Environ Earth Sci (2017) 76:652
DOI 10.1007/s12665-017-6981-2

ORIGINAL ARTICLE

A novel hybrid integration model using support vector machines and random subspace for weather-triggered landslide susceptibility assessment in the Wuning area (China)

Haoyuan Hong[1,2,3] · Junzhi Liu[1,2,3] · A-Xing Zhu[1,2,3] · Himan Shahabi[4] · Binh Thai Pham[5] · Wei Chen[6] · Biswajeet Pradhan[7,8] · Dieu Tien Bui[9]

Received: 20 March 2017 / Accepted: 16 September 2017 / Published online: 6 October 2017

Abstract This study proposed a hybrid modeling approach using two methods, support vector machines and random subspace, to create a novel model named random subspace-based support vector machines (RSSVM) for assessing landslide susceptibility. The newly developed model was then tested in the Wuning area, China, to produce a landslide susceptibility map. With the purpose of achieving the objective of the study, a spatial dataset was initially constructed that includes a landslide inventory map consisting of 445 landslide regions. Then, various landslide-influencing factors were defined, including slope angle, aspect, altitude, topographic wetness index, stream power index, sediment transport index, soil, lithology, normalized difference vegetation index, land use, rainfall, distance to roads, distance to rivers, and distance to faults. Next, the result of the RSSVM model was validated using statistical index-based evaluations and the receiver operating characteristic curve approach. Then, to evaluate the performance of the suggested RSSVM model, a comparison analysis was performed to other existing approaches such as artificial neural network, Naïve Bayes (NB) and support vector machine (SVM). In general, the performance of the RSSVM model was better than the other models for spatial prediction of landslide susceptibility. The AUC results of the applied models are as follows: RSSVM (AUC = 0.857), followed by MLP (AUC = 0.823), SVM (AUC = 0.814) and NB (AUC = 0.783). The present study indicates that RSSVM can be used for landslide susceptibility evaluation, and the results are very useful for local governments and people living in the Wuning area.

Keywords Landslides · GIS · Support vector machines · Random subspace

✉ A-Xing Zhu
azhu@wisc.edu

✉ Wei Chen
chenwei.0930@163.com; chenwei0930@xust.edu.cn

1 Key Laboratory of Virtual Geographic Environment (Nanjing Normal University), Ministry of Education, Nanjing 210023, China

2 State Key Laboratory Cultivation Base of Geographical Environment Evolution (Jiangsu Province), Nanjing 210023, China

3 Jiangsu Center for Collaborative Innovation in Geographic Information Resource Development and Application, Nanjing 210023, Jiangsu, China

4 Department of Geomorphology, Faculty of Natural Resources, University of Kurdistan, Sanandaj, Iran

5 Department of Geotechnical Engineering, University of Transport Technology, 54 Trieu Khuc, Thanh Xuan, Hanoi, Vietnam

6 College of Geology and Environment, Xi'an University of Science and Technology, Xi'an 710054, China

7 Department of Civil Engineering, Geospatial Information Science Research Center (GISRC), Faculty of Engineering, University Putra Malaysia, 43400 Serdang, Selangor Darul Ehsan, Malaysia

8 School of Systems, Management and Leadership, Faculty of Engineering and Information Technology, University of Technology Sydney, Building 11, Level 06, 81 Broadway, PO Box 123, Ultimo, NSW 2007, Australia

9 Geographic Information System group, Department of Business and IT, University College of Southeast Norway, N-3800 Bøi Telemark, Norway

Introduction

Landslides are considered one of the most extensively distributed mass developments in hilly terrain all over the world (Blothe et al. 2015; Carey et al. 2015; Nicolussi et al. 2015; Paulin et al. 2015; Pham et al. 2016a). Due to their unexpected and seasonal characteristics (Vilimek and Smolikova 2015; Yang et al. 2015), landslides always present great risks to human life and economic stability (Kritikos and Davies 2015; Panek 2015), especially in industrial and other social activities such as mines, water resources facilities and hydropower stations (Vranken et al. 2015; Zhang et al. 2015). Therefore, over the last two decades, many scientists have engaged themselves to predict landslide locations spatially, which is critical for development and planning purposes (Peng et al. 2015; Posner and Georgakakos 2015). In the formation of landslides, many factors play key roles; some are natural factors, and some are human factors. It has also been observed that climate change and landslide activities have a significant relationship (Sewell et al. 2015). Some recent studies have revealed that turbidity currents and submarine landslides will be the main cause of future rapid global climate warming. However, some research has indicated that accelerated climate change does not undoubtedly increase activity and thus does not provide a direct confirmation of climate change as a dominant triggering factor (Clare et al. 2015). In fact, heavy rainfall and snow melt have been supported as great indirect triggering factors due to complex hydrogeological effects and isolated groundwater organization. As such, it is difficult to determine the triggering parameter, especially when it is particularly affiliated with oblique river erosion in a landslide appendage (Abolmasov et al. 2015; Kirschbaum et al. 2015).

Preparation of a high-quality landslide inventory map is very critical in landslide studies. Field surveys and remote sensing-based analysis are commonly used, but both of these approaches have merits and demerits. Field investigation is highly time-consuming and very laborious. On the other hand, satellite image processing techniques also need rigorous validation steps to produce an authentic landslide inventory map (Hong et al. 2016; Lin et al. 2015). High-resolution imagery is an essential information base for properly evaluating and certifying landslide features (Akcay 2015; Yusof et al. 2015). Similarly, selecting landslide-conditioning factors is also an important and complex process. In this regard, the statistical index method is generally used for the validation of different landslide-conditioning factor sets. The method expresses how forceful every single landslide-conditioning factor enhances or diminishes the objective task and then excludes a few factors to acquire superior inputs (Bellugi et al. 2015; Meinhardt et al. 2015).

In recent years, extensive application of Geographical Information System (GIS) and Remote Sensing (RS) technologies has been used to assess landslide susceptibility, hazards and risk extent (Ciabatta et al. 2015; Oliveira et al. 2015; Pham et al. 2016c; Shahabi and Hashim 2015; Tan et al. 2015). There are numerous state-of-the-art approaches and methods that have been adopted in landslide susceptibility mapping. For example, the step-wise weight assessment ratio (Dehnavi et al. 2015), multivariate adaptive regression spline (MARS) (Chen et al. 2017d; Conoscenti et al. 2015; Wang et al. 2015b), artificial neural network (ANN) (Chen et al. 2017e; Dou et al. 2015; Bui et al. 2016), random forest (RF) (Chen et al. 2017g; Trigila et al. 2015), multicriteria evaluation (Ahmed 2015), kernel logistic regression (Chen et al. 2017f; Hong et al. 2015; Bui et al. 2016), spatial multicriteria evaluation (SMCE) (Gaprindashvili and Van Westen 2016), and adaptive neuro-fuzzy inference system (ANFIS) (Chen et al. 2017a; Nasiri Aghdam et al. 2016; Bui et al. 2012) have all been used. However, there still remains a question of the selection of the best method in landslide susceptibility assessment (Boue et al. 2015; Goetz et al. 2015; Pham et al. 2015). Thus, in order to achieve the most suitable and optimum model, many scientists have used various models in geographically different study areas worldwide (Chen et al. 2017g; Bui et al. 2017; Tsangaratos et al. 2017; Youssef et al. 2016).

The principal aim of the present study is to analyze probable utilization of the novel hybrid integration method of support vector machines (SVM) and random subspace (RS), named RSSVM, for landslide susceptibility assessment. Random subspace is an integrated algorithm, whereas RSSVM is a SVM-based classifier. Using statistical index-based assessment and receiver operating characteristic curve (ROC) means, the accuracy of the RSSVM model has been assessed. In addition, other approaches, such as Artificial Neural Network (ANN), Naïve Bayes (NB) and Support Vector Machine (SVM), have been applied, and the results were compared to the results of the RSSVM model. All the aforementioned models were applied in the Wuning area situated in the Jiangxi province (China). The analysis process was carried out using Weka 3.7 and ArcMap 10.0.

Model background

Support vector machines

A support vector machine (SVM) is defined as a supervised learning method and is widely applied in the statistical categorization and regression test (Chen et al. 2017b; Pham et al. 2016b). It leads to a map vector in a higher dimension space, within which there is a maximum separation hyperplane (Feng et al. 2016; Promper and Glade 2016). On the opposite side of the hyperplane, SVM separates data from two parallel hyperplanes from each other (Chen et al. 2017c). In addition, SVM separates a super flat surface to maximize the distance of two parallel hyperplanes (Bezak et al. 2016; Iadanza et al. 2016). Assuming that the distance between parallel gaps is greater, the classifier of the total error is smaller. The main idea of SVM can be summarized as follows (Li et al. 2016; Ma et al. 2016):

1. It is used to solve the problem of linear separable analysis, and in the case of a linear integral, nonlinear mapping algorithms are generally used, which makes a low spatial input space linear case into a high spatial feature space linear separable case (Alvioli and Baum 2016; Bennett et al. 2016; Mertens et al. 2016). Therefore, it makes possible a high spatial feature space by using a linear algorithm on nonlinear characteristics.
2. It is based on a structural opportunity minimization theory to create an excellent aggressive plane segmentation in the feature space (LaHusen et al. 2016; Osadchiev et al. 2016; Yamao et al. 2016).
3. More detailed information about the SVM algorithm can be seen in Chen et al. (2016a, 2016b, 2017h) and Pradhan (2013). The equation of SVM algorithm can be described as follows:

$$y_i(w \cdot x_i + b)/ \geq 1 - \xi_i \tag{1}$$

Random subspace ensemble

Random subspace is defined as a classic integrated algorithm which was proposed by Ho in 1998 (Tin Kam 1998). The algorithm is similar to the bagging algorithm and is randomly selected by the original training set to construct the training subset (Kotsiantis 2011; Kuncheva et al. 2010; Mielniczuk and Teisseyre 2014). However, the difference is that random subspace is randomly chosen from the original training set of features (Bertoni et al. 2005; Skurichina and Duin 2001, 2002). Then, the series features a subset of each subclassifier training at the final forecast results obtained by a combination of voting methods (Lai et al. 2006; Sun and Zhang 2007; Tao et al. 2006). The performance of the results depends on integrated learning differences, the bagging method subcategories to obtain the difference between the subclassification among different samples of each subclassifier training, and the random subspace ensemble learning method to take samples at different spatial characteristics to obtain differences between subclassification (Nanni and Lumini 2008; Zhang and Jia 2007; Zhu et al. 2009).

Novel hybrid integration of support vector machines (SVM) and random subspace (RS) (RSSVM)

The novel hybrid integration of the support vector machines (SVM) approach and the random subspace (RS), ensemble (RSSVM), for spatial forecasting of landslides appearing in the Wuning area, is displayed in Fig. 1.

1. *Data collection and processing in GIS* The landslide database is created for building and certifying the proposed integrated landslide susceptibility model. The landslide inventory map with 445 landslides was randomly split into 70% (311 landslides shown in the yellow color) for coaching the models and the remaining 30% (134 landslides shown in the red color) for validation purposes (Fig. 2). The coaching and validating datasets were applied to build the landslide susceptibility model, whereas the testing dataset was employed to check the RSSVM model. For this purpose, by employing the frequency ratio method, fourteen landslide-influencing factors were reclassified from categorical classes into continuous values. (Bui et al. 2015). Subsequently, all 14 landslide-conditioning factor maps were transformed to a raster format with a pixel of 25 m.
2. *Random Subspace ensemble (RSS) optimization* Random subspace (RSS) can improve the performance of the SVM classifier by dividing a dataset with a large dimensional feature space to lower datasets, such as in training data. Generally, these ensemble data mining classifiers have better accuracy than a single predictor. These methods first produce via training the dataset base classifications. Second, the real classification is achieved by combining the results of base classifiers with the previous base classifier (Piao et al. 2015). The process (1) divides the original feature space (FR) into L feature subsets (FS) of p-dimensionality, (2) submits each subset to a base classifier (BC) in the ensemble, and (3) makes a final decision on the class of the form achieved by connecting the decisions of these BCs using a connecting order such as “most votes” (Kuncheva and Plumpton 2010). Because the feature subsets submitted to the BCs are diverse, an aligned ensemble is a common selection (Kuncheva and Plumpton 2010). By evaluating the fitness of each result, the algorithm constantly searches the solution space and pursues the most appropriate set of the model criteria.

