A Remote Sensing-Based Approach for Debris-Flow Susceptibility Assessment Using Artificial Neural Networks and Logistic Regression Modeling

Racha Elkadiri, Mohamed Sultan, Ahmed M. Youssef, Tamer Elbayoumi, Ronald Chase, Ali B. Bulkhi, and Mohamed M. Al-Katheeri

***Abstract*—Efforts to map the distribution of debris flows, to assess the factors controlling their development, and to identify the areas susceptible to their occurrences are often hampered by the paucity of monitoring systems and historical databases in many parts of the world. In this paper, we develop and successfully apply methodologies that rely heavily on readily available remote-sensing datasets over the Jazan province in the Red Sea hills of Saudi Arabia. A fivefold exercise was conducted: 1) a geographical information system (GIS) with a Web interface was generated to host and analyze relevant coregistered remote-sensing data and derived products; 2) an inventory was compiled for debris flows identified from satellite datasets (e.g., GeoEye, Orbview), a subset of which was field verified; 3) spatial analyses were conducted in a GIS environment and 10 predisposing factors were identified; 4) an artificial neural network (ANN) model and a logistic regression (LR) model were constructed, optimized, and validated; and 5) the generated models were used to produce debris-flow susceptibility maps. Findings include: 1) excellent prediction performance for both models (ANN: 96.1%; LR: 96.3%); 2) the high correspondence between model outputs (91.5% of the predictions were common) reinforces the validity of the debris-flow susceptibility results; 3) the variables with the highest predictive power were topographic position index (TPI), slope, distance to drainage line (DTDL), and normalized difference vegetation index (NDVI); and 4) the adopted methodologies are reliable, cost-effective, and could potentially be applied over many of the world's data-scarce mountainous lands, particularly along the Red Sea Hills.**

***Index Terms*—Artificial neural networks (ANN), data mining, data-scarce field regions, debris flows, geographical information system (GIS), logistic regression (LR), remote sensing.**

I. Introduction

DEBRIS FLOWS are sudden mass movements that are usually initiated by intense precipitation on unconsolidated steep mountain channels [1], [2]. They pose a substantial threat to human life and property in mountainous areas, especially in populous regions that are witnessing unplanned urbanization activities. These mass movements are often associated with losses of human lives and property [3]–[5]. Increasing awareness of the social, financial, and economic impacts of these environmental hazards is stimulating worldwide investigations aimed at understanding the factors controlling their distribution and predicting the areas susceptible to their development [6], [7].

In many vulnerable areas of the world, hazard susceptibility analyses are hampered by the scarcity or unavailability of historical mass movement archives and relevant spatial datasets. Monitoring mass movement systems typically requires extensive networks of ground-based stations. Although such systems are very useful and may save lives and protect property in many populated areas [4], this level of data collection and analysis is not available in most of the vulnerable regions of developing countries. Moreover, debris flows typically develop in rugged and inaccessible terrains where field work is difficult, expensive, and time consuming. Fortunately, recent advances in remote sensing hold the promise to address these inadequacies. In this paper, we develop cost-effective approaches that rely on observations extracted from a range of remote-sensing datasets, procedures that could be readily applicable to large vulnerable sectors of the Earth's mountainous regions.

We chose our study area within the Jazan province in the Red Sea Hills. The Jazan study area is characterized by rugged terrain, restricted accessibility, large areal extent (>2700 km^2), frequent debris-flow occurrences, and limited field data for the study area. As is the case for many developing countries, Jazan lacks ground-based systems to monitor and archive historical mass movement occurrences.

We used inferred and validated remote-sensing debris-flow events alongside the extracted spatial relevant datasets to derive debris-flow susceptibility models for the area. Several modeling approaches were advanced to assess the susceptibility to mass movement in an area. These models could be grouped into three main categories: 1) heuristic; 2) deterministic; and 3) statistical. Heuristic models rely on expert knowledge to assign weights to the various controlling factors [8], [9]. Although some of these models prove to be successful in specific locations, they are highly subjective and site specific [10]. Deterministic models are largely based on well-characterized mathematical relationships encompassing the physical laws driving mass movements. The required

Manuscript received February 07, 2014; accepted May 28, 2014. Date of publication August 10, 2014; date of current version January 21, 2015.

R. Elkadiri, M. Sultan, and R. Chase are with the Department of Geosciences, Western Michigan University, Kalamazoo, MI 49008 USA (e-mail: racha.elkadiri@wmich.edu; mohamed.sultan@wmich.edu; ronald.chase@wmich.edu).

A. M. Youssef, A. B. Bulkhi, and M. M. Al-Katheeri are with the Saudi Geological Survey, Jeddah 21514, Saudi Arabia (e-mail: Ahmed.AM@sgs.org.sa; Bulkhi.ab@sgs.org.sa; Katheri.mm@sgs.org.sa).

