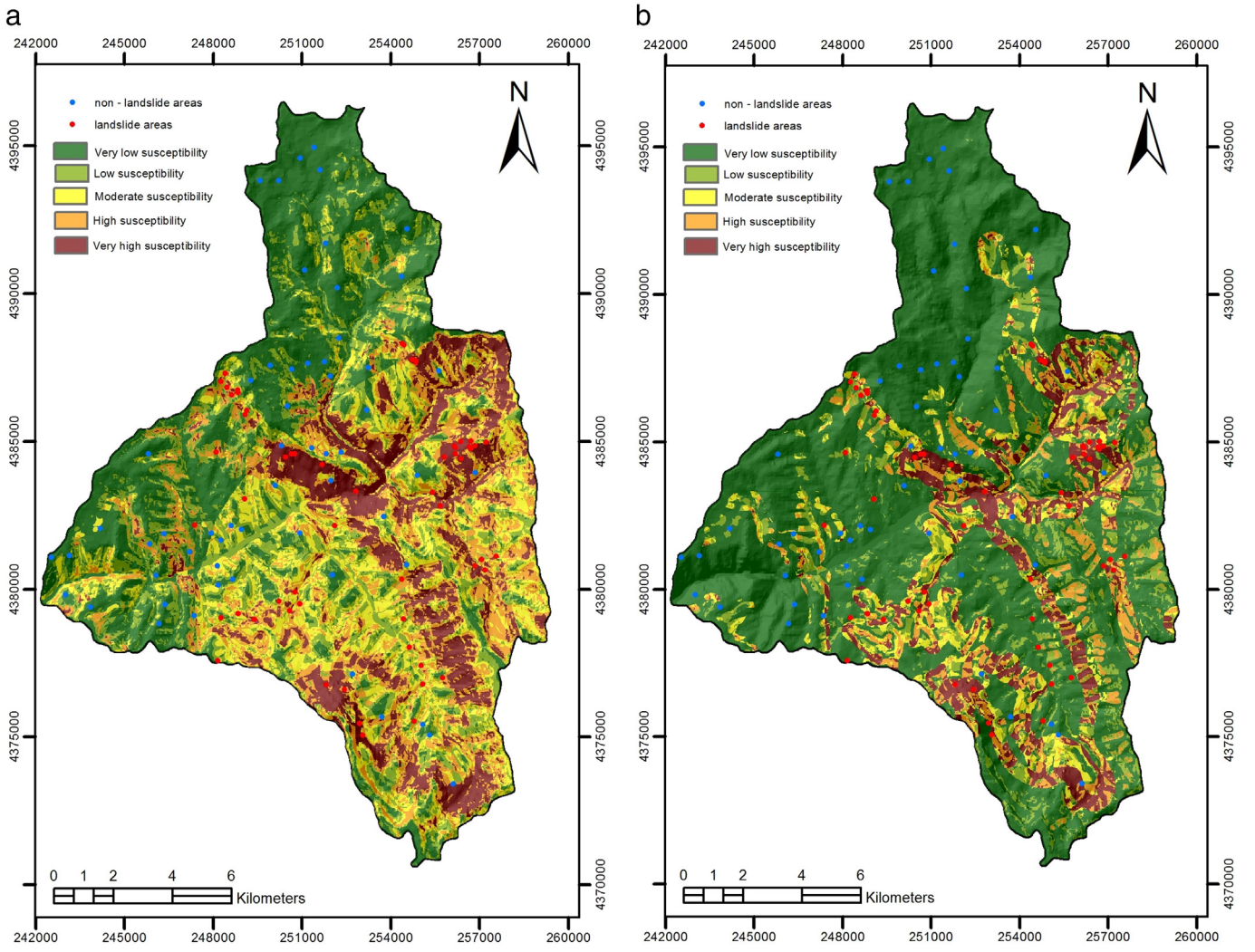

Fig. 6. Landslide susceptibility maps of the prediction model 8. a. LR classifier, b. NB classifier.

A general conclusion from applying both models can be extracted that the reduction of feature dimensionality helped in decreasing the size of the training samples and consequently improved the generalization performance of the classification algorithms.

Inspecting the produced susceptibility maps one can detect differences in the extent of the areas classified in low to very low susceptibility. Specifically, the two classes cover about 71% of the total area in the susceptibility map produced by the NB classifier, while for the LR classifier the two classes cover about 48%. The two models however give similar results about the less susceptible areas that are located at the north and northwest parts, while the central part of the study area shows a significant extent characterized by very high degree of susceptibility. From the visual analysis of the landslide susceptibility map produced by both classifiers, it is obvious that the spatial pattern of susceptibility follows the spatial distribution of the landslide conditioning variables, distance from road network and tectonic features. Their influence is also highlighted by the FSP analysis.