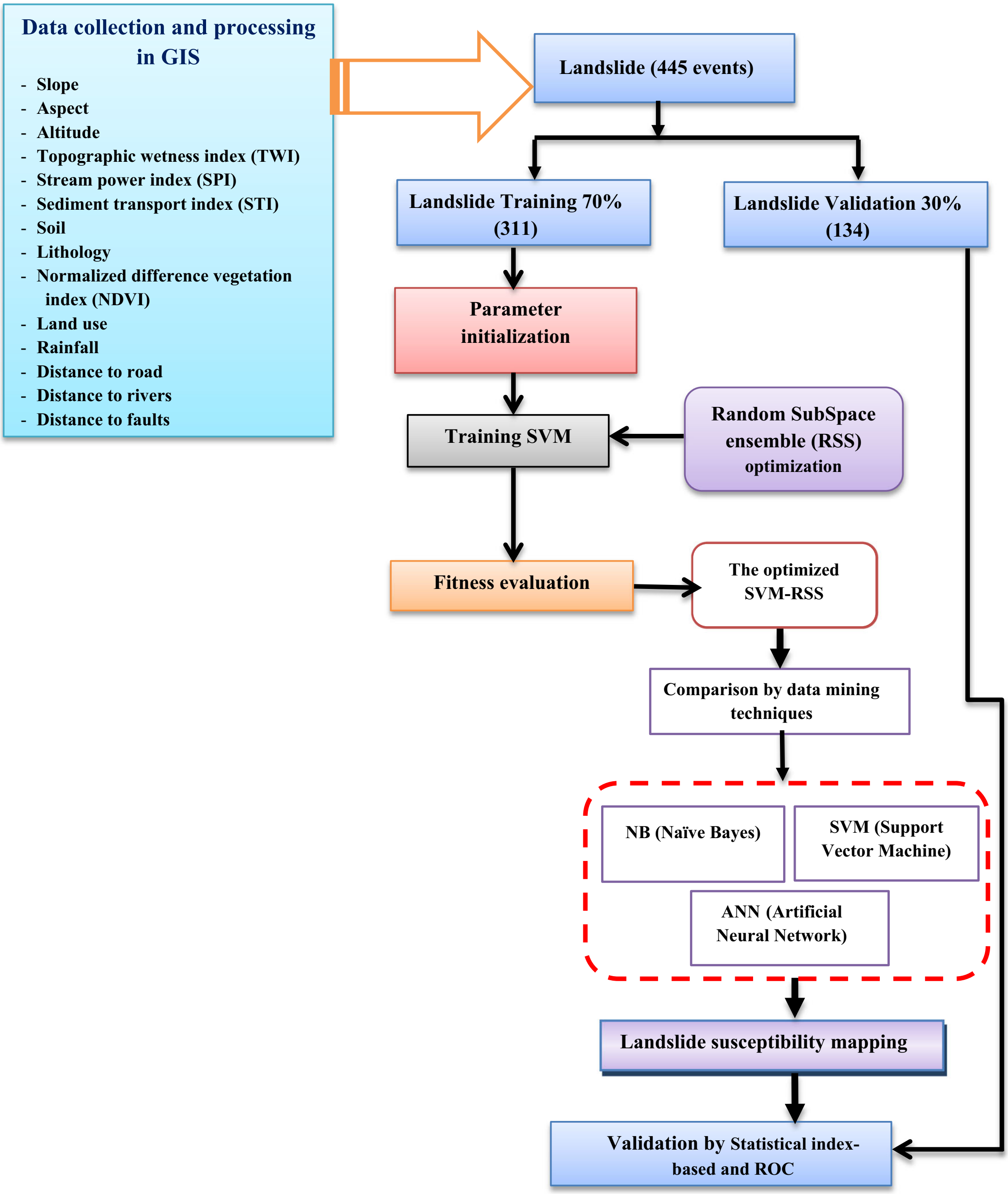


Fig. 1 The flowchart of the proposed RSSVM model

3. *Training support vector machines model (SVM)* As we know, the SVM algorithm is a supervised learning method; a training dataset with input components and matching desired class labels must be supplied. Using the RBF kernel function and based on the training dataset, the SVM algorithm draws the input data from the original input space into a high-dimensional feature space (Chen et al. 2016a). Accordingly, the result of this algorithm builds up a hyperplane. SVM can separate the input data of influencing factors into two distinctive decision areas: "landslide" ($Y = 1$) and "non-landslide" ($Y = -1$). When the training case is finished, the SVM algorithm for classifying input patterns is displayed.
4. *The optimized RSSVM model* The SVM-based boost method continues until the maximum number of generations is obtained (Piao et al. 2015). When the exploring process completes, an appropriate set of the RSSVM adapt parameters is established. Then, it is used to compose a classification model for spatial

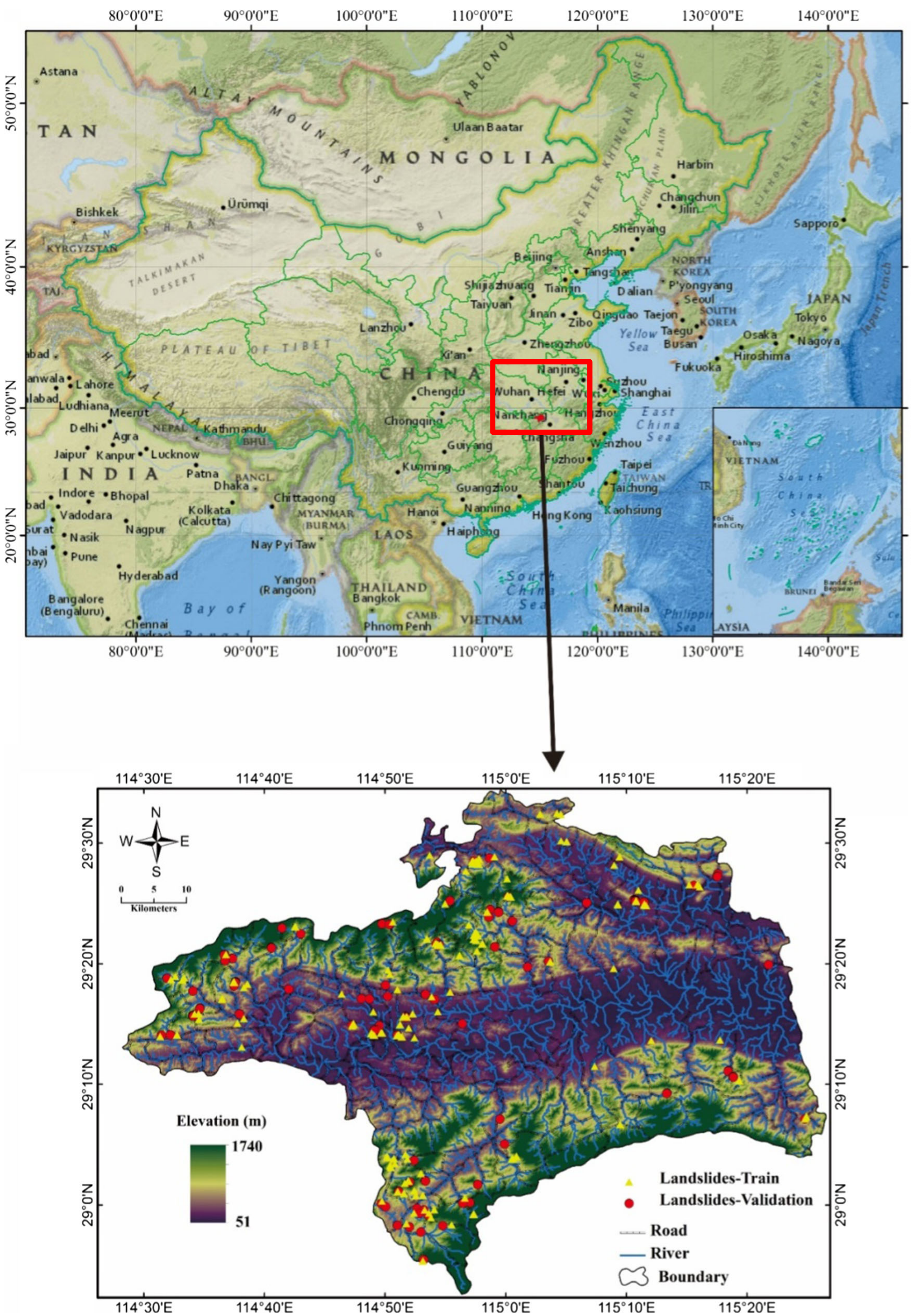


Fig. 2 Landslide location map of the Wuning area [Map of China from the National Geographic World Map (ESRI 2010)]

Table 1 Types of geological formations in the study area

No.	Unit name	Lithology
A	Hua yansi group; Xi yanshan group; Yang liugan group; Guan yintang group	Argillaceous limestone; nodular limestone; gray-black shale; stone seam; siltstone; limestone; carbonate rock
B	Shuang shan group; Lao hudong group	Pure quartz sandstone; hematite; limestone; dolomite
C	Wuning group; Mou xia group; Zheng jiazu group	Conglomerate; pebbly sandstone; siltstone; limestone; sandy shale
	Shuang qiaoshan group; Xiu shui group; An lelin group	Tuffaceous slate; phyllite; spilite; quartz keratophyre
D	Jiu xiantang group; Mu bushan group; Xia changshan group; Huang xie group	(Two long; K-feldspar) granite
	Hai hui group; Xi huashan group	Two-mica adamellite
	Nan tuo group	Adamellite
	Liu ken unit	Porphyry
E	Lian cang group; Dou shantuo group	Gray conglomerate; sandstone; conglomerate bottom; moraine mud conglomerate
	Lan tian group; Pi yuancun group; Deng cai group	Clay gravel moraine; shale; limestone and manganese; dark gray chert; siliceous limestone; dolomitic limestone
F	Hu le group; Xin kailing group; Tan shan group; Tan tou group; Hong hua group	Dolomite; limestone; nodular limestone; mudstone; shale silicon
G	Long tan group; Wu jiacheng group; Da long group; Liang shan group; Xi xia group; Xiao jiabian group	Limestone; dark gray siltstone; clay shale; coal seams; flint limestone; chert siliceous shale
	Mao kou group; Yun taishuang group; Huang long group	Sandstone; dolomite; limestone; coal-bearing shale; limestone flint; chert
H	Luo kedong group	Conglomerate; volcanic breccia; tuff; slate; lava
	Jiu lin unit; Kuai bu unit; Shi huajian unit	Tonalite-diorite; granodiorite; monzonite granite
	Hu qiaoli unit	Monzonite granite; adamellite
	Tuo lin unit	Sandstone
I	Fengtou group; Mao shan group; Xin tang group; Dian bei group; Li shuwo group	Sandy mudstone; argillaceous siltstone and fine sandstone; sandstone
J	Zhou chongcun group; Qing long group	Micrite; silty shale; calcareous mudstone; marl
	Lan tiancun group; Pi yuancunoup	Dark gray silty shale; shale; limestone and manganese; dark gray chert; siliceous limestone; dolomitic limestone

A, B, C, D, E, F, G, H, I, J represent the class of lithology

evaluation of landslide susceptibility in the Wuning area. The suggested model is prepared to forecast invisible input arrangements in the validation dataset.