T. Elbayoumi is with the Department of Statistics, Oklahoma State University, Stillwater, OK 74074 USA (e-mail: tamer.m.elbayoumi@okstate.edu).

Digital Object Identifier 10.1109/JSTARS.2014.2337273

1939-1404 © 2014 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

data include geotechnical characteristics of the rocks and soils, slope measurements, and hydrological conditions [11]–[13]. The advantage of deterministic models is that they permit rigorous quantification of stability variables; the main problem is their need for exhaustive data from individual slopes, which makes these methods effective for studying only small areas [14], [15]. Statistical models are data driven; they can be used to analyze the effect of each controlling factor separately, as in bivariate statistical models [16]–[18] or all factors jointly as in the case of multivariate statistical models [15], [19]. In recent years, soft computing techniques have been used in mass movement susceptibility analyses, such as neuro-fuzzy [20], [21], back-propagation artificial neural network (ANN) [22], and fuzzy logic models [23].

We adopted a statistical model approach to assess debris-flow susceptibility in our study area. The statistical approach was selected for a number of reasons. It is suitable for regional-scale applications [24] and is thus more suited for the use of remote-sensing datasets rather than extensive field data, as in the case of deterministic models. In addition, the approach is highly objective because it is built on actual cases rather than expert knowledge [25]. Moreover, by definition, statistical methods readily allow independent evaluation and cross validation [26].

Model performance is becoming of increasing importance as the numbers of statistical models addressing mass movements continue to increase. Comparisons between models, quality assessment, and evaluation of their performances have been the focus of multiple studies. Given the multiple successes reported of ANN and logistic regression (LR) models in a wide range of geo-environments [24], [27]–[30] compared to other models, ANN (an artificial intelligence statistical method) and LR (a traditional statistical method) were used in this study to investigate debris-flow susceptibility in the Jazan area. The two methods have different theoretical backgrounds and if we were to find agreement between the results generated from the two data-mining methodologies, this would provide an additional strong indication of the validity of the obtained results.

The study area, Jazan, is typical of many of the landslide locations worldwide. Such areas are found in steep mountainous terrains that are inaccessible. Installing and maintaining adequate monitoring systems in these locations are not an easy task. The significance of this study emerges from the implementation of methodologies that rely heavily on remote-sensing datasets to account for the paucity or lack of ground systems and historical databases. Moreover, Jazan climate, landscape, and hydrologic settings resemble those of surrounding areas along the Red Sea hills in Yemen, Saudi Arabia, Sinai, and Sudan (2.7 million km^2, 2% of the Earth's continental area); hence, results could potentially be applicable to neighboring countries and are ideal for application in developing countries that have limited resources and lack adequate field observations.

We first report the characteristics (geologic, hydrogeologic, climatic) of the study area (Section II), followed by a description of the available data and the adopted methodology (Section III), and finally, we report our results and findings (Section IV).

II. Study Area Settings

The study area (2717 km^2) covers most of the mountainous areas within the Jazan province. It is located at the extreme southwestern part of the Red Sea hills of Saudi Arabia (approximately $42°45'00''$E to $43°12'00''$E and $17°11'30''$N to $17°41'00''$N), bordering Yemen to the south and the Red Sea to the west (Fig. 1).

The discovery of oil in the Kingdom of Saudi Arabia signaled an important era in the construction of new urban centers and networks of highways. The Jazan area is no exception. However, the economic development of the area has been hampered by intensified debris-flow occurrences that put the population and property (buildings, roads, infrastructure, etc.) at risk.

The Jazan area has all the ingredients to be at high risk for debris-flow development. In fact, the area is characterized by highly elevated mountains (up to 2.5 km amsl) [Fig. 4(a)] with steep topographic reliefs (up to 84°) [Fig. 4(b)]. Rainfall is the main triggering mechanism for debris flows in the study area. Jazan receives around 550 mm/year of rainfall, with a peak record of 1400 mm in 1979 [31]. The high precipitation levels in Jazan are largely related to the presence of tropical air masses that reach the extreme southwestern part of Saudi Arabia and to the high elevations that induce orographic precipitation [32]. In addition, the study area is characterized by intense structural deformation (e.g., folds, faults, fractures). The Red Sea hills of the Arabian Peninsula are composed of volcanic, volcaniclastic, plutonic, and metamorphic rocks that were formed by the accretion of island arcs and closure of interleaving oceanic arcs $700-1000$ Ma [33], [34]. These rocks were elevated with the opening of the Red Sea some 30 million years ago, exposing this complex along the length of the Red Sea on the African and Arabian sides [35]. The basement complex is unconformably overlain by Paleozoic, Mesozoic, and Tertiary sedimentary successions [36], [37].