6. Conclusions

The present study evaluates the influence of models complexity and the size of training data used on the performance of a LR and a NB classifier that classifies areas in reference to their landslide susceptibility. Specifically, the performance of nine different models was calculated.

The pre-processing phase that involves the reduction of the number of feature that are used, assists in decreasing the required training samples and consequently improves the generalization performance of both classification algorithms. The reduction of the number of variables used should be an option and there evaluation should be a significant step in landslide susceptibility assessments, especially when one utilizes algorithms and methods that suffer from the curse of dimensionality. In our study the best performance was achieved in both classifiers for models that used five landslide related variables; slope angle, mean annual rainfall, distance from river network, distance from tectonic features and distance from road network, while engineering geological units and aspect were left aside as less informative. In general, the outcomes of the study showed that the behavior of high complex models with low size training data implies over fitting of data and improvement could be achieved by reducing the number of

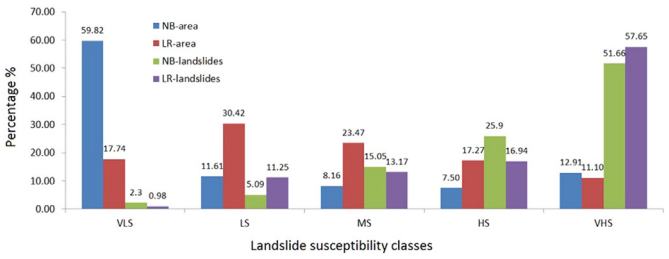

Fig. 7. Bar graphs showing the relative distribution of landslide susceptibility levels and landslide density.

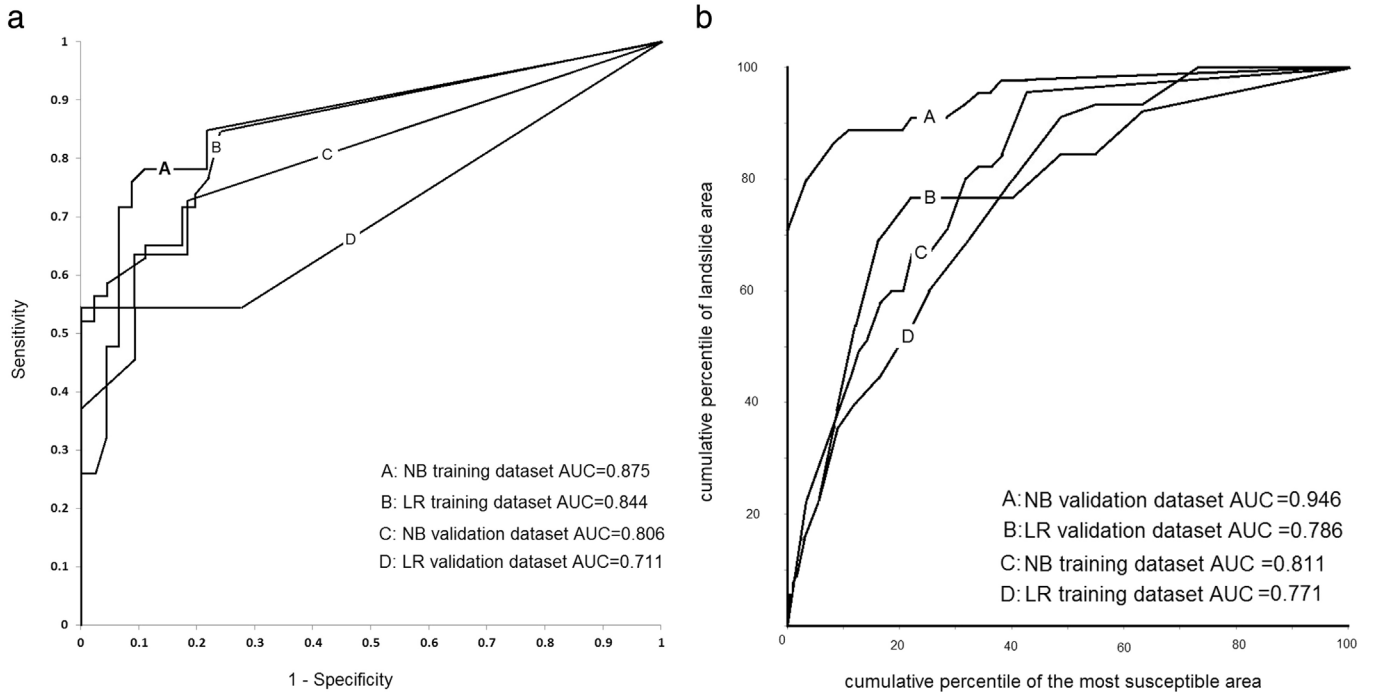

Fig. 8. a., b. ROC curves and success and predictive rate curves.

landslide-related variables that the models use. In particular, the NB classifier was estimated to have better generalization ability against LR classifier. It is safe enough to assume that when low complex models are initialized and low sized data sets, the NB classifier could be considered as a more reliable tool for constructing landslide susceptibility maps. In this context, the produced landslide susceptibility map could be regarded as a useful tool for local and national authorities and also scientific institutions in order to evaluate strategies to prevent and mitigate the impact landslides.