5. *Comparison with data mining techniques* The obtained results from the novel hybrid integration of support vector machines and random subspace ensemble (RSSVM) were compared to other well-known data mining techniques such as Multiple Perceptron Neural Network (MLPN), Naïve Bayes (NB) and Support Vector Machine (SVM) for spatial assessment of landslide susceptibility in the Wuning area.

Experiment and analysis

Description of study region

The Wuning area is almost 3507 km^2 and overlays the entire area of the Wuning region. The Wuning area is situated in the western part of Jiangxi province. It lies in the South hilly area of Mubu between the longitudes of 114°29′E and 115°27′E and the latitudes of 28°53′N and 29°35′N (Fig. 2). According to a report of geological survey results (http://www.cgs.gov.cn/), there are more than 56 geologic groups and units recognized (Table 1). In the Wuning area, the main lithologies are Tuffaceous slate, phyllite, spilite, quartz keratophyre, (two long; K-feldspar) granite, and two-mica adamellite (Fig. 3). Geologic data for the Wuning area were acquired from the China Geology Survey (http://www.cgs.gov.cn/). Over the last 20 years, many mines have been developed in the Wuning area.

The Wuning area is in the subtropics monsoon climate region. The wet period is usually from April to August. The total average precipitation for 3 months (April–June) is approximately 638 mm. The largest daily precipitation is more than 100 mm during the wet period. The dry period is commonly from September to January with average precipitation approximately 65.6 mm/month. The annual

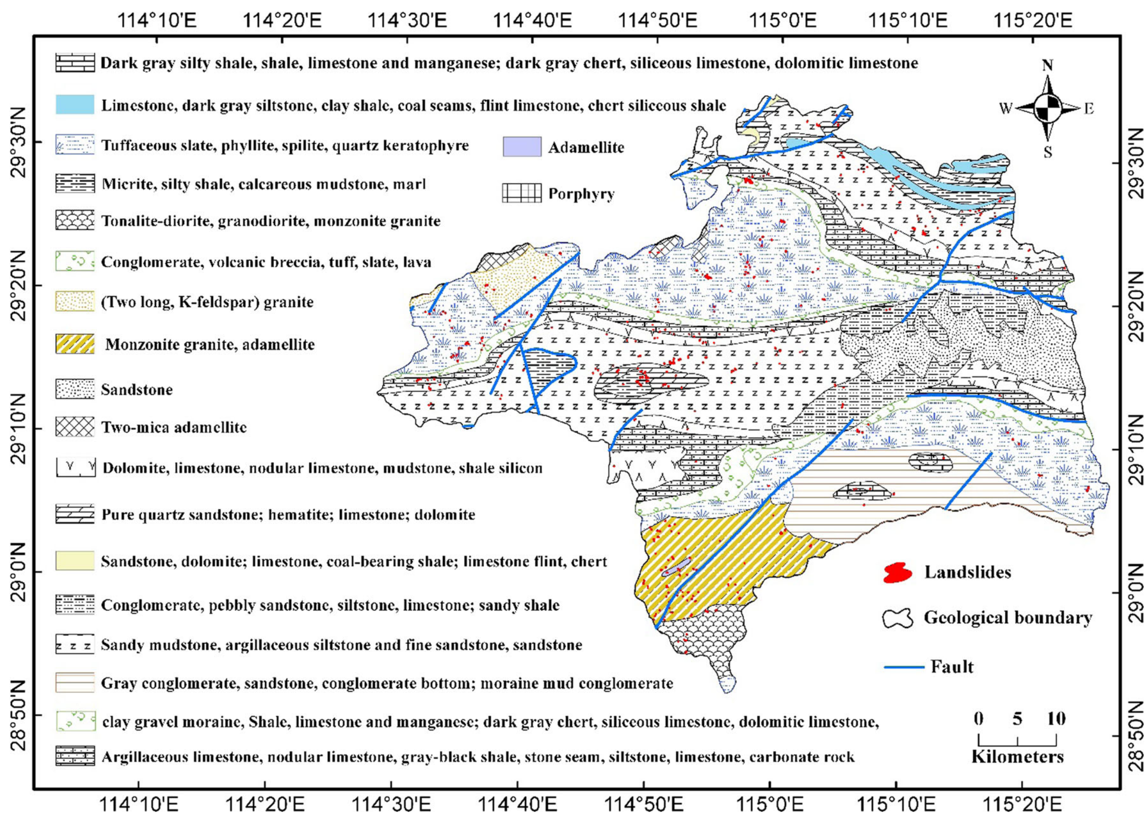


Fig. 3 Geology map of the study area

average precipitation in the Wuning area varies from 578 to 1183 mm, with an average of 162.4 wet days. The annual average temperature is 16.6 °C. The highest and lowest average temperatures are 39.1 and − 5.2 °C, respectively. The annual average sunshine in the Wuning area is 1700.5 h.

Data preparation

Historical landslide events

There are numerous different methods and techniques being used to construct landslide inventory maps. These methods are field survey, satellite image interpretation, aerial photography and historical landslide records (Benoit et al. 2015; Palamakumbure et al. 2015; Varilova et al. 2015; Wang et al. 2015a). However, until now, there has been no agreement among scientists on the best suitable primary method for producing an accurate landslide inventory map ((Hong et al. 2017; Uhlemann et al. 2016).

In this research, the landslide inventory map was generated by combining field survey data with satellite image data. The landslide inventory map for the Wuning area had 445 landslide locations, which were provided by the Jiangxi Province Meteorological Bureau (http://www.weather.org.cn) and the Department of Land and Resources of Jiangxi Province (http://www.cgs.gov.cn; Fig. 2). Figure 4 shows Google Earth photographs of landslides in the study area. The landslide inventory map demonstrates that the volume of the smallest landslide is 20 m^3, the volume of the largest is 96,000 m^3, and the average volume of landslides is 1761.3 m^3. In the study area, large-sized landslides (> 1000 m^3) account for 8.1% of the total landslides, and 254 people have been threatened. Medium-sized (200–1000 m^3) landslides account for approximately 16.0% of the total landslides and have threatened 121 people in the study area. Small-sized landslides (< 200 m^3) have threatened 551 people and account for 75.9% of the total landslides. Heavy rainfall is major reason for these landslide occurrences; there have been no reports about earthquake-induced landslides. Approximately 38.5% of the landslides occurred when the measured rainfall was approximately 100 mm per day. The other landslides occurred when the daily rainfall was greater than 105 mm.

Landslide-conditioning parameters

In the current study, based on the analysis of the landslide inventory map and some literature, a total of 14 landslide-conditioning factors were selected as follows: slope, aspect, altitude, topographic wetness index (TWI), stream power index (SPI), sediment transport index (STI), soil,

Fig. 4 Google Earth photographs showing landslides of the study area

lithology, normalized difference vegetation index (NDVI), land use, rainfall, distance to roads, distance to rivers and distance to faults.

A digital elevation model (DEM) for the Wuning area was produced from ASTER GDEM Version 2 (http://gdem.ersdac.jspacesystems.or.jp). Based on this DEM, slope, altitude, aspect, TWI, SPI and STI were extracted Using Arcgis 10.2 software (Fig. 5a–f). The soil map was compiled in 1995 by the Institute of Soil Science, Chinese Academy of Sciences (ISSCAS) China (http://www.issas.ac.cn; Fig. 5g). The lithology map was categorized into eight groups (A, B, C, D, E, F, G, H, I and J) (Fig. 5h). The NDVI and land use map were obtained from Landsat 7 ETM+ satellite images that were acquired on 10 December 1999 (Fig. 5i, j). These images were obtained from the Computer Network Information Centre of Chinese Academy of Sciences (http://www.gscloud.cn). The NDVI was calculated using the common formula:

$$\mathrm{NDVI} = (\mathrm{NIR} - R)/(\mathrm{NIR} + R) \tag{2}$$

where NIR and Red are the near infrared and red bands, which are from 0.7 to 1 and 0.6 to 0.7 lm, respectively, of the electromagnetic spectrum. The maximum likelihood supervised method was used for the land use classification, with a classification accuracy of 90.7%. Rainfall data are from the Jiangxi Province Meteorological Bureau (http://www.weather.org.cn). For the period 1960–2012, there were 37 rainfall stations that were used to create the rainfall map. The mean annual precipitation was divided into five classes (Fig. 5k) using the inverse distance weighted method (Zhu et al. 2012). Distance to roads and rivers and distance to faults were produced from topographic maps and geological maps, respectively (Fig. 5l–n). The detailed information of the classes for each landslide-influencing factors is provided in Table 2. Finally, all these maps were transformed into the same resolution of 25 m × 25 m.

Results and discussion

Feature selection of linear support vector machine (LSVM)

The Linear Support Vector Machine (LSVM) is a classifier of the Support Vector Machine (SVM) algorithm, which has been widely used in landslide susceptibility modeling (Chou et al. 2016; Conte et al. 2016; Gullà et al. 2016). According to a one-vs-the-rest arrangement, this class holds both sparse and dense input, and the multiclass hold is controlled (Fan et al. 2016; Lora et al. 2016; Romano et al. 2016).

It is very meaningful to assess the predictive ability of an assembling training dataset using fourteen landslide-conditioning causes. In this study, we use the Linear Support Vector Machine method with tenfold cross-validation. Figure 6 presents the predictive ability of landslide-conditioning factors in the Wuning area. It was demonstrated that slope angle has the best predictive ability in landslide susceptibility models (AM = 13.8). Rainfall also has a very high offering in landslide susceptibility models (AM = 12.8). TWI (AM = 12) and STI (AM = 11.4) have relatively high offerings in landslide susceptibility models as well. Slope aspect (AM = 10) and distance to road (AM = 9) have moderate offerings in the modeling.

Fig. 5 Landslide-conditioning factor maps of the Wuning area: **a** slope degree, **b** aspect, **c** altitude, **d** topographic wetness index (TWI), **e** stream power index (SPI), **f** sediment transport index (STI), **g** soil, **h** lithology, **i** normalized difference vegetation index (NDVI), **j** land use, **k** rainfall, **l** distance to roads, **m** distance to rivers, **n** distance to faults