III. Data and Methods

A fivefold methodology was adopted. First, we generated a database that incorporates relevant coregistered remote-sensing datasets and derived products. The generated datasets were compiled and hosted in a geographical information system (GIS) for archival and spatial data analysis and a Web-based GIS [38] for data distribution (Step I). The compiled datasets were essential for the implementation of the remaining steps. Second, we compiled an inventory of known locations of debris flows that were observed in the field or were clearly identified from high-resolution satellite imagery (Step II). Third, we investigated the factors that could potentially control the occurrence of the identified debris flows (Step III). Fourth, we constructed and validated debris-flow susceptibility models for the Jazan area using LR and ANN pattern-recognition modeling (Step IV). For each model, we tested model sensitivity to input selection using a stepwise approach and receiver operating characteristics (ROC) test. The models were constructed

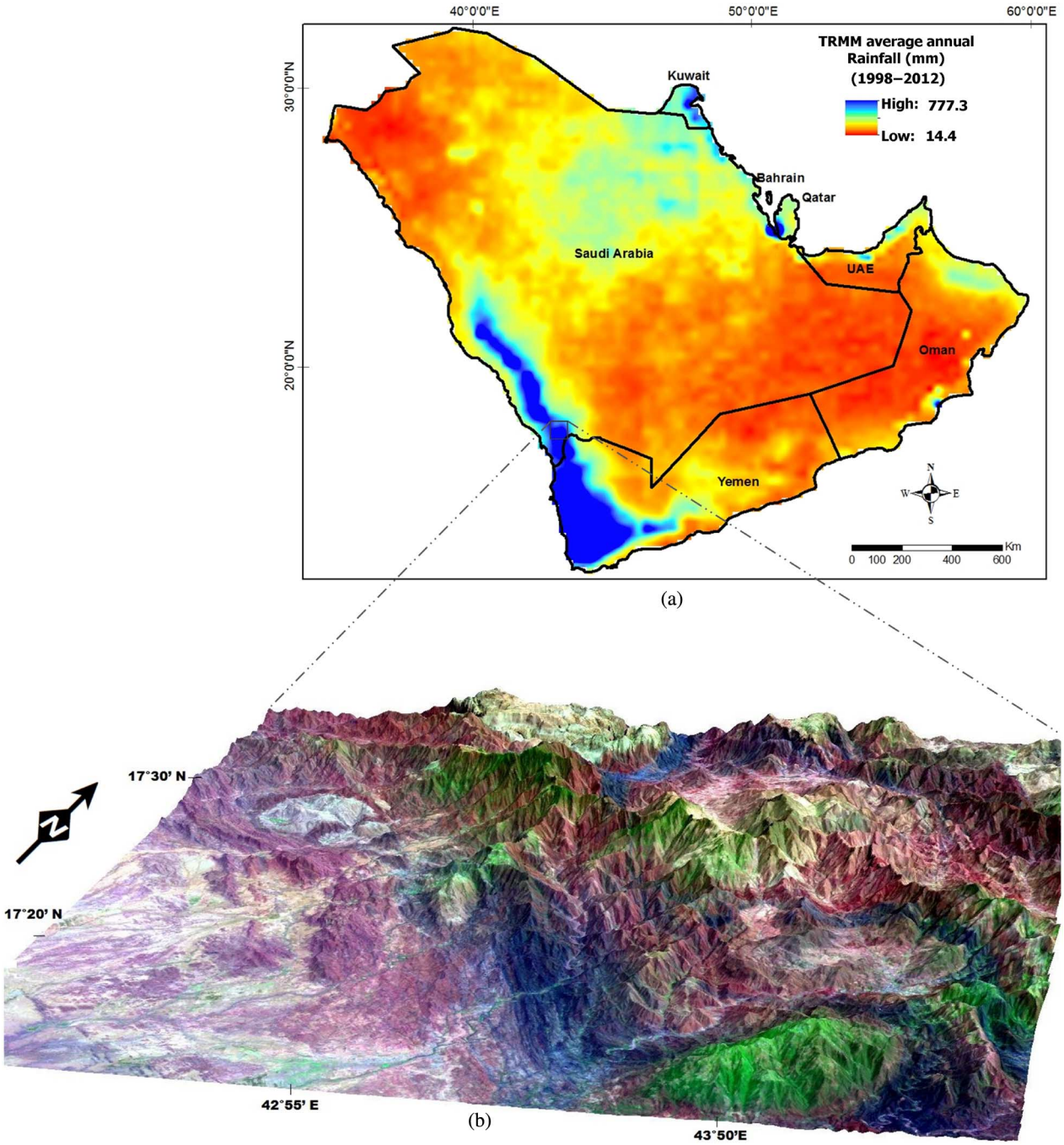


Fig. 1. (a) Tropical Rainfall Measuring Mission (TRMM) average annual rainfall (from 1998 to 2012) showing that the highest amounts of precipitation are in the extreme southwestern parts of the Arabian Peninsula, the study area. (b) Three-dimensional surface model of the study area overlain by Landsat 8 false color composite (Band 7: red, Band 4: green, and Band 2: blue) showing the lithological, topographical, and morphological heterogeneity of the area.

using a randomly selected subset (80%) of the debris-flow inventory, whereas the remaining 20% were used for cross validation. The validated models were then used to predict susceptibility to debris flows across the study area (Step V).