References

Aguado, E., Burt, J., 2012. Understanding Weather and Climate. sixth ed. Prentice Hall, Upper Saddle River, New Jersey, p. 576.

Akgun, A., 2012. A comparison of landslide susceptibility maps produced by logistic regression, multi-criteria decision, and likelihood ratio methods: a case study at İzmir, Turkey. Landslides 9 (1), 93–106.

Akgun, A., Kincal, C., Pradhan, B., 2012. Application of remote sensing data and GIS for landslide risk assessment as an environmental threat to Izmir city (west Turkey). Environ. Monit. Assess. 184, 5453–5470.

Aleotti, P., Chowdhury, R., 1999. Landslide hazard assessment: summary review and new perspectives. Bull. Eng. Geol. Environ. 58 (1), 21–44.

Alexander, D., 1992. On the causes of landslides: human activities, perception, and natural processes. Environ. Geol. Water Sci. 20 (3), 165–179.

Alimohammadlou, Y., Najafi, A., Yalcin, A., 2013. Landslide process and impacts: a proposed classification method. Catena 104, 219–232.

Alimohammadlou, Y., Najafi, A., Gokceoglu, C., 2014. Estimation of rainfall-induced landslides using ANN and fuzzy clustering methods: a case study in Saeen Slope, Azerbaijan province, Iran. Catena 120, 149–162.

Aubouin, J., 1959. Contribution a l' etude geologique de la Grece septentrionale: les confins de l' Epire et de la Thessalie. Ann. Geol. Pays Hellen. 10, 1–483.

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

Baioni, D., 2011. Human activity and damaging landslides and floods on Madeira Island. Nat. Hazards Earth Syst. Sci. 11, 3035–3046.

Ballabio, C., Sterlacchini, S., 2012. Support vector machines for landslide susceptibility mapping: the Staffora River Basin case study, Italy. Math. Geosci. 44 (1), 47–70.

Bijukchhen, S.M., Kayastha, P., Dhital, M.R., 2013. A comparative evaluation of heuristic and bivariate statistical modeling for landslide susceptibility mappings in Ghurmi-Dhad Khola, east Nepal. Arab. J. Geosci. 6 (8), 2727–2743.

Brenning, A., 2005. Spatial prediction models for landslide hazards: review, comparison and evaluation. Nat. Hazards Earth Syst. Sci. 5, 853–862.

Brunn, J.H., 1956. Contribution à l'étude géologique du Pinde septentrional et d'une partie de la Macédoine occidentale. Annales géol. pays hellén.,1re série, t. 7, 358 p., 20 pl.

Can, T., Nefeslioglu, H.A., Gokceoglu, C., Sonmez, H., Duman, T.Y., 2005. Susceptibility assessments of shallow earthlows triggered by heavy rainfall at three catchments by logistic regression analysis. Geomorphology 72 (1–4), 250–271.

Caniani, D., Pascale, S., Sdao, F., Sole, A., 2008. Neural networks and landslide susceptibility: a case study of the urban area of Potenza. Nat. Hazards 45, 55–72.

Castellanos Abella, E.A., van Westen, C.J., 2007. Generation of a landslide risk index map for Cuba using spatial multi-criteria evaluation. Landslides 4, 311–325.

Cestnik, B., Kononenko, I., Bratko, I., 1987. Assistant 86: a knowledge-elicitation tool for sophisticated users. Proceedings of the Second European Working Session on Learning. Sigma Press, Wulmslow, UK, pp. 31–45.

Chacon, J., Irigaray, C., Fernandez, T., El Hamdouni, R., 2006. Engineering geology maps: landslides and geographical information systems. Bull. Eng. Geol. Environ. 65, 341–411.

Champatiray, P.K., Dimri, S., Lakhera, R.C., Sati, S., 2007. Fuzzy-based method for landslide hazard assessment in active seismic zone of Himalaya. Landslides 4, 101–111.

Cheeseman, P., Stutz, J., 1996. Bayesian classification (autoclass): theory and results. Advances in Knowledge Discovery and Data Mining. AAAI Press, Menlo Park, CA, pp. 153–180.