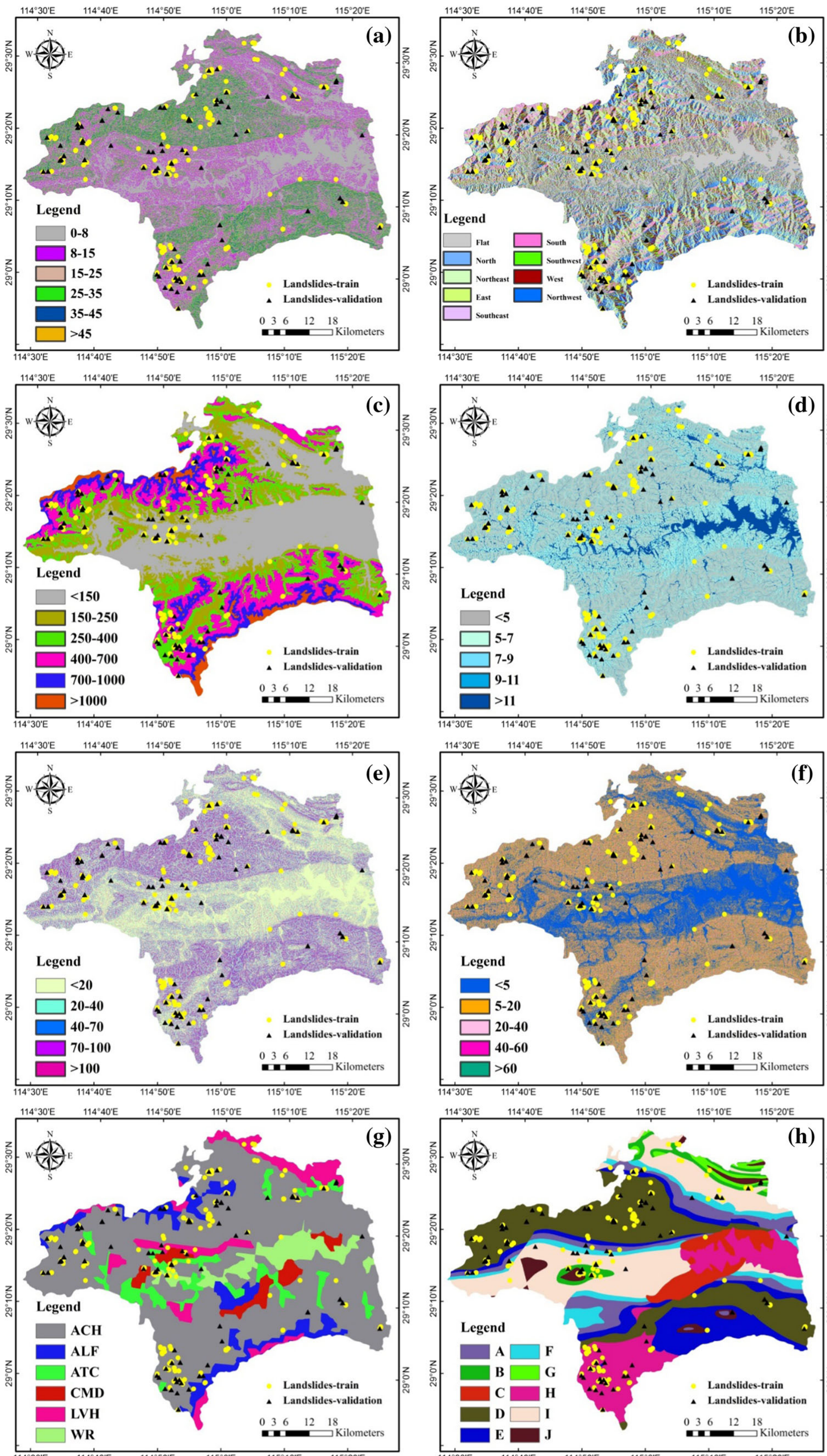
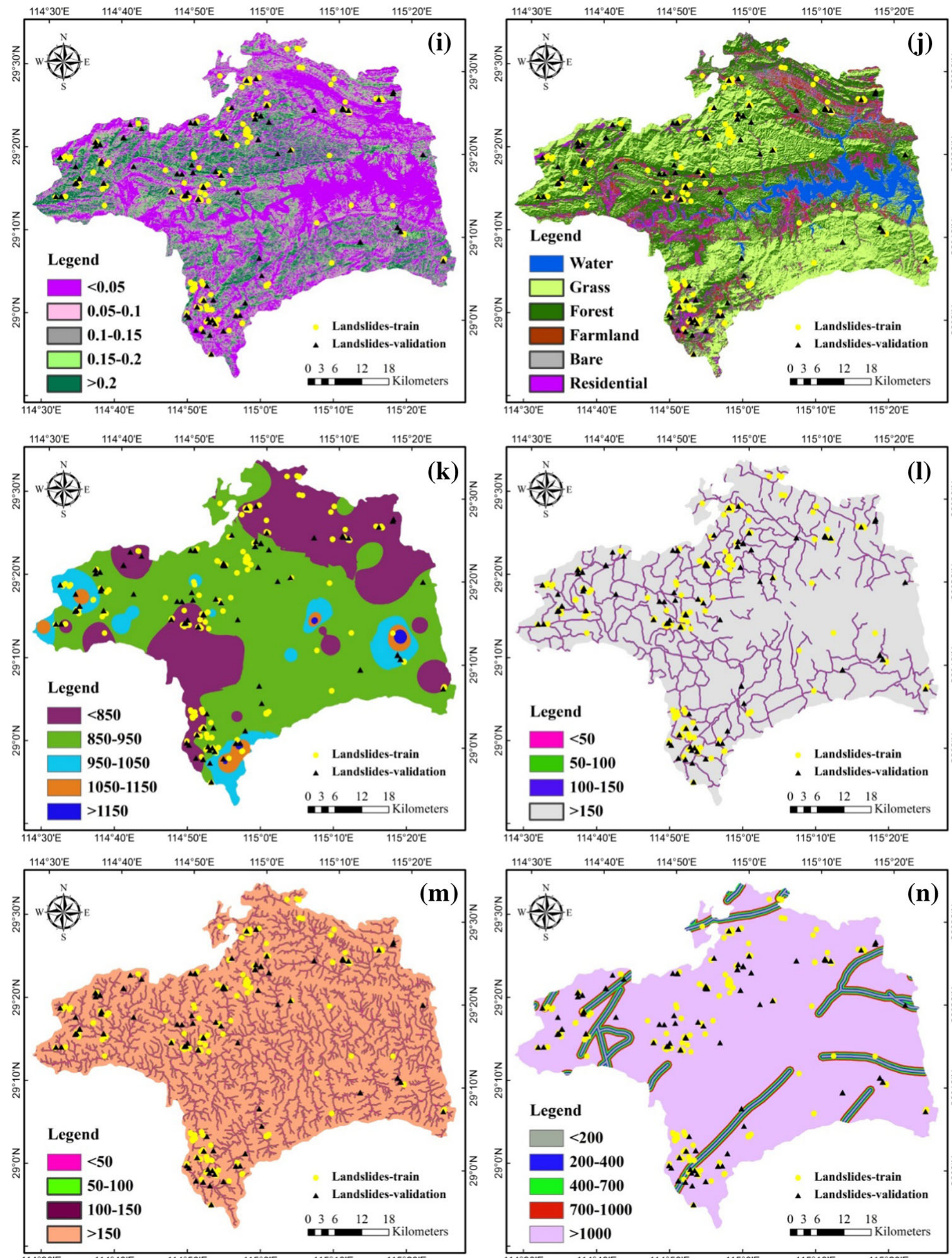


Fig. 5 continued



Other factors, such as distance to faults (AM = 7.3) and distance to rivers (AM = 7.2), had nearly similar predictive abilities. Altitude (AM = 5.4), land use (AM = 4.9) and lithology (AM = 4.9) had similar offerings. In contrast, NDVI (AM = 2.3) and SPI (AM = 2.1) had low predictive ability, and soil type (AM = 1.9) had the lowest predictive ability.

In sum, all fourteen landslide-conditioning factors contributed to the landslide susceptibility models (AM > 0). Overall, these fourteen landslide-conditioning factors have been used in landslide susceptibility.

Preparation of dataset and training the RSSVM model

Performance of the RSSVM model significantly depended on the selection of the calculating parameter, which is the number of iterations. Thus, a test of the performance of the RSSVM model was accomplished with different numbers of iterations to filter the optimal parameter. For this purpose, the ROC curve method was used to evaluate the performance of RSSVM.

Figure 7 shows the analytical results of the ROC curve with various numbers of iterations for training and testing

Table 2 Classes of landslide-influencing factors

No.	Classes	Method
Slope degree (°)	(1) 0–8, (2) 8–15, (3) 15–25, (4) 25–35, (5) 35–45, (6) > 45	Equal interval
Aspect	(1) Flat, (2) North, (3) Northeast, (4) East, (5) Southeast, (6) South, (7) Southwest, (8) West, (9) Northwest	Equal interval
Altitude	(1) 150, (2) 150–250, (3) 250–400, (4) 400–700, (5) 700–1000, (6) > 1000	Defined interval
TWI	(1) < 5, (2) 5–7, (3) 7–9, (4) 9–11, (5) > 11	Equal interval
SPI	(1) < 20, (2) 20–40, (3) 40–70, (4) 70–100, (5) > 100	Defined interval
STI	(1) < 5, (2) 5–20, (3) 20–40, (4) 40–60, (5) > 60	Defined interval
Soil	(1) ACH, (2) ALF, (3) ATC, (4) CMD, (5) LVH, (6) WR	Soil
Lithology	(1) A, (2) B, (3) C, (4) D, (5) E, (6) F, (7) G, (8) H, (9) I, (10) J	Litho facies
NDVI	(1) < 0.05, (2) 0.05–0.1, (3) 0.1–0.15, (4) 0.15–0.2, (5) > 0.2	Equal interval
Landuse	(1) Water, (2) residential area, (3) forest land, (4) bare land, (5) farm land, (6) grass land	Supervised classification
Rainfall	(1) < 850, (2) 850–950, (3) 950–1050, (4) 1050–1150, (5) > 1150	Equal interval
Distance to roads	(1) < 50, (2) 50–100, (3) 100–150, (4) > 150	Equal interval
Distance to rivers	(1) < 50, (2) 50–100, (3) 100–150, (4) > 150	Equal interval
Distance to faults	(1) < 200, (2) 200–400, (3) 400–700, (4) 700–1000, (5) > 1000	Defined interval

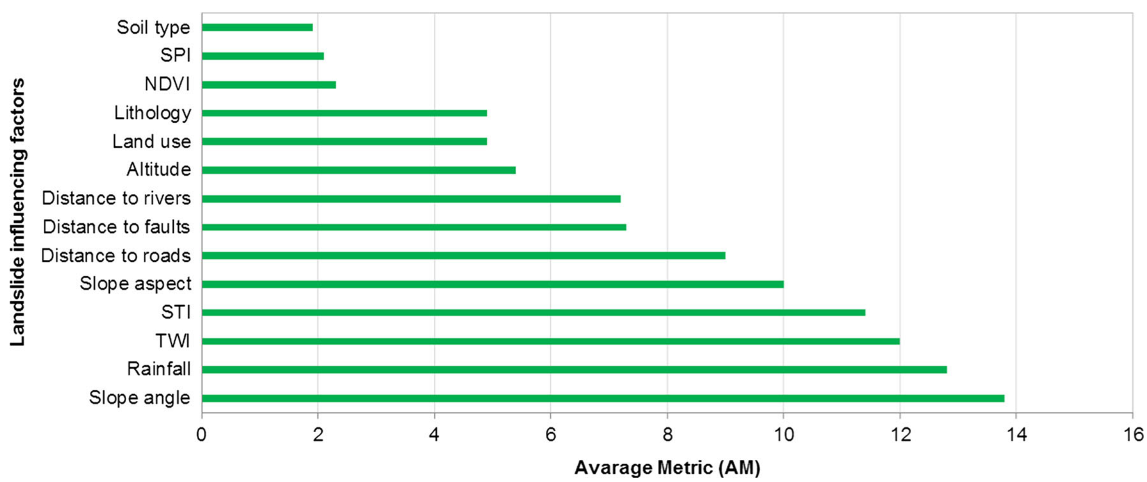


Fig. 6 Predictive ability of landslide trigging factors using the LSVM approach

the RSSVM model. It can be seen that when the number of iterations is 14, this results in the best performance of the RSSVM model. Thus, in the present study, the number of iterations has been set to 14 for training the RSSVM classifier in the novel classifier framework.

Validation of predictive ability of the RSSVM model

Performance of the RSSVM model for landslide susceptibility assessment using statistical index-based evaluations is shown in Table 3. It can be seen that the RSSVM model achieved good classification of both landslide and non-landslide pixels. The positive predictive value is 77.39% in the training dataset and 78.18% in the testing dataset, indicating that the probability the RSSVM model accurately classifies pixels to the landslide class is 77.39 and 78.18%. The sensitivity is 88.62% for the training dataset and 78.18% for the testing dataset indicating that 88.62 and 78.18% of the landslide pixels are accurately classified into the landslide class. Overall, the performance of the RSSVM model for classification of landslide pixels (specificity = 79.93%) is slightly better than those of non-landslide pixels (specificity = 78.18%).