A. Generation of GIS Platforms (Step 1)

The GIS platforms (desktop and Web-based) were developed to organize and manage the datasets and to provide a platform for users to access and visualize the accumulated datasets for the study area. The GIS database incorporates coregistered and orthorectified digital mosaics generated from relevant datasets with a unified projection (UTM-Zone 38, WGS84 datum). The image processing and GIS analyses were conducted using commercial software (ENVI 4.8 and ArcGIS 10.1). The Web-based GIS is a hybrid system that takes advantage of the existing tools and datasets in Google Maps, applies Python scripts to generate custom tools, and uses the ArcGIS server to host the services. Fig. 2 shows the various data sources, datasets, and their derived products that compose the GIS database. The data used in this study originate mainly from remote-sensing datasets along with

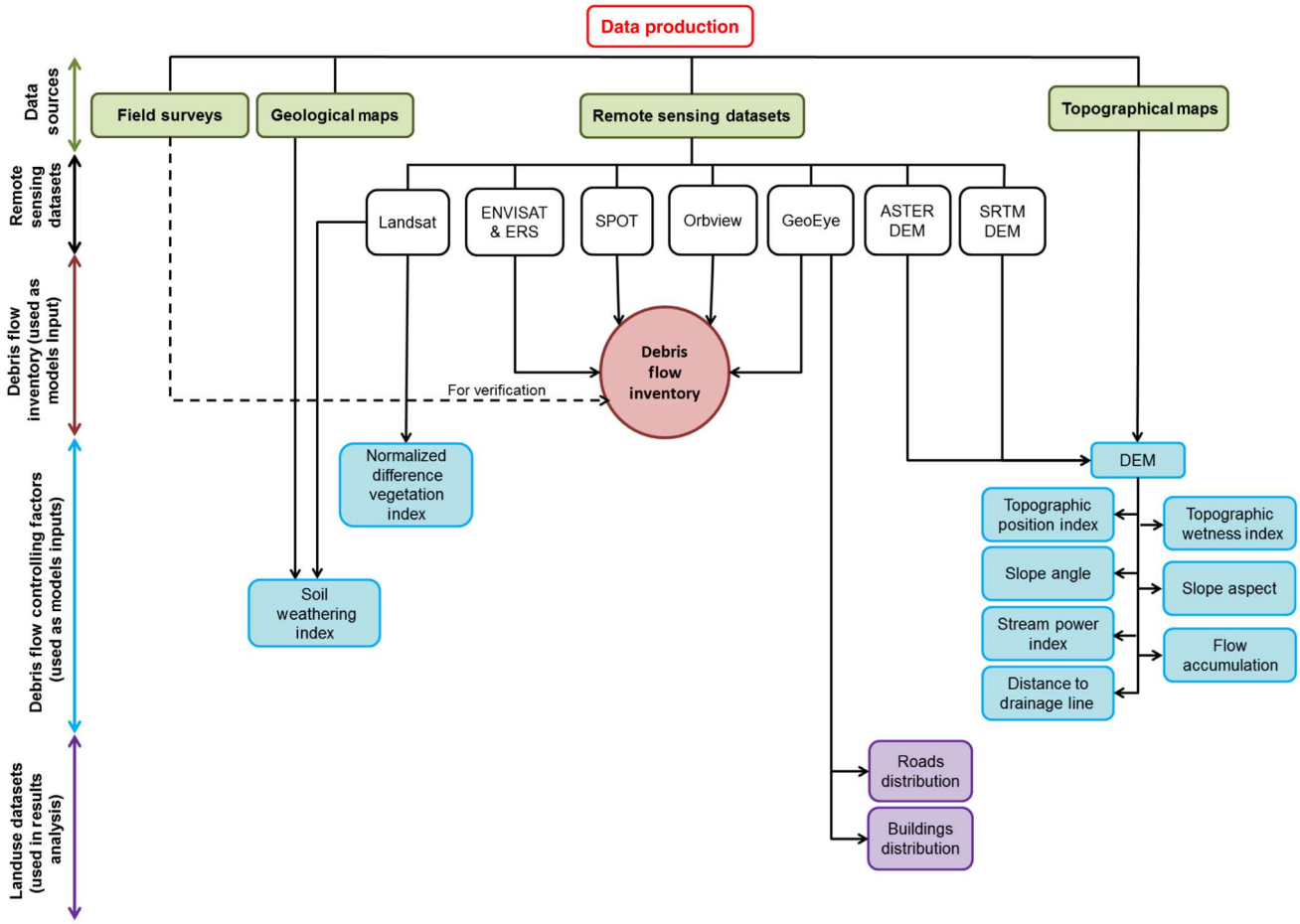


Fig. 2. Schematic diagram of the various datasets, their derived products, and their usage. The generated data products constitute the GIS platform's content.

geologic maps, topographical maps, and limited field surveys. Details on these datasets (e.g., spatial resolution, date of acquisition) are available in Appendix I.