Choi, J., Oh, H.J., Won, J.S., Lee, S., 2010. Validation of an artificial neural network model for landslide susceptibility mapping. Environ. Earth Sci. 60, 473–483.

Chung, C.J.F., Fabbri, A.G., 2003. Validation of spatial prediction models for landslide hazard mapping. Nat. Hazards 30 (3), 451–472.

Conforti, M., Pascale, S., Robustelli, G., Sdao, F., 2014. Evaluation of prediction capability of the artificial neural networks for mapping landslide susceptibility in the Turbolo River catchment (northern Calabria, Italy). Catena 113, 236–250.

Cooper, G.F., Herskovits, E.A., 1992. Bayesian method for the induction of probabilistic networks from data. Mach. Learn. 9 (4), 309–347.

Costanzo, D., Rotigliano, E., Irigaray, C., Jiménez-Perálvarez, J.D., Chacón, J., 2012. Factors selection in landslide susceptibility modelling on large scale following the GIS matrix method: application to the River Beiro Basin (Spain). Nat. Hazards Earth Syst. Sci. 12, 327–340.

Dai, F.C., Lee, C.F., Ngai, Y.Y., 2002. Landslide risk assessment and management: an overview. Eng. Geol. 64 (1), 65–87.

Das, I., Sahoo, S., van Westen, C., Stein, A., Hack, R., 2010. Landslide susceptibility assessment using logistic regression and its comparison with a rock mass classification system, along a road section in the northern Himalayas (India). Geomorphology 114, 627–637.

Dash, M., Liu, H., 1997. Feature selection for classification. Intell. Data Anal. 1, 679–693.

Domingos, P., Pazzani, M., 1997. Beyond independence: conditions for the optimality of the simple Bayesian classifier. Mach. Learn. 29, 103–130.

Dormann, C.F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J.R.G., Gruber, B., Lafourcade, B., Leitão, P.J., Münkemüller, T., McClean, C., Osborne, P.E., Reineking, B., Schröder, B., Skidmore, A.K., Zurell, D., Lautenbach, S., 2013. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. Ecography 36, 27–46.

Ercanoglu, M., Gokceoglu, C., 2002. Assessment of landslide susceptibility for a landslide prone area (north of Yenice, NW Turkey) by fuzzy approach. Environ. Geol. 41, 720–730.

Ercanoglu, M., Gokceoglu, C., 2004. Use of fuzzy relations to produce landslide susceptibility map of a landslide prone area West Black Sea region, Turkey. Eng. Geol. 75 (3–4), 229–250.

Ermini, L., Catani, F., Casagli, N., 2005. Artificial neural networks applied to landslide susceptibility assessment. Geomorphology 66, 327–343.