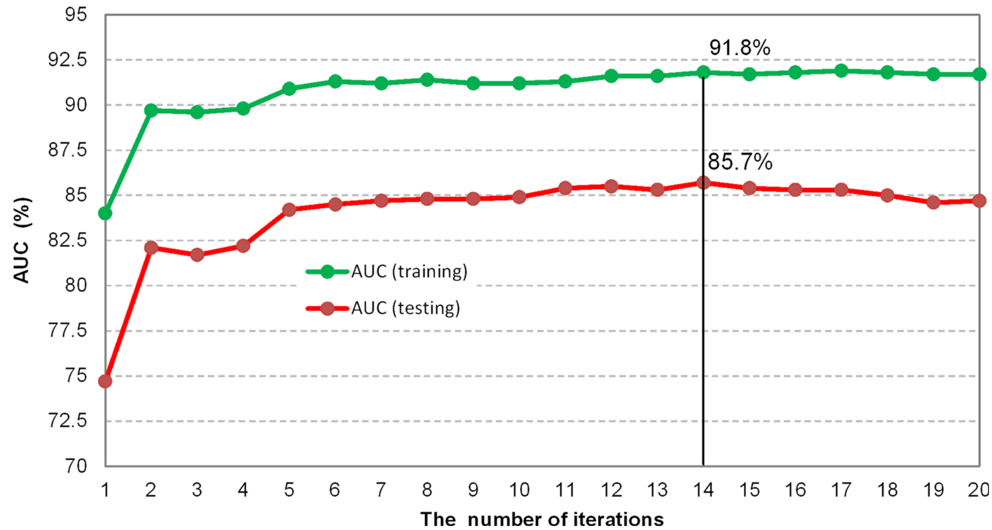


Fig. 7 Analysis of the results of the RSSVM model using ROC curve with various numbers of iterations

Table 3 Performance of the RSSVM model using statistical index-based evaluations

No.	Parameter	Training dataset	Testing dataset
1	True-positive	397.00	172.00
2	True-negative	462.00	172.00
3	False-positive	116.00	48.00
4	False-negative	51.00	48.00
5	Positive predictive value (%)	77.39	78.18
6	Negative predictive value (%)	90.06	78.18
7	Sensitivity (%)	88.62	78.18
8	Specificity (%)	79.93	78.18
9	Accuracy (%)	83.72	78.18

The receiver operating characteristic (ROC) curve was also applied to evaluate the general performance of the RSSVM model. The ROC curve is widely used in landslide susceptibility mapping. In general, the AUC value varies from 0.5 to 1; if the AUC value is equal to 1, the result of the landslide model is excellent; otherwise, if the AUC value is equal to 0.5, the result of the landslide model is imprecise. Figures 8 and 9 show the results of the RSSVM model for landslide susceptibility assessment using the ROC curve technique. In this study, the RSSVM model executed very well based on the analysis of the ROC curve (AUC = 0.918); additionally, the ROC curve for testing the RSSVM model is 0.857, which is reasonably satisfactory.

Comparison of the RSSVM model with popular landslide models

In this study, other popular landslide susceptibility models, such as Multiple Perceptron Neural Network, Naïve Bayes and Support Vector Machine, have been applied and compared to the result of the proposed hybrid model.

Multiple Perceptron Neural Network (MLP) Based on the technique of biological nervous systems, which contain the brain and process information, artificial neural networks are defined as an information processing method (Haeberli et al. 2001; Satorra and Bentler 2001). The real function and effectiveness of neural network algorithms demonstrate their capability to perform both linear and nonlinear connections and to master these connections directly from the modeling data (Carlini et al. 2016; Gutiérrez and Lizaga 2016). Classic linear models are naturally poor when they input modeling data that includes nonlinear information (Andrews 1988; Ye and Chen 2001). As we know, neural network algorithms are being adapted to an expanding number of real-world problems (Dickson and Perry 2016; Wang et al. 2016). Their basic convenience is that they can address issues that are too complicated for normal methods (Leung et al. 2000; Rao and Scott 1987). In general, neural network algorithms are well suited to address problems that include pattern recognition of trends in data (Ogneva-Himmelberger et al. 2009; Song et al. 2014). The most ordinary neural network algorithm is the multiple perceptron. Due to its need for a desired output

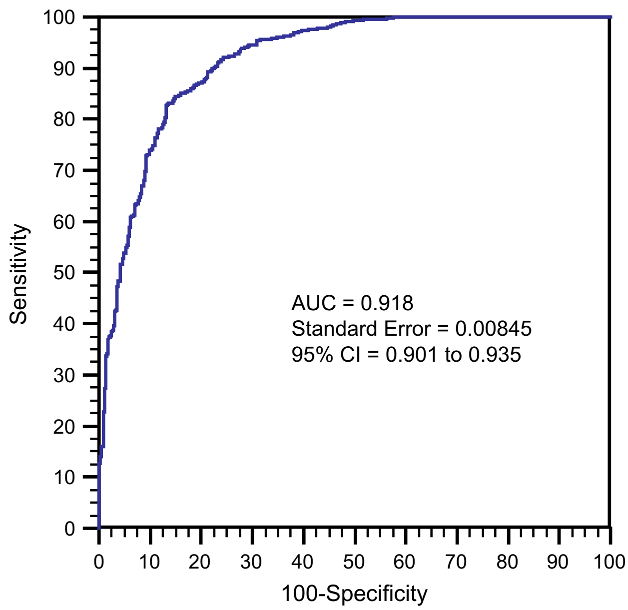


Fig. 8 Analysis of the ROC curve for training the RSSVM model

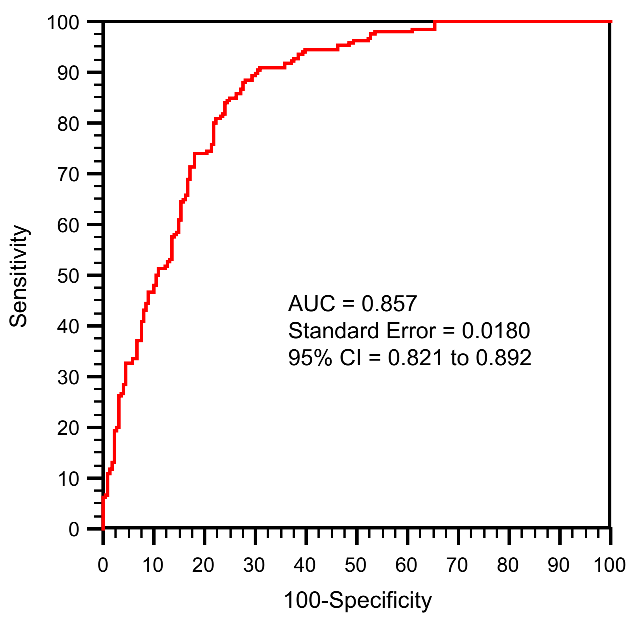


Fig. 9 Analysis of the ROC curve for testing the RSSVM model

to study, this type of neural network algorithm is famous as a supervised system (Moore and Sawyer 2016). The goal of it is to build a model that correctly maps the input and output data so that it can then be applied to obtain the output result even if the output is unknown (Webster et al. 2016). Back propagation is the most common algorithm adopted by the multiple perceptron neural network algorithm (Shi et al. 2016). With back propagation, the input data are often given to the neural network algorithm (Cirurleo et al. 2016). The neural network algorithm always

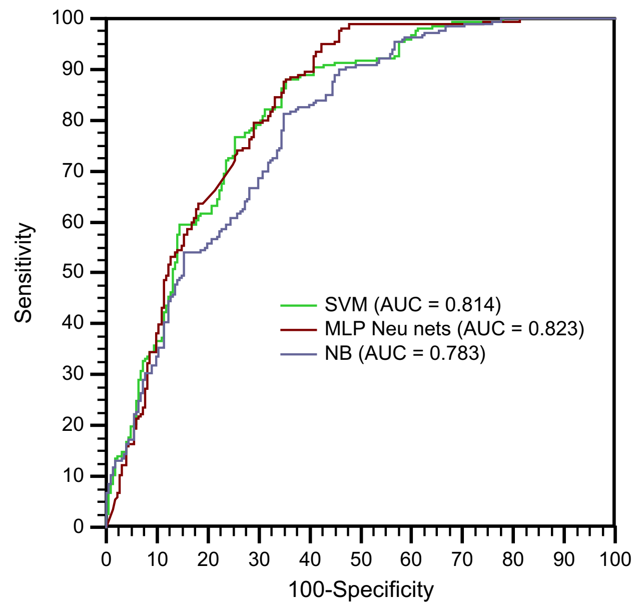


Fig. 10 Comparison of predictive ability of landslide models

adjusts the weights and decreases with each iteration error, moving closer and closer to acquiring the coveted result (Gutiérrez and Lizaga 2016).

Naïve Bayes (NB) The definition of a Bayesian network is a directed acyclic diagram with a probability annotation where every node in the graph represents random variables; two nodes in the graph occur if there is an arc, and the two nodes correspond to the probability of whether a random variable is dependent and conversely indicates that two independent random variables are the conditions (Pirdavani et al. 2014a, b). An arbitrary node in the network of X has a corresponding Conditional Probability Table (Conditional aim-listed Probability Table, CPT), and nodes X in the father obtain all the possible values of the Conditional Probability (Chen et al. 2012; Harris et al. 2010). If nodes are without father X, then X CPT for the prior probability is the distribution (Wei and Qi 2012). A Bayesian network structure of the nodes and the CPT define the likelihood of allocation of each variable in the system. Naïve Bayes models originated through classical mathematics theory and have a solid mathematical foundation as well as the stability of classification efficiency (Zhang et al. 2011). At the same time, the NB model requires only a few estimated parameters, is less sensitive to missing data, and the algorithm is simpler (Zhang and Mei 2011). In theory, the NB model of minimum error rate compares well with other classification methods (Hadayeghi et al. 2010). However, this is not always the case because the NB model assumptions are independent of each other between attributes. This assumption is often not established in practical applications, and this brings a certain influence to the correct classification of the NB model (Koutsias et al.

2010). Depending on the number of attributes or when the attribute correlation is large, the NBC model classification efficiency is better than the decision tree model. If the attribute correlation is small, the performance of the NB model is good (Sharma et al. 2011). The Naïve Bayes theorem is a kind of unsupervised learning with no iteration; the learning efficiency is high, and it is easy to implement under large sample sizes with better performance (Kumar et al. 2012; Lukawska-Matuszewska and Urbanski 2014). However, because the conditional independence assumption is too strong, assumption features associated with the condition of the characteristics of the input vector of scenarios do not apply (Martinez-Fernandez et al. 2013; Paez et al. 2011).

Using the results from analyzing the performance of the ROC curve for different landslide models (Fig. 10), it can be found that the RSSVM model (AUC = 0.857) has the highest performance, followed by the SVM model (AUC = 0.814), MLPN model (AUC = 0.823) and NB model (AUC = 0.783).