The assembled digital products assisted in addressing the following:

1) Data analysis and distribution via our Web-based GIS [38].
2) Compilation of the inventory for known debris-flow locations (Section III-B) (the latter were used as inputs for the debris-flow susceptibility models).
3) Identification of factors that control the occurrence of the identified debris flows (Section III-C) (the factors served as inputs for the debris-flow susceptibility models).
4) Risk analysis by assessing the effect of the delineated debris flows on the roads and buildings in the study area (Section IV-B).

B. Compilation of Inventory for a Suite of Debris Flows (Step II)

Debris flows are more likely to occur under the same conditions that caused earlier instabilities [39]. Thus, the development and acquisition of a well-documented debris-flow inventory for the study area represent a fundamental requirement for statistical debris-flow modeling [40]. Unfortunately, efforts to construct such inventories are hampered in many of the vulnerable parts of the world, including Jazan area, by the absence or paucity of appropriate *in situ* monitoring systems and historical databases, given the high costs involved in developing and maintaining these systems and the inaccessibility of large sectors of these mountainous areas [5]. For those reasons, we applied methodologies that rely heavily on observations extracted from remote-sensing imagery to collect sample locations of areas that witness the development of debris flows (Fig. 3). The various remote-sensing datasets that were used to compile the debris-flow inventory are illustrated in Fig. 2 and described in Appendix I. The following characteristics were used to visually identify historical debris flows from these images.

1) Debris flows are composed largely of bare soils and rocks that appear bright in the visible wavelength region compared to their surrounding vegetated areas, which are dark in these wavelength regions.
2) Debris flows have specific morphological features that are identifiable from high-spatial-resolution imagery, topographical three-dimensional models [41], and radar images [42]. These morphological features include the presence of a torrent/gully/stream where surficial roughness is higher than in the surrounding area. The roughness was inferred from European remote-sensing satellite (ERS) and environmental satellite (Envisat) radar backscatter coefficient images. Other identifiable

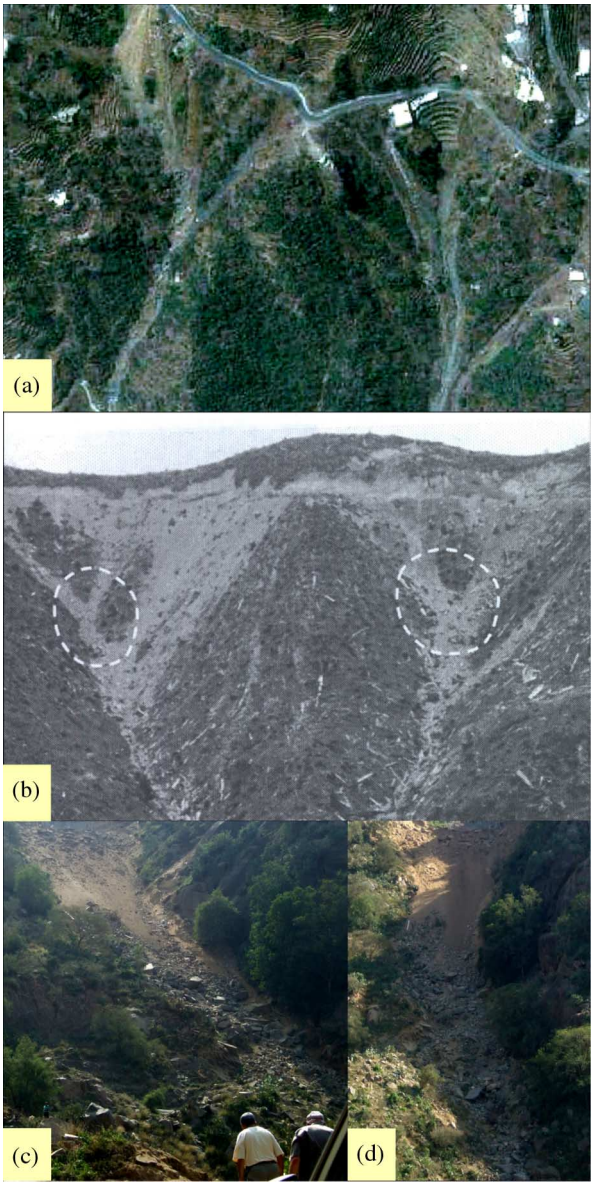


Fig. 3. (a) Satellite image (GeoEye) of debris flows in the study area. (b) Field shot of typical debris flows (modified from reference [51]). (c), (d) Field shots of debris flows from study area.

morphological features are the presence of deposition fans, channelized paths at steep mountains that have debris deposit downward, and bare breaks and scars [1].