ESRI, 2013. ArcGIS Desktop: Release 10.1. Environmental Systems Research Institute, Redlands, CA.
Fawcett, T., 2006. An introduction to ROC analysis. Pattern Recogn. Lett. 27, 861–874.
Feizizadeh, B., Blaschke, T., 2013. GIS-multicriteria decision analysis for landslide susceptibility mapping: comparing three methods for the Urmia lake basin, Iran. Nat. Hazards 65 (3), 2105–2128.
Feizizadeh, B., Blaschke, T., Roodposhti, M.S., 2013. Integrating GIS based fuzzy set theory in multicriteria evaluation methods for landslide susceptibility mapping. Int. J. Geoinf. 9 (3), 49–57.
Feizizadeh, B., Roodposhti, M.S., Jankowski, P., Blaschke, T., 2014. A GIS-based extended fuzzy multi-criteria evaluation for landslide susceptibility mapping. Comput. Geosci. 73, 208–221.
Felicisimo, A.M., Cuartero, A., Remondo, J., Quiros, E., 2013. Mapping landslide susceptibility with logistic regression, multiple adaptive regression splines, classification and regression trees, and maximum entropy methods: a comparative study. Landslides 10 (2), 175–189.
Fell, R., Corominas, J., Bonnard, C., Cascini, L., Leroi, E., Savage, W., 2008. Guidelines for landslide susceptibility, hazard and risk zoning for land-use planning. Eng. Geol. 102, 99–111.
Ferentinou, M., Sakellariou, M., 2007. Computational intelligence tools for the prediction of slope performance. Comput. Geotech. 34, 362–384.
Flentje, P., Stirling, D., Chowdhury, R.N., 2007. Landslide susceptibility and hazard derived from a landslide inventory using data mining — an Australian case study. Proceedings of the First north American Landslide Conference, Landslides and Society: Integrated Science, Engineering, Management and Mitigation, pp. 1–10.
Goetz, J.N., Brenning, A., Petschko, H., Leopold, P., 2015. Evaluating machine learning and statistical prediction techniques for landslide susceptibility modeling. Comput. Geosci. 81, 1–11.
Gorsevski, P.V., Gessler, P.E., Foltz, R.B., Elliot, W.J., 2006. Spatial prediction of landslide hazard using logistic regression and ROC analysis. Trans. GIS 10 (3), 395–415.
Guns, M., Vanacker, V., 2012. Logistic regression applied to natural hazards: rare event logistic regression with replications. Nat. Hazards Earth Syst. Sci. 12, 1937–1947.
Guyon, I., Elisseeff, A., 2003. An introduction to variable and feature selection. J. Mach. Learn. Res. 3, 1157–1182.
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, 181–216.
Guzzetti, F., Reichenbach, P., Cardinali, M., Galli, M., Ardizzone, F., 2005. Probabilistic landslide hazard assessment at the basin scale. Geomorphology 72, 272–299.
Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, H.I., 2009. The WEKA Data Mining Software: An Update; SIGKDD Explorations, Volume 11, Issue 1.
Halvorson, M., 2010. Microsoft Visual Basic 2010 Step by Step. Pearson Education, p. 608.
Hanley, J.A., McNeil, B.J., 1982. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 143 (1), 29–36.
Heckmann, T., Gegg, K., Gegg, A., Becht, M., 2014. Sample size matters: investigating the effect of sample size on a logistic regression susceptibility model for debris flows. Nat. Hazards Earth Syst. Sci. 14, 259–278.
Hong, H., 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.
Hutchinson, J.N., 1995. Keynote paper: landslide hazard assessment. Proceedings 6th International Symposium on Landslides, Christchurch. Balkema, Rotterdam, pp. 1805–1841.
Ilia, I., Tsangaratos, P., 2015. Applying weight of evidence method and sensitivity analysis to produce a landslide susceptibility map. Landslides http://dx.doi.org/10.1007/s10346-015-0576-3.
Irigaray, C., Fernández, T., El Hamdouni, R., Chacón, J., 2007. Evaluation and validation of landslide-susceptibility maps obtained by a GIS matrix method: examples from the Betic Cordillera (southern Spain). Nat. Hazards 41, 61–79.