Delineating landslide susceptibility maps

In this study, 311 landslides were used for training data (70%), and 134 landslides were used for validation data (30%). First, the training dataset of the SVM and NB models were run using Weka 3.7.12 software. The polygon of the study area in ArcGIS 10.2 was then converted to rasters with a pixel size of 25 × 25 m, which was similar for all conditioning factors. The raster polygon was sampled in GIS with all conditioning factors, and this layer is the test dataset used in the Weka program for landslide

Table 4 Landslide bulk on landslide susceptibility map

Class	LSIs intervals	Pixels of class	Pixels of landslides	% Class	% Landslides	Landside density (LD)
Very low	(0.019–0.153)	1,431,546	0	25.47	0.00	0
Low	(0.153–0.301)	1,330,357	7	23.67	3.57	0.15
Moderate	(0.301–0.459)	1,145,280	23	20.37	11.73	0.58
High	(0.459–0.622)	984,530	56	17.51	28.57	1.63
Very high	(0.622–0.939)	729,472	110	12.98	56.12	4.32

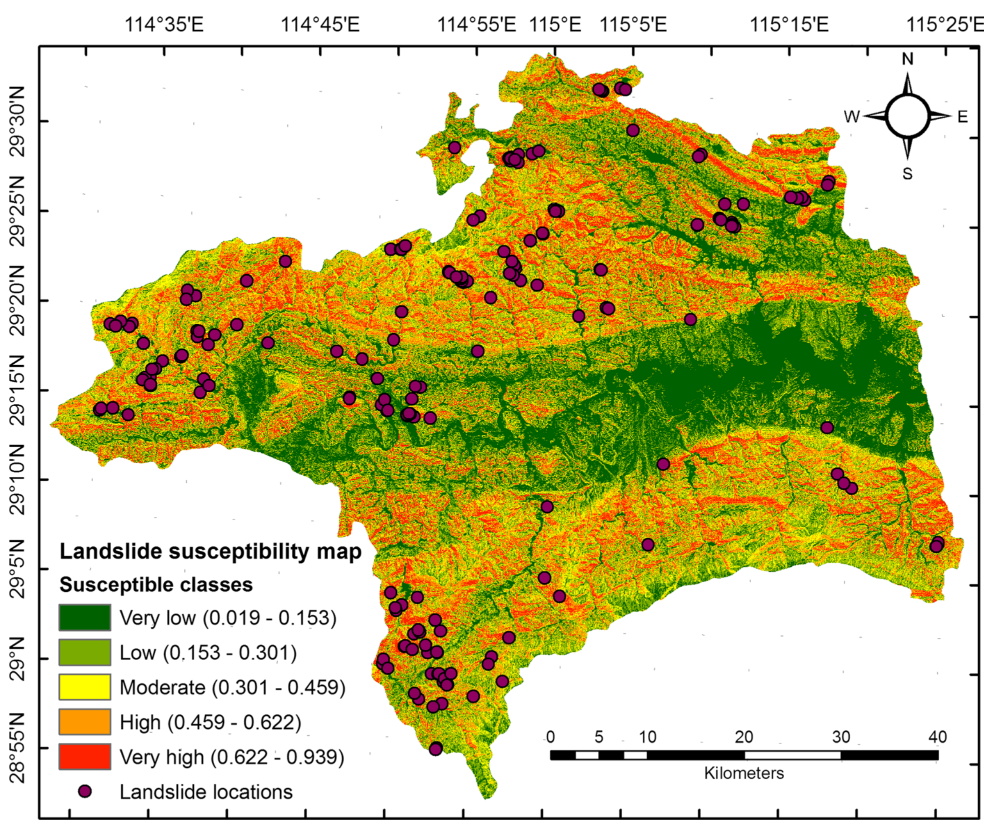


Fig. 11 Landslide susceptibility map generated using the RSSVM model

susceptibility mapping. Consequently, according to the samples for each point raster of the study area, a probability of landslide incidence was obtained and transferred to GIS, and a landslide susceptibility map for the RSSVM model was produced. Ultimately, the range of values of the susceptibility map was classified into five categories based on the natural break classification method (Basofi et al. 2015), including very low susceptibility (VLS), low susceptibility (LS), moderate susceptibility (MS), high susceptibility (HS) and very high susceptibility (VLS) (Dai and Lee 2002; Table 4). The landslide susceptibility map of the RSSVM model is shown in Fig. 11.

Interpretation of the landslide susceptibility map generated using the RSSVM model shows that the very high class covers only 12.98% of the total study area but contains 56.12% of the total landslide locations. In contrast, the low and very low classes account for 49.14% of the total study area; however, they contain only 3.57% of the total landslide locations. This indicates that the RSSVM model produces a high accuracy result and that the map fits well with the landslide inventories.

Conclusions

Landslides are a most dangerous and hugely destructive disaster all over the world. For this reason, landslide susceptibility research is most important for local management and town planners. Many scientists have utilized different methods to develop landslide susceptibility maps in various regions worldwide. However, until now there has been no agreement about the best method to use in landslide susceptibility modeling. Thus, the current study aimed to discover a new ensemble method to achieve this target.

In this study, a novel hybrid integration approach was used by integrating support vector machines (SVM) and random subspace (RS) for landslide susceptibility assessment. The result shows that the RSSVM model performs very well in landslide susceptibility mapping in the Wuning area (China). The predictive ability of a base classifier of SVM is significantly improved through the newly proposed RSSVM model. In comparison with Multiple Perceptron Neural Network (MLP), Naïve Bayes (NB) and Support Vector Machine (SVM), the RSSVM model has the best performance. In general, the landslide susceptibility map of this model is very beneficial for decision makers and land use planners in the Wuning area.

Acknowledgements The authors would like to thank Editor-in-chief James W. LaMoreaux and three anonymous reviewers for their meaningful comments on the primary version of the manuscript. This research was supported by the National Natural Science Foundation of China (Project Nos: 41431177, 41601413), the Natural Science Research Programme of Jiangsu (Project No: BK20150975), the Natural Science Research Programme of Jiangsu (Project No: 14KJA170001), and the National Basic Research Programme of China (Project No: 2015CB954102). Support for A-Xing Zhu through the Vilas Associate Award, the Hammel Faculty Fellow Award, The Manasse Chair Professorship from the University of Wisconsin-Madison, and The "One-Thousand Talents" Programme of China is greatly appreciated.

References

Abolmasov B, Milenkovic S, Marjanovic M, Duric U, Jelisavac B (2015) A geotechnical model of the Umka landslide with reference to landslides in weathered Neogene marls in Serbia. Landslides 12:689–702. doi:10.1007/s10346-014-0499-4

Ahmed B (2015) Landslide susceptibility mapping using multi-criteria evaluation techniques in Chittagong Metropolitan Area, Bangladesh. Landslides 12:1077–1095. doi:10.1007/s10346-014-0521-x

Akcay O (2015) Landslide fissure inference assessment by ANFIS and logistic regression using UAS-based photogrammetry. ISPRS Int J Geo-Inf 4:2131–2158. doi:10.3390/ijgi4042131

Alvioli M, Baum RL (2016) Parallelization of the TRIGRS model for rainfall-induced landslides using the message passing interface. Environ Model Softw 81:122–135. doi:10.1016/j.envsoft.2016.04.002

Andrews DW (1988) Chi square diagnostic tests for econometric models: introduction and applications. J Econ 37:135–156

Basofi A, Fariza A, Ahsan AS, Kamal IM (2015) A comparison between natural and Head/tail breaks in LSI (Landslide Susceptibility Index) classification for landslide susceptibility mapping: a case study in Ponorogo, East Java, Indonesia. International Conference on Science in Information Technology, Yogyakarta, Indonesia, pp 337–342

Bellugi D, Milledge DG, Dietrich WE, McKean JA, Perron JT, Sudderth EB, Kazian B (2015) A spectral clustering search algorithm for predicting shallow landslide size and location. J Geophys Res Earth Surf 120:300–324. doi:10.1002/2014jf003137

Bennett GL, Miller SR, Roering JJ, Schmidt DA (2016) Landslides, threshold slopes, and the survival of relict terrain in the wake of the Mendocino Triple Junction. Geology 44:363–366. doi:10.1130/g37530.1

Benoit L, Briole P, Martin O, Thom C, Malet JP, Ulrich P (2015) Monitoring landslide displacements with the Geocube wireless network of low-cost GPS. Eng Geol 195:111–121. doi:10.1016/j.enggeo.2015.05.020

Bertoni A, Folgieri R, Valentini G (2005) Bio-molecular cancer prediction with random subspace ensembles of support vector machines. Neurocomputing 63:535–539

Bezak N, Šraj M, Mikoš M (2016) Copula-based IDF curves and empirical rainfall thresholds for flash floods and rainfall-induced landslides. J Hydrol. doi:10.1016/j.jhydrol.2016.02.058

Blothe JH, Korup O, Schwanghart W (2015) Large landslides lie low: excess topography in the Himalaya–Karakoram ranges. Geology 43:523–526. doi:10.1130/g36527.1

Boue A, Lesage P, Cortes G, Valette B, Reyes-Davila G (2015) Real-time eruption forecasting using the material failure forecast method with a Bayesian approach. J Geophys Res Solid Earth 120:2143–2161. doi:10.1002/2014jb011637

Bui DT, Pradhan B, Lofman O, Revhaug I, Dick OB (2012) Landslide susceptibility mapping at Hoa Binh province (Vietnam) using an adaptive neuro-fuzzy inference system and GIS. Comput Geosci 45:199–211

Bui DT, Pradhan B, Revhaug I, Nguyen DB, Pham HV, Bui QN (2015) A novel hybrid evidential belief function-based fuzzy logic model in spatial prediction of rainfall-induced shallow landslides in the Lang Son city area (Vietnam). Geomat Nat Hazards Risk 6:243–271. doi:10.1080/19475705.2013.843206

Bui DT, Tuan TA, Klempe H, Pradhan B, Revhaug I (2016) Spatial prediction models for shallow landslide hazards: a comparative assessment of the efficacy of support vector machines, artificial neural networks, kernel logistic regression, and logistic model tree. Landslides 13:361–378. doi:10.1007/s10346-015-0557-6

Bui DT, Nguyen QP, Hoang ND, Klempe H (2017) A novel fuzzy K-nearest neighbor inference model with differential evolution for spatial prediction of rainfall-induced shallow landslides in a tropical hilly area using GIS. Landslides 14(1):1–17. doi:10.1007/s10346-016-0708-4

Carey JM, Moore R, Petley DN (2015) Patterns of movement in the Ventnor landslide complex, Isle of Wight, southern England. Landslides 12:1107–1118. doi:10.1007/s10346-014-0538-1

Carlini M, Chelli A, Vescovi P, Artoni A, Clemenzi L, Tellini C, Torelli L (2016) Tectonic control on the development and distribution of large landslides in the Northern Apennines (Italy). Geomorphology 253:425–437. doi:10.1016/j.geomorph.2015.10.028

Chen G, Zhao KG, McDermid GJ, Hay GJ (2012) The influence of sampling density on geographically weighted regression: a case study using forest canopy height and optical data. Int J Remote Sens 33:2909–2924. doi:10.1080/01431161.2011.624130

Chen W, Chai H, Zhao Z, Wang Q, Hong H (2016a) Landslide susceptibility mapping based on GIS and support vector machine models for the Qianyang County, China. Environ Earth Sci 75:474

Chen W, Wang J, Xie X, Hong H, Trung NV, Tien Bui D, Wang G, Li X (2016b) Spatial prediction of landslide susceptibility using integrated frequency ratio with entropy and support vector machines by different kernel functions. Environ Earth Sci 75:1344

Chen W, Panahi M, Pourghasemi HR (2017a) Performance evaluation of GIS-based new ensemble data mining techniques of adaptive neuro-fuzzy inference system (ANFIS) with genetic algorithm (GA), differential evolution (DE), and particle swarm optimization (PSO) for landslide spatial modelling. CATENA 157:310–324

Chen W, Pourghasemi HR, Kornejady A, Zhang N (2017b) Landslide spatial modeling: introducing new ensembles of ANN, MaxEnt, and SVM machine learning techniques. Geoderma 305:314–327. doi:10.1016/j.geoderma.2017.06.020

Chen W, Pourghasemi HR, Naghibi SA (2017c) A comparative study of landslide susceptibility maps produced using support vector machine with different kernel functions and entropy data mining models in China. Bull Eng Geol Environ. doi:10.1007/s10064-017-1010-y

Chen W, Pourghasemi HR, Naghibi SA (2017d) Prioritization of landslide conditioning factors and its spatial modeling in Shangnan County, China using GIS-based data mining algorithms. Bull Eng Geol Environ. doi:10.1007/s10064-017-1004-9