3) The use of high-spatial-resolution imagery (i.e., 0.5 and 1 m spatial-resolution images) allowed us to distinguish even very small events.

Using these criteria, a total of 5100 debris-flow pixels and 5234 debris-flow-absent pixels were mapped in the study area. Out of the selected locations, a suite of 86 debris flows (equivalent to $\approx$430 pixels) were visited in the field for verification purposes. All of the field-investigated sites were successfully verified, giving us confidence in the applied methodology. An additional 55 (equivalent to $\approx$275 pixels) debris flows were confirmed from field reports [43]–[48].

Unfortunately, difficulties are involved in applying these methodologies to extract the distribution of debris flows across the entire study area. These include:

1) the time and effort needed to identify the large numbers (tens to hundreds of thousands) of debris flows across the study area [49];
2) the subjective element involved in visual identification of debris flows; and
3) the difficulty in identifying small flows and those that have been inactive for long time periods. With time, vegetation encroaches over the bare soils within the debris flows and makes it difficult to discriminate the spectrally bright flows from the surrounding dark vegetated areas.

C. Identification of Factors Controlling the Distribution of Debris Flows (Step III)

The variables that determine the debris-flow hazard of an area can be grouped into two general categories: 1) the intrinsic or static variables that contribute to debris-flow susceptibility, such as geology, slope gradient, slope aspect, long-term drainage patterns, elevation, and material cohesion; and 2) the dynamic or extrinsic variables that tend to trigger debris flows in an area of given susceptibility, such as heavy rainfall events and earthquakes [3], [5], [12], [50]. In this study, we are carrying on a debris-flow susceptibility investigation that is by definition the spatial analysis of the intrinsic factors; it expresses the likelihood of debris-flow occurrences at each spatial element [21].

As described earlier, intrinsic variables can be used as controlling factors in predicting future occurrences of debris flows in investigated areas [25]; however, there are no common guiding principles for selecting these variables [15]. In this study, the intrinsic factors were selected from among those most commonly used in literature to evaluate debris-flow susceptibility. Initial spatial analyses were conducted to ascertain whether each of these factors played a role in controlling the distribution of debris flows in the Jazan study area. The correspondence between the distribution of debris flows and various indexes suggested a causal effect.

The 10 most relevant controlling variables that were selected in the first stage are slope angle, slope aspect, topographic position index (TPI), stream power index (SPI), elevation, flow accumulation, topographic wetness index (TWI), normalized difference vegetation index (NDVI), distance to drainage line (DTDL), and soil weathering index (SWI). All 10 variables were established as thematic maps with a 10-m spatial resolution grid. The study area grid is 4871 rows by 6496 columns (i.e., total number is 31 642 016). Each thematic map and its relevance to the modeling are described as follows.

1) Slope Angle: Wide variations ($0^\circ - 84^\circ$) in slope angles are observed in the Jazan province [Fig. 4(b)]. We expect an increase of debris-flow occurrences with increase in slope angle up to a point where the steepness of the slope prohibits soil layer development and debris accumulation.

2) Slope Aspect: In the Jazan province, slope aspect generally trends in all directions [Fig. 4(c)]. It is commonly reported that the aspect is an indicator of exposition to preferential wind directions, precipitation regimes, sunlight impact, and discontinuity orientations [29], [40].

3) Normalized Difference Vegetation Index: Vegetation immobilizes a large amount of water and increases the shear