Jain, A.K., Duin, R.P.W., Mao, J., 2000. Statistical pattern recognition: a review. IEEE Trans. Pattern Anal. Mach. Intell. 22, 4–37.
John, G., Langley, P., 1995. Estimating continuous distributions in Bayesian classifiers. Proceeding of the Eleventh Conference on Uncertainty in Artificial Intelligence, pp. 338–345.
Jones, P.D., Harris, I., 2008. Climatic Research Unit (CRU) time-series datasets of variations in climate with variations in other phenomena. University of East Anglia Climatic Research Unit, NCAS British Atmospheric Data Centre.
Kayastha, P., Dhital, M.R., De Smedt, F., 2012. Landslide susceptibility mapping using the weight of evidence method in the Tinau watershed, Nepal. Nat. Hazards 63 (2), 479–498.
Korup, O., Stolle, A., 2014. Landslide prediction from machine learning. Geol. Today 30 (1), 26–33.
Koukis, G., Sabatakakis, N., Nikolaou, N., Loupasakis, C., 2005. Landslides Hazard Zonation in Greece. Proc of Open Symp. on Landslides Risk Analysis and Sustainable Disaster Management by International Consortium on Landslides, Washington USA, Chapter 37, pp. 291–296.
Kouli, M., Loupasakis, C., Soupios, P., Rozos, D., Vallianatos, F., 2014. Landslide susceptibility mapping by comparing the WLC and WofE multi-criteria methods in the West Crete Island, Greece. Environ. Earth Sci. 72 (12), 5197–5219.
Lee, S., 2004. Application of likelihood ratio and logistic regression models to landslide susceptibility mapping using GIS. Environ. Manag. 34, 223–232.
Lee, S., Pradhan, B., 2007. Landslide hazard mapping at Selangor, Malaysia using frequency ratio and logistic regression models. Landslides 4 (1), 33–41.
Lee, S., Sambath, T., 2006. Landslide susceptibility mapping in the Damrei Romel area, Cambodia using frequency ratio and logistic regression models. Environ. Geol. 50 (6), 847–855.
Lee, S., Ryu, J.H., Min, K., Won, J.S., 2003. Landslide susceptibility analysis using GIS and artificial neural network. Earth Surf. Process. Landf. 28 (12), 1361–1376.
Lee, S., Ryu, J., Won, J., Park, H., 2004. Determination and application of the weights for landslide susceptibility mapping using an artificial neural network. Eng. Geol. 71 (3–4), 289–302.
Mancini, F., Ceppi, C., Ritrovato, G., 2010. GIS and statistical analysis for landslide susceptibility mapping in the Daunia area, Italy. Nat. Hazards Earth Syst. Sci. 10, 1851–1864.
Marjanovic, M., Kovacevic, M., Bajat, B., Vozenilek, V., 2011. Landslide susceptibility assessment using SVM machine learning algorithm. Eng. Geol. 123, 225–234.
Marquardt, D., 1970. Generalized inverses, ridge regression, biased linear estimation, and non – linear estimation. Technometrics 12, 605–607.
Melchiorre, C., Matteucci, M., Azzoni, A., Zanchi, A., 2008. Artificial neural networks and cluster analysis in landslide susceptibility zonation. Geomorphology 94 (3–4), 379–400.
Miner, A.S., Vamplew, P., Windle, D.J., Flentje, P., Warner, P., 2010. A comparative study of various data mining techniques as applied to the modeling of landslide susceptibility on the Bellarine Peninsula, Victoria, Australia. In: Williams, A.L., Pinches, G.M., Chin, C.Y., McMorran, T.J. (Eds.), Geologically Active. CRC Press, New York, NY, USA, p. 352.
Montgomery, D.C., Peck, E.A., Vining, G.G., 2001. Introduction to Linear Regression Analysis. third ed. Wiley, New York, NY.
Murakami, Y., Mizuguchi, K., 2010. Applying the Naïve Bayes classifier with kernel density estimation to the prediction of protein-protein interaction sites. Bioinformatics 26 (15), 1841–1848.
Muthu, K., Petrou, M., Tarantino, C., Blonda, P., 2008. Landslide possibility mapping using fuzzy approaches. IEEE Trans. Geosci. Remote Sens. 46, 1253–1265.
Neaupane, K.M., Achet, S.H., 2004. Use of back propagation neural network for landslide monitoring: a case study in the higher Himalaya. Eng. Geol. 74 (3–4), 213–226.
Nefeslioglu, H.A., 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, 171–191.