Chen W, Pourghasemi HR, Zhao Z (2017e) A GIS-based comparative study of Dempster-Shafer, logistic regression and artificial neural network models for landslide susceptibility mapping. Geocarto Int 32:367–385. doi:10.1080/10106049.2016.1140824

Chen W, Xie X, Peng J, Wang J, Duan Z, Hong H (2017f) GIS-based landslide susceptibility modelling: a comparative assessment of kernel logistic regression, Naïve-Bayes tree, and alternating decision tree models. Geomat Nat Hazards Risk. doi:10.1080/19475705.2017.1289250

Chen W, Xie X, Wang J, Pradhan B, Hong H, Bui DT, Duan Z, Ma J (2017g) 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

Chen W, Pourghasemi HR, Panahi M, Kornejady A, Wang J, Xie X, Cao S (2017h) Spatial prediction of landslide susceptibility using an adaptive neuro-fuzzy inference system combined with frequency ratio, generalized additive model, and support vector machine techniques. Geomorphology 297:69–85. doi:10.1016/j.geomorph.2017.09.007

Chou H-T, Lee C-F, Lo C-M (2016) The formation and evolution of a coastal alluvial fan in eastern Taiwan caused by rainfall-induced landslides. Landslides. doi:10.1007/s10346-016-0678-6

Ciabatta L, Brocca L, Massari C, Moramarco T, Puca S, Rinollo A, Gabellani S, Wagner W (2015) Integration of satellite soil moisture and rainfall observations over the italian territory. J Hydrometeorol 16:1341–1355. doi:10.1175/jhm-d-14-0108.1

Ciurleo M, Calvello M, Cascini L (2016) Susceptibility zoning of shallow landslides in fine grained soils by statistical methods. CATENA 139:250–264. doi:10.1016/j.catena.2015.12.017

Clare MA, Talling PJ, Hunt JE (2015) Implications of reduced turbidity current and landslide activity for the initial eocene thermal maximum: evidence from two distal, deep-water sites. Earth Planet Sci Lett 420:102–115. doi:10.1016/j.epsl.2015.03.022

Conoscenti C, Ciaccio M, Caraballo-Arias NA, Gomez-Gutierrez A, Rotigliano E, Agnesi V (2015) Assessment of susceptibility to earth-flow landslide using logistic regression and multivariate adaptive regression splines: a case of the Bence River basin (western Sicily, Italy). Geomorphology 242:49–64. doi:10.1016/j.geomorph.2014.09.020

Conte E, Donato A, Troncone A (2016) A simplified method for predicting rainfall-induced mobility of active landslides. Landslides. doi:10.1007/s10346-016-0692-8

Dai FC, Lee CF (2002) Landslide characteristics and slope instability modeling using GIS, Lantau Island, Hong Kong. Geomorphology 42:213–228. doi:10.1016/S0169-555X(01)00087-3

Dehnavi A, Aghdam IN, Pradhan B, Varzandeh MHM (2015) A new hybrid model using step-wise weight assessment ratio analysis (SWAM) technique and adaptive neuro-fuzzy inference system (ANFIS) for regional landslide hazard assessment in Iran. CATENA 135:122–148. doi:10.1016/j.catena.2015.07.020

Dickson ME, Perry GLW (2016) Identifying the controls on coastal cliff landslides using machine-learning approaches. Environ Model Softw 76:117–127. doi:10.1016/j.envsoft.2015.10.029

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

Fan L, Lehmann P, Or D (2016) Effects of soil spatial variability at the hillslope and catchment scales on characteristics of rainfall-induced landslides. Water Resour Res 52:1781–1799. doi:10.1002/2015WR017758

Feng Z-Y, Lo C-M, Lin Q-F (2016) The characteristics of the seismic signals induced by landslides using a coupling of discrete element and finite difference methods. Landslides. doi:10.1007/s10346-016-0714-6

Gaprindashvili G, Van Westen CJ (2016) Generation of a national landslide hazard and risk map for the country of Georgia. Nat Hazards 80:69–101. doi:10.1007/s11069-015-1958-5

Goetz JN, Brenning A, Petschko H, Leopold P (2015) Evaluating machine learning and statistical prediction techniques for landslide susceptibility modeling. Comput Geosci 81:1–11. doi:10.1016/j.cageo.2015.04.007

Gullà G, Peduto D, Borrelli L, Antronico L, Fornaro G (2016) Geometric and kinematic characterization of landslides affecting urban areas: the Lungro case study (Calabria, Southern Italy). Landslides 4:1–18. doi:10.1007/s10346-015-0676-0

Gutiérrez F, Lizaga I (2016) Sinkholes, collapse structures and large landslides in an active salt dome submerged by a reservoir: the unique case of the Ambal ridge in the Karun River, Zagros Mountains, Iran. Geomorphology 254:88–103. doi:10.1016/j.geomorph.2015.11.020

Hadayeghi A, Shalaby AS, Persaud BN (2010) Development of planning level transportation safety tools using geographically weighted poisson regression. Accid Anal Prev 42:676–688. doi:10.1016/j.aap.2009.10.016

Haeberli W, Schaub Y, Huggel C (2001) Increasing risks related to landslides from degrading permafrost into new lakes in de-glaciating mountain ranges. Geomorphology. doi:10.1016/j.geomorph.2016.02.009

Harris P, Fotheringham AS, Crespo R, Charlton M (2010) The use of geographically weighted regression for spatial prediction: an evaluation of models using simulated data sets. Math Geosci 42:657–680. doi:10.1007/s11004-010-9284-7

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

Hong H, Pourghasemi HR, Pourtaghi ZS (2016) Landslide susceptibility assessment in Lianhua County (China): a comparison between a random forest data mining technique and bivariate and multivariate statistical models. Geomorphology 259:105–118

Hong H, Chen W, Xu C, Youssef AM, Pradhan B, Tien Bui D (2017) Rainfall-induced landslide susceptibility assessment at the Chongren area (China) using frequency ratio, certainty factor, and index of entropy. Geocarto Int 32:139–154. doi:10.1080/10106049.2015.1130086

Iadanza C, Trigila A, Napolitano F (2016) Identification and characterization of rainfall events responsible for triggering of debris flows and shallow landslides. J Hydrol. doi:10.1016/j.jhydrol.2016.01.018

Kirschbaum D, Stanley T, Zhou YP (2015) Spatial and temporal analysis of a global landslide catalog. Geomorphology 249:4–15. doi:10.1016/j.geomorph.2015.03.016

Kotsiantis S (2011) Combining bagging, boosting, rotation forest and random subspace methods. Artif Intell Rev 35:223–240

Koutsias N, Martinez-Fernandez J, Allgower B (2010) Do factors causing wildfires vary in space? Evidence from geographically weighted regression. Gisci Remote Sens 47:221–240. doi:10.2747/1548-1603.47.2.221

Kritikos T, Davies T (2015) Assessment of rainfall-generated shallow landslide/debris-flow susceptibility and runout using a GIS-based approach: application to western Southern Alps of New Zealand. Landslides 12:1051–1075. doi:10.1007/s10346-014-0533-6

Kumar S, Lal R, Liu DS (2012) A geographically weighted regression kriging approach for mapping soil organic carbon stock. Geoderma 189:627–634. doi:10.1016/j.geoderma.2012.05.022

Kuncheva LI, Plumpton CO (2010) Choosing parameters for random subspace ensembles for fMRI classification. In: El Gayar N, Kittler J, Roli F (eds) Multiple classifier systems: 9th international workshop, MCS 2010, Cairo, Egypt, 7–9 April 2010 Proceedings. Springer, Berlin, pp 54–63

Kuncheva LI, Rodríguez JJ, Plumpton CO, Linden DE, Johnston SJ (2010) Random subspace ensembles for fMRI classification. IEEE Trans Med Imaging 29:531–542

LaHusen SR, Duvall AR, Booth AM, Montgomery DR (2016) Surface roughness dating of long-runout landslides near Oso, Washington (USA), reveals persistent postglacial hillslope instability. Geology 44:111–114. doi:10.1130/g37267.1

Lai C, Reinders MJT, Wessels L (2006) Random subspace method for multivariate feature selection. Pattern Recogn Lett 27:1067–1076. doi:10.1016/j.patrec.2005.12.018

Leung Y, Mei C-L, Zhang W-X (2000) Testing for spatial autocorrelation among the residuals of the geographically weighted regression. Environ Plan A 32:871–890

Li G, West AJ, Densmore AL, Hammond DE, Jin Z, Zhang F, Wang J, Hilton RG (2016) Connectivity of earthquake-triggered landslides with the fluvial network: implications for landslide sediment transport after the 2008 Wenchuan earthquake. J Geophys Res Earth Surf 121:703–724

Lin CH, Jan JC, Pu HC, Tu Y, Chen CC, Wu YM (2015) Landslide seismic magnitude. Earth Planet Sci Lett 429:122–127

Lora M, Camporese M, Troch PA, Salandin P (2016) Rainfall-triggered shallow landslides: infiltration dynamics in a physical hillslope model. Hydrol Process. doi:10.1002/hyp.10829

Lukawska-Matuszewska K, Urbanski JA (2014) Prediction of near-bottom water salinity in the Baltic Sea using ordinary least squares and geographically weighted regression models. Estuar Coast Shelf Sci 149:255–263. doi:10.1016/j.ecss.2014.09.003

Ma H-R, Cheng X, Chen L, Zhang H, Xiong H (2016) Automatic identification of shallow landslides based on Worldview2 remote sensing images. APPRES 10:016008–016008. doi:10.1117/1.JRS.10.016008

Martinez-Fernandez J, Chuvieco E, Koutsias N (2013) Modelling long-term fire occurrence factors in Spain by accounting for local variations with geographically weighted regression. Nat Hazards Earth Syst Sci 13:311–327. doi:10.5194/nhess-13-311-2013

Meinhardt M, Fink M, Tunschel H (2015) Landslide susceptibility analysis in central Vietnam based on an incomplete landslide inventory: comparison of a new method to calculate weighting factors by means of bivariate statistics. Geomorphology 234:80–97. doi:10.1016/j.geomorph.2014.12.042

Mertens K, Jacobs L, Maes J, Kabaseke C, Maertens M, Poesen J, Kervyn M, Vranken L (2016) The direct impact of landslides on household income in tropical regions: a case study from the Rwenzori Mountains in Uganda. Sci Total Environ 550:1032–1043. doi:10.1016/j.scitotenv.2016.01.171

Mielniczuk J, Teisseyre P (2014) Using random subspace method for prediction and variable importance assessment in linear regression. Comput Stat Data Anal 71:725–742. doi:10.1016/j.csda.2012.09.018

Moore ZT, Sawyer DE (2016) Assessing post-failure mobility of submarine landslides from seismic geomorphology and physical properties of mass transport deposits: an example from seaward of the Kumano Basin, Nankai Trough, offshore Japan. Mar Geol 374:73–84. doi:10.1016/j.margeo.2016.02.003

Nanni L, Lumini A (2008) Random subspace for an improved biohashing for face authentication. Pattern Recogn Lett 29:295–300

Nasiri Aghdam I, 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 75:1–20

Nicolussi K, Spotl C, Thurner A, Reimer PJ (2015) Precise radiocarbon dating of the giant Weis landslide (Eastern Alps, Austria). Geomorphology 243:87–91. doi:10.1016/j.geomorph.2016.05.001

Ogneva-Himmelberger Y, Pearsall H, Rakshit R (2009) Concrete evidence and geographically weighted regression: a regional analysis of wealth and the land cover in Massachusetts. Appl Geogr 29:478–487

Oliveira SC, Zezere JL, Catalao J, Nico G (2015) The contribution of PSInSAR interferometry to landslide hazard in weak rock-dominated areas. Landslides 12:703–719. doi:10.1007/s10346-014-0522-9