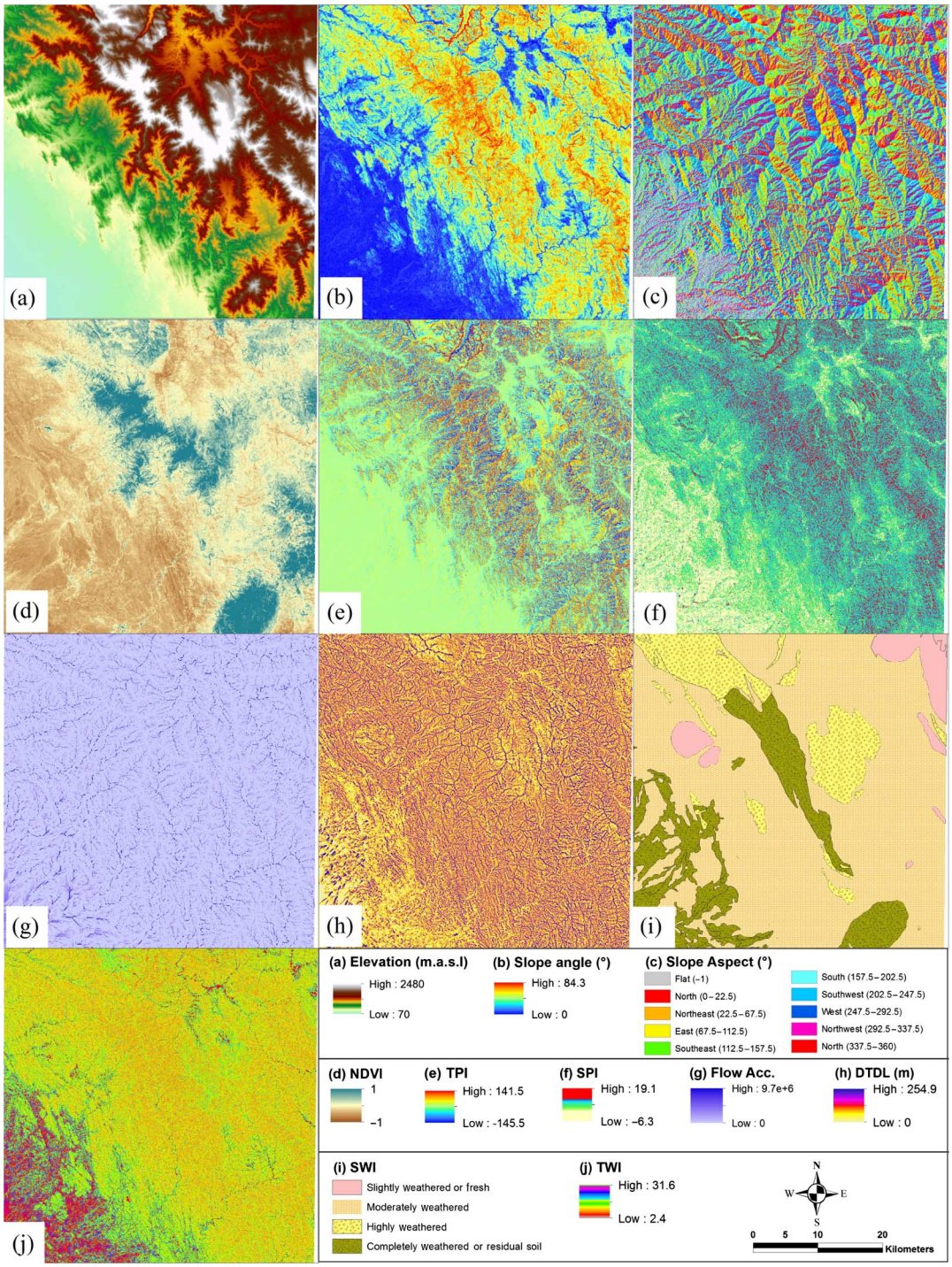


Fig. 4. Controlling variables maps used in this study: (a) elevation map; (b) slope angle map; (c) slope aspect map; (d) normalized difference vegetation index (NDVI) map; (e) topographic position index (TPI) map; (f) stream power index (SPI) map; (g) flow accumulation map; (h) distance to drainage line (DTDL) map; (i) soil weathering index (SWI) map; and (j) topographic wetness index (TWI) map. The extent of each map is the study area (from 42°45′00″E to 43°12′00″E and from 17°11′30″N to 17°41′30″N).

resistance and soil cohesion of the mass [39], [51]. The spatial variations in plant density across the study area were extracted from NDVI images, an image that is sensitive to the concentration of chlorophyll present at the surface [52]. NDVI images were calculated using the following equation:

$$\text{NDVI} = (NIR - R) / (NIR + R) \quad (1)$$

where *NIR* is the energy reflected in the near-infrared portion of the electromagnetic spectrum and *R* is the energy reflected in the red portion of the electromagnetic spectrum [53], [54]. A composite NDVI [Fig. 4(d)] was obtained by averaging the NDVI images extracted from Landsat ETM and TM temporal images (from December 06, 1994 to May 24, 2007). We averaged multiple NDVIs to capture the vegetative density of each pixel throughout the year instead of using a single NDVI

TABLE I
CONTINGENCY TEST SUCCESS PERCENTAGE FOR THE DIFFERENT TPI SCALES

	TPI 20 m	TPI 30 m	TPI 40 m	TPI 50 m	TPI 100 m	TPI 200 m	TPI 300 m	TPI 400 m	TPI 500 m
Contingency test success percentage	54.73 %	80.98 %	65.43 %	87.75 %	**88.71 %**	84.55 %	76.31 %	69.55 %	64.25 %

image that may represent spatial vegetation variations relative to a specific season of that year. The NDVI is dimensionless and it varies in the study area from -1 (barren land) to 1 (forest coverage).