Nefeslioglu, H.A., Sezer, E., Gokceoglu, C., Bozkir, A.S., Duman, T.Y., 2010. Assessment of landslide susceptibility by decision trees in the metropolitan area of Istanbul, Turkey. Math. Probl. Eng. http://dx.doi.org/10.1155/2010/901095 (Article ID 901095).
Negnevitsky, M., 2002. Artificial Intelligence: A Guide to Intelligent Systems. Addison–Wesley/Pearson Education, Harlow, England, p. 394.
Ng, A.Y., Jordan, M.I., 2001. On discriminative vs. generative classifiers: a comparison of logistic regression and naive Bayes. In: Dietterich, T.G., Becker, S., Ghahramani, Z. (Eds.), NIPS. MIT Press, MA, pp. 841–848.
O'brien, R.M., 2007. A caution regarding rules of thumb for variance inflation factors. Qual. Quant. 41, 673–690.
Oh, H.J., Lee, S., 2010. Cross-validation of logistic regression model for landslide susceptibility mapping at Ganeoung areas, Korea. Disaster Adv. 3 (2), 44–55.
Oh, H.J., Lee, S., 2011. Landslide susceptibility mapping on Panaon Island, Philippines using a geographic information system. Environ. Earth Sci. 62, 935–951.
Oh, H.J., Pradhan, B., 2011. Application of a neuro-fuzzy model to landslide susceptibility mapping in a tropical hilly area. Comput. Geosci. 37 (3), 1264–1276.
Poudyal, C.P., Chang, C., Oh, H.J., Lee, S., 2010. Landslide susceptibility maps comparing frequency ratio and artificial neural networks: a case study from the Nepal Himalaya. Environ. Earth Sci. 61 (5), 1049–1064.
Pourghasemi, H.R., Pradhan, B., Gokceoglu, C., 2012a. Application of fuzzy logic and analytical hierarchy process (AHP) to landslide susceptibility mapping at Haraz watershed, Iran. Nat. Hazards 63 (2), 965–996.
Pourghasemi, H.R., Mohammady, M., Pradhan, B., 2012b. Landslide susceptibility mapping using index of entropy and conditional probability models in GIS: Safarood Basin, Iran. Catena 97, 71–84.
Pourghasemi, H.R., Moradi, H.R., Fatemi Aghda, S.M., 2013a. 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.
Pourghasemi, H.R., Jirandeh, A.G., Pradhan, B., Xu, C., Gokceoglu, C., 2013b. Landslide susceptibility mapping using support vector machine and GIS at the Golestan Province, Iran. J. Earth Syst. Sci. 122 (2), 349–369.
Pradhan, B., 2010. Application of an advanced fuzzy logic model for landslide susceptibility analysis. Int. J. Comput. Intell. Syst. 3 (3), 370–381.
Pradhan, B., 2011a. Use of GIS-based fuzzy logic relations and its cross application to produce landslide susceptibility maps in three test areas in Malaysia. Environ. Earth Sci. 63 (2), 329–349.
Pradhan, B., 2011b. Manifestation of an advanced fuzzy logic model coupled with geoinformation techniques for landslide susceptibility analysis. Environ. Ecol. Stat. 18 (3), 471–493.
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.
Pradhan, B., Lee, S., 2010a. Delineation of landslide hazard areas on Penang Island, Malaysia, by using frequency ratio, logistic regression, and artificial neural network models. Environ. Earth Sci. 60, 1037–1054.
Pradhan, B., Lee, S., 2010b. Landslide susceptibility assessment and factor effect analysis: back-propagation artificial neural networks and their comparison with frequency ratio and bivariate logistic regression modelling. Environ. Model. Softw. 25, 747–759.
Pradhan, B., Lee, S., 2010c. Regional landslide susceptibility analysis using back – propagation neural network model at Cameron highland, Malaysia. Landslides 7 (1), 13–30.
Pradhan, B., Lee, S., Buchroithner, M.B., 2009. Use of geospatial data for the development of fuzzy algebraic operators to landslide hazard mapping: a case study in Malaysia. Appl. Geomatics 1 (1), 3–15.
Pradhan, B., Lee, S., Buchroithner, M.F., 2010a. A GIS-based back-propagation neural network model and its cross application and validation for landslide susceptibility analyses. Comput. Environ. Urban. Syst. 34, 216–235.