Osadchiev AA, Korotenko KA, Zavialov PO, Chiang WS, Liu CC (2016) Transport and bottom accumulation of fine river

sediments under typhoon conditions and associated submarine landslides: case study of the Peinan River, Taiwan. Nat Hazards Earth Syst Sci 16:41–54. doi:10.5194/nhess-16-41-2016
Paez A, Farber S, Wheeler D (2011) A simulation-based study of geographically weighted regression as a method for investigating spatially varying relationships. Environ Plan A 43:2992–3010. doi:10.1068/a44111
Palamakumbure D, Flentje P, Stirling D (2015) Consideration of optimal pixel resolution in deriving landslide susceptibility zoning within the Sydney Basin, New South Wales, Australia. Comput Geosci 82:13–22. doi:10.1016/j.cageo.2015.05.002
Panek T (2015) Recent progress in landslide dating: a global overview. Prog Phys Geogr 39:168–198. doi:10.1177/0309133314550671
Paulin GL, Bursik MI, Orozco JJZ, Garcia JEF (2015) Landslide susceptibility of volcanic landforms in the Rio El Estado watershed, Pico de Orizaba volcano, Mexico. Nat Hazards 77:559–574. doi:10.1007/s11069-015-1612-2
Peng L, Xu SN, Hou JW, Peng JH (2015) Quantitative risk analysis for landslides: the case of the Three Gorges area, China. Landslides 12:943–960. doi:10.1007/s10346-014-0518-5
Pham BT, Bui DT, Indra P, Dholakia MB (2015) Landslide susceptibility assessment at a part of Uttarakhand Himalaya, India using GIS-based statistical approach of frequency ratio method. Int J Eng Res Technol 4:338–344. doi:10.17577/IJERTV4IS110285
Pham BT, Bui DT, Prakash I, Dholakia M (2016a) Evaluation of predictive ability of support vector machines and naive Bayes trees methods for spatial prediction of landslides in Uttarakhand state (India) using GIS. J Geomat 10:71–79
Pham BT, Pradhan B, Bui DT, Prakash I, Dholakia M (2016b) A comparative study of different machine learning methods for landslide susceptibility assessment: a case study of Uttarakhand area (India). Environ Model Softw 84:240–250
Pham BT, Bui DT, Prakash I, Dholakia MB (2016c) Rotation forest fuzzy rule-based classifier ensemble for spatial prediction of landslides using GIS. Nat Hazards 83:1–31. doi:10.1007/s11069-016-2304-2
Piao Y, Piao M, Jin CH, Shon HS, Chung J-M, Hwang B, Ryu KH (2015) A new ensemble method with feature space partitioning for high-dimensional data classification. Math Probl Eng 2015:590678. doi:10.1155/2015/590678
Pirdavani A, Bellemans T, Brijs T, Kochan B, Wets G (2014a) Assessing the road safety impacts of a teleworking policy by means of geographically weighted regression method. J Transp Geogr 39:96–110. doi:10.1016/j.jtrangeo.2014.06.021
Pirdavani A, Bellemans T, Brijs T, Wets G (2014b) Application of geographically weighted regression technique in spatial analysis of fatal and injury crashes. J Transp Eng. doi:10.1061/(asce)te.1943-5436.0000680
Posner AJ, Georgakakos KP (2015) Normalized landslide index method for susceptibility map development in El Salvador. Nat Hazards 79:1825–1845. doi:10.1007/s11069-015-1930-4
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. Comp Geosci 51:350–365
Promper C, Glade T (2016) Multilayer-exposure maps as a basis for a regional vulnerability assessment for landslides: applied in Waidhofen/Ybbs, Austria. Nat Hazards 82:111–127. doi:10.1007/s11069-016-2311-3
Rao J, Scott A (1987) On simple adjustments to Chi-square tests with sample survey data. Ann Stat 15:385–397
Romano A, Di Risio M, Bellotti G, Molfetta MG, Damiani L, De Girolamo P (2016) Tsunamis generated by landslides at the coast of conical islands: experimental benchmark dataset for mathematical model validation. Landslides. doi:10.1007/s10346-016-0696-4
Satorra A, Bentler PM (2001) A scaled difference Chi square test statistic for moment structure analysis. Psychometrika 66:507–514
Sewell RJ, Parry S, Millis SW, Wang N, Rieser U, DeWitt R (2015) Dating of debris flow fan complexes from Lantau Island, Hong Kong, China: the potential relationship between landslide activity and climate change. Geomorphology 248:205–227. doi:10.1016/j.geomorph.2015.07.041
Shahabi H, Hashim M (2015) Landslide susceptibility mapping using GIS-based statistical models and remote sensing data in tropical environment. Sci Rep. doi:10.1038/srep09899
Sharma V, Irmak S, Kabenge I (2011) Application of GIS and geographically weighted regression to evaluate the spatial non-stationarity relationships between precipitation vs. irrigated and rainfed maize and soybean yields. Trans Asabe 54:953–972
Shi JS, Wu LZ, Wu SR, Li B, Wang T, Xin P (2016) Analysis of the causes of large-scale loess landslides in Baoji, China. Geomorphology 264:109–117. doi:10.1016/j.geomorph.2016.04.013
Skurichina M, Duin RP (2001) Bagging and the random subspace method for redundant feature spaces. Multiple classifier systems. Springer, Berlin, pp 1–10
Skurichina M, Duin RP (2002) Bagging, boosting and the random subspace method for linear classifiers. Pattern Anal Appl 5:121–135
Song WZ, Jia HF, Huang JF, Zhang YY (2014) A satellite-based geographically weighted regression model for regional PM2.5 estimation over the Pearl River Delta region in China. Remote Sens Environ 154:1–7. doi:10.1016/j.rse.2014.08.008
Sun S, Zhang C (2007) The selective random subspace predictor for traffic flow forecasting. IEEE Trans Intell Transp Syst 8:367–373
Tan YM, Guo D, Xu B (2015) A geospatial information quantity model for regional landslide risk assessment. Nat Hazards 79:1385–1398. doi:10.1007/s11069-015-1909-1
Tao D, Tang X, Li X, Wu X (2006) Asymmetric bagging and random subspace for support vector machines-based relevance feedback in image retrieval. IEEE Trans Pattern Anal Mach Intell 28:1088–1099
Tin Kam H (1998) The random subspace method for constructing decision forests. IEEE Trans Pattern Anal Mach Intell 20:832–844. doi:10.1109/34.709601
Trigila A, Iadanza C, Esposito C, Scarascia-Mugnozza G (2015) Comparison of logistic regression and random forests techniques for shallow landslide susceptibility assessment in Giampilieri (NE Sicily, Italy). Geomorphology 249:119–136. doi:10.1016/j.geomorph.2015.06.001
Tsangaratos P, Ilia I, Hong H, Chen W, Xu C (2017) Applying information theory and GIS-based quantitative methods to produce landslide susceptibility maps in Nancheng County, China. Landslides 14:1091–1111. doi:10.1007/s10346-016-0769-4
Uhlemann S, Smith A, Chambers J, Dixon N, Dijkstra T, Haslam E, Meldrum P, Merritt A, Gunn D, Mackay J (2016) Assessment of ground-based monitoring techniques applied to landslide investigations. Geomorphology 253:438–451. doi:10.1016/j.geomorph.2015.10.027
Varilova Z, Kropacek J, Zvelebil J, St'astny M, Vilimek V (2015) Reactivation of mass movements in Dessie graben, the example of an active landslide area in the Ethiopian Highlands. Landslides 12:985–996. doi:10.1007/s10346-015-0613-2
Vilimek V, Smolikova J (2015) Scientific research for landslide risk analysis and international education for mitigation and preparedness. Landslides 12:1227–1231. doi:10.1007/s10346-015-0636-8

Vranken L, Vantilt G, Eeckhaut MVD, Vandekerckhove L, Poesen J (2015) Landslide risk assessment in a densely populated hilly area. Landslides 12:787–798

Wang GQ, Bao Y, Cuddus Y, Jia XY, Serna J, Jing Q (2015a) A methodology to derive precise landslide displacement time series from continuous GPS observations in tectonically active and cold regions: a case study in Alaska. Nat Hazards 77:1939–1961. doi:10.1007/s11069-015-1684-z

Wang LJ, Guo M, Sawada K, Lin J, Zhang JC (2015b) Landslide susceptibility mapping in Mizunami City, Japan: a comparison between logistic regression, bivariate statistical analysis and multivariate adaptive regression spline models. CATENA 135:271–282. doi:10.1016/j.catena.2015.08.007

Wang Y, Song C, Lin Q, Li J (2016) Occurrence probability assessment of earthquake-triggered landslides with Newmark displacement values and logistic regression: the Wenchuan earthquake, China. Geomorphology 258:108–119. doi:10.1016/j.geomorph.2016.01.004

Webster JM, George NPJ, Beaman RJ, Hill J, Puga-Bernabéu Á, Hinestrosa G, Abbey EA, Daniell JJ (2016) Submarine landslides on the Great Barrier Reef shelf edge and upper slope: a mechanism for generating tsunamis on the north-east Australian coast? Mar Geol 371:120–129. doi:10.1016/j.margeo.2015.11.008

Wei CH, Qi F (2012) On the estimation and testing of mixed geographically weighted regression models. Econ Model 29:2615–2620. doi:10.1016/j.econmod.2012.08.015

Yamao M, Sidle RC, Gomi T, Imaizumi F (2016) Characteristics of landslides in unwelded pyroclastic flow deposits, southern Kyushu, Japan. Nat Hazards Earth Syst Sci 16:617–627. doi:10.5194/nhess-16-617-2016

Yang HQ, Lan YF, Lu L, Zhou XP (2015) A quasi-three-dimensional spring-deformable-block model for runout analysis of rapid landslide motion. Eng Geol 185:20–32. doi:10.1016/j.enggeo.2014.11.016

Ye N, Chen Q (2001) An anomaly detection technique based on a Chi square statistic for detecting intrusions into information systems. Qual Reliab Eng Int 17:105–112

Youssef AM, Pourghasemi HR, El-Haddad BA, Dhahry BK (2016) Landslide susceptibility maps using different probabilistic and bivariate statistical models and comparison of their performance at Wadi Itwad Basin, Asir Region, Saudi Arabia. Bull Eng Geol Environ 75:63–87. doi:10.1007/s10064-015-0734-9

Yusof NM, Pradhan B, Shafri HZM, Jebur MN, Yusoff Z (2015) Spatial landslide hazard assessment along the Jelapang Corridor of the North-South Expressway in Malaysia using high resolution airborne LiDAR data. Arab J Geosci 8:9789–9800. doi:10.1007/s12517-015-1937-x

Zhang X, Jia Y (2007) A linear discriminant analysis framework based on random subspace for face recognition. Pattern Recognit 40:2585–2591

Zhang HG, Mei CL (2011) Local least absolute deviation estimation of spatially varying coefficient models: robust geographically weighted regression approaches. Int J Geogr Inf Sci 25:1467–1489. doi:10.1080/13658816.2010.528420

Zhang CS, Tang Y, Xu XL, Kiely G (2011) Towards spatial geochemical modelling: use of geographically weighted regression for mapping soil organic carbon contents in Ireland. Appl Geochem 26:1239–1248. doi:10.1016/j.apgeochem.2011.04.014

Zhang J, Gurung DR, Liu R, Murthy MSR, Su F (2015) Abe Barek landslide and landslide susceptibility assessment in Badakhshan Province, Afghanistan. Landslides 12:597–609. doi:10.1007/s10346-015-0558-5

Zhu Y, Liu J, Chen S (2009) Semi-random subspace method for face recognition. Image Vis Comput 27:1358–1370

Zhu JX, Zhang LZ, Zhou XY, Liang GL, Wang Q, Cai ZZ (2012) Application of inverse distance weighted method to regional landslide hazards assessment. Bull Soil Water Conserv 32(3):136–140