Pradhan, B., Sezer, E.A., Gokceoglu, C., Buchroithner, M.F., 2010b. Landslide susceptibility mapping by neuro-fuzzy approach in a landslide-prone area (Cameron Highlands, Malaysia). IEEE Trans. Geosci. Remote Sens. 48 (12), 4164–4177.
Regmi, N.R., Giardino, J.R., Vitek, J.D., 2010a. Modeling susceptibility to landslides using the weight of evidence approach: Western Colorado, USA. Geomorphology 115, 172–187.
Regmi, N.R., Giardino, J.R., Vitek, J.D., 2010b. Assessing susceptibility to landslides: using models to understand observed changes in slopes. Geomorphology 122, 25–38.
Regmi, N.R., Giardino, J.R., McDonald, E.V., Vitek, J.D., 2014. A comparison of logistic regression-based models of susceptibility to landslides in western Colorado, USA. Landslides 11, 247–262.
Rozos, D., Pyrgiotis, L., Skias, S., Tsagaratos, P., 2008. An implementation of rock engineering system for ranking the instability potential of natural slopes in Greek territory: an application in Karditsa County. Landslides 5 (3), 261–270.
Sabatakakis, N., Koukis, G., Vassiliades, E., Lainas, S., 2013. Landslide susceptibility zonation in Greece. Nat. Hazards 65 (1), 523–543.
Saito, H., Nakayama, D., Matsuyama, H., 2009. Comparison of landslide susceptibility based on a decision-tree model and actual landslide occurrence: the Akaishi mountains, Japan. Geomorphology 109 (3–4), 108–121.
Sdao, F., Lioi, D.S., Pascale, S., Caniani, D., Mancini, I.M., 2013. Landslide susceptibility assessment by using a neuro-fuzzy model: a case study in the rupestrian heritage rich area of Matera. Nat. Hazards Earth Syst. Sci. 13, 395–407.
Sezer, A.E., Pradhan, B., Gokceoglu, C., 2011. Manifestation of an adaptive neuro-fuzzy model on landslide susceptibility mapping: Klang valley, Malaysia. Expert Syst. Appl. 38 (7), 8208–8219.
Shahabi, H., Khezri, S., Bin Ahmad, B., Hashim, M., 2014. Landslide susceptibility mapping at central Zab basin, Iran: a comparison between analytical hierarchy process, frequency ratio and logistic regression models. Catena 115, 55–70.
Soeters, R.S., van Westen, C.J., 1996. Slope instability recognition, analysis, and zonation. In: Turner, A.K., Schuster, R.L. (Eds.), Landslides: Investigation and Mitigation. Special Report vol. 247.
Soria, D., Garibaldi, J.M., Ambrogi, F., Biganzoli, E.M., Ellis, I.O., 2011. A "non-parametric" version of the naive Bayes classifier. Knowl.-Based Syst. 24 (6), 775–784.
SPSS, Inc. Released, 2007. SPSS for Windows, Version 16.0. SPSS Inc., Chicago.
Suzen, M.L., Kaya, B.S., 2011. Evaluation of environmental parameters in logistic regression models for landslide susceptibility mapping. Int. J. Digit. Earth 5, 338–355.
Tien Bui, D., Pradhan, B., Lofman, O., Revhaug, I., Dick, O.B., 2012a. Landslide susceptibility assessment in the Hoa Binh province of Vietnam using artificial neural network. Geomorphology 171-172, 12–19.
Tien Bui, D., Pradhan, B., Lofman, O., Revhaug, I., Dick, O.B., 2012b. 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.
Tien Bui, D., Pradhan, B., Lofman, O., Revhaug, I., 2012c. Landslide Susceptibility Assessment in Vietnam Using Support Vector Machines, Decision Tree, and Naïve Bayes Models. Mathematical Problems in Engineering vol. 2012. http://dx.doi.org/10.1155/2012/974638 (Article ID 974638, 26 pages, 2012).
Tien Bui, D., Anh Tuan, T., Klempe, H., Pradhan, B., Revhaug, I., 2015a. 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 http://dx.doi.org/10.1007/s10346-015-0557-6.
Tien Bui, D., Tuan, T., Klempe, H., Pradhan, B., Revhaug, I., 2015b. 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 http://dx.doi.org/10.1007/s10346-015-0557-6.
Tsangaratos, P., Benardos, A., 2014. Estimating landslide susceptibility through an artificial neural network classifier. Nat. Hazards 74 (3), 1489–1516.
Tsangaratos, P., Ilia, I., 2015. Landslide susceptibility mapping using a modified decision tree classifier in the Xanthi Prefecture, Greece. Landslides http://dx.doi.org/10.1007/s10346-015-0565-6.
Tsangaratos, P., Ilia, I., Rozos, D., 2013. Case event system for landslide susceptibility analysis. In: Margottini, Canuti, Sassa (Eds.), Landslide Science and Practice. Springer, Berlin Heidelberg, pp. 585–593.
Vahidnia, M.H., Alesheikh, A.A., Alimohammadi, A., Hosseinali, F., 2010. A GIS-based neuro-fuzzy procedure for integrating knowledgeand data in landslide susceptibility mapping. Comput. Geosci. 36 (29), 1101–1114.
Van Den Eeckhaut, M., Vanwalleghem, T., Poesen, J., Govers, G., Verstraeten, G., Vandekerckhove, L., 2006. Prediction of landslide susceptibility using rare events logistic regression: a case-study in the Flemish Ardennes (Belgium). Geomorphology 76, 392–410.
Van Den Eeckhaut, M., Marre, A., Poesen, J., 2010. Comparison of two landslide susceptibility assessments in the Champagne–Ardenne region (France). Geomorphology 115, 141–155.
Van Westen, J., Van Asch, J., Soeters, R., 2006. Landslide hazard and risk zonation — why is still so difficult? Bull. Eng. Geol. Environ. 65, 167–184.
Varnes, D.J., 1984. International Association of Engineering Geology Commission on Landslides and Other Mass Movements on Slopes: Landslide Hazard Zonation: A Review of Principles and Practice. UNESCO, Paris (63 pp.).
Wang, L.J., Sawada, K., Moriguchi, S., 2013. Landslide susceptibility analysis with logistic regression model based on FCM sampling strategy. Comput. Geosci. 57, 81–92.
Weisberg, S., Fox, J., 2010. An R Companion to Applied Regression. Sage Publications, Incorporated, Los Angeles, London, New Delhi, Singapore, Washington, D.C.
Xu, C., Dai, F., Xu, X., Lee, Y.H., 2012. GIS-based support vector machine modeling of earthquake-triggered landslide susceptibility in the Jianjiang River watershed, China. Geomorphology 145-146, 70–80.
Yalcin, A., Reis, S., Aydinoglu, A.C., 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, 274–287.
Yao, X., Tham, L.G., Dai, F.C., 2008. Landslide susceptibility mapping based on support vector machine: a case study on natural slopes of Hong Kong, China. Geomorphology 101, 572–582.
Yeon, Y.K., Han, J.G., Ryu, K.H., 2010. Landslide susceptibility mapping in Injae, Korea, using a decision tree. Eng. Geol. 16 (3–4), 274–283.
Yesilnacar, 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 (3–4), 251–266.
Yilmaz, I., 2010. Comparison of landslide susceptibility mapping methodologies for Koyulhisar, Turkey: conditional probability, logistic regression, artificial neural networks, and support vector machine. Environ. Earth Sci. 61, 821–836.
Zare, M., Pourghasemi, H., 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 (8), 2873–2888.
Zhu, A.-X., Wang, R., Qiao, J., Qin, C.-Z., Chen, Y., Liu, J., Du, F., Lin, Y., Zhu, T., 2014. An expert knowledge-based approach to landslide susceptibility mapping using GIS and fuzzy logic. Geomorphology 214, 128–138.