4) Topographic Position Index: The TPI is an index that reflects the morphology of the topography [55]. Debris flows frequently occur in slope depressions, shallow ducts, and gullies that are characterized by terrain concavities [51]. The TPI image [Fig. 4(e)] was generated by applying the Jenness algorithm [56], which considers the surrounding cells up to a certain distance. We calculated the TPI using varying neighboring radiuses (range: $20-500$ m) and conducted a contingency test to identify the most suitable radius, which was found to be 100 m (Table I). The TPI is dimensionless; it ranges in the study area from -145.5 in valley bottoms to 141.5 in mountain hilltops.

5) Stream Power Index: SPI measures the erosive power of surface runoff [27]. An SPI thematic map [Fig. 4(f)] was generated using the following equation:

$$\mathrm{SPI} = \ln\left(A^* \tan\left(S\right)\right) \tag{2}$$

where *A* is the local upslope catchment area draining through a certain pixel in square meters and *S* is the local slope gradient in degrees [27], [57]. The SPI is dimensionless. In the study area, SPI reaches its minimum (-6.3) in the southwestern valley.

6) Elevation (Above Mean Sea Level): Elevation ranges between 70 and 2480 m in the study area [Fig. 4(a)]. Slope instability is believed to represent an environmental disturbance in lower and intermediate altitudinal ranges [55].

7) Flow Accumulation: The flow accumulation is used as a predisposing factor because it indicates the processes of water flow, convergence, and infiltration. The flow accumulation for the study area [Fig. 4(g)] was derived from the digital elevation model (DEM) using the 8-flow direction algorithm of Jensen and Domingue [58]. In the study area, the flow accumulation reaches its maximum (9.7×10^6 contributing pixels) at the outlet of the drainage basins (i.e., southwestern valley) and reaches its minimum at the uppermost pixels of the drainage basins (0 contributing pixel).

8) Distance to Drainage Line: Runoff plays an important role in the initiation and propagation of debris flows [21], [40]; therefore, the distance to drainage channels was investigated as being a potential indicator for the distribution of debris flows. The drainage network was delineated using standard stream delineation techniques (topographic parameterization (TOPAZ) technique; [59]), and the DTDL was computed as a continuous function quantifying the proximity of each pixel across the study area; it ranges between 0 and 255 m [Fig. 4(h)].

9) Soil Weathering Index: Lithological variations often contribute to differences in strength, weight, and permeability of rocks and soils [30]. A SWI map was generated using lithologic and textural information portrayed in geologic maps (e.g., [60]), topographic characteristics (e.g., slope image), and inferences from remote sensing to classify each pixel in the study area into one of four groups in accordance with the inferred levels of weathering and soil thicknesses [Fig. 4(i)]. Examples of these inferences include the following.

1) Highly vegetated areas (bright on NDVI images) were considered to have highly weathered thick soils [61], [62].
2) Foliated mafic rock units (e.g., amphibolite schist) on gentle slopes or flat areas are more likely to give rise to highly weathered thick soils than massive granitoids exposed on steep slopes.
3) Areas displaying spectral characteristics indicative of the presence of clay minerals were assumed to be highly weathered soils. Clay minerals have distinctive strong absorption features within the TM band 7 wavelength region ($2.08-2.35\,\mu\mathrm{m}$) and have higher reflectance values within TM band 5 wavelength region ($1.55-1.75\,\mu\mathrm{m}$) [63], [64]. Thus, soils rich in hydroxyl-bearing clay minerals can be readily mapped as bright areas on the Landsat TM band 5/7 ratio image [63].

10) Topographic Wetness Index: TWI is commonly used to quantify topographic control on hydrological processes [27], [55] and as an indicator for soil conditions and sediment and matter accumulation [55]. A TWI thematic map was generated [Fig. 4(j)] using the following equation:

$$\mathrm{TWI} = \ln\left(\frac{A}{\tan(S)}\right) \tag{3}$$

where *A* is the local upslope catchment area draining through a certain pixel in square meters and *S* is the local slope gradient in degrees [65]. The TWI is dimensionless. It reaches its maximum at the southwestern valley (31.6), where most of the sediments accumulate in the study area.

D. Debris-Flow Susceptibility Model Development and Evaluation (Step IV)

Two independent data mining methods, ANNs and LR, were applied. Our rationale for adopting these two methods is that they are based on two different algorithms; an agreement between obtained results from the two methodologies will provide strong support for the obtained results. This additional validation technique is particularly important in statistic-based approaches, because they are developed using available trends in the data and not explicitly based on physical processes [66]. The modeling process is summarized and illustrated in Fig. 5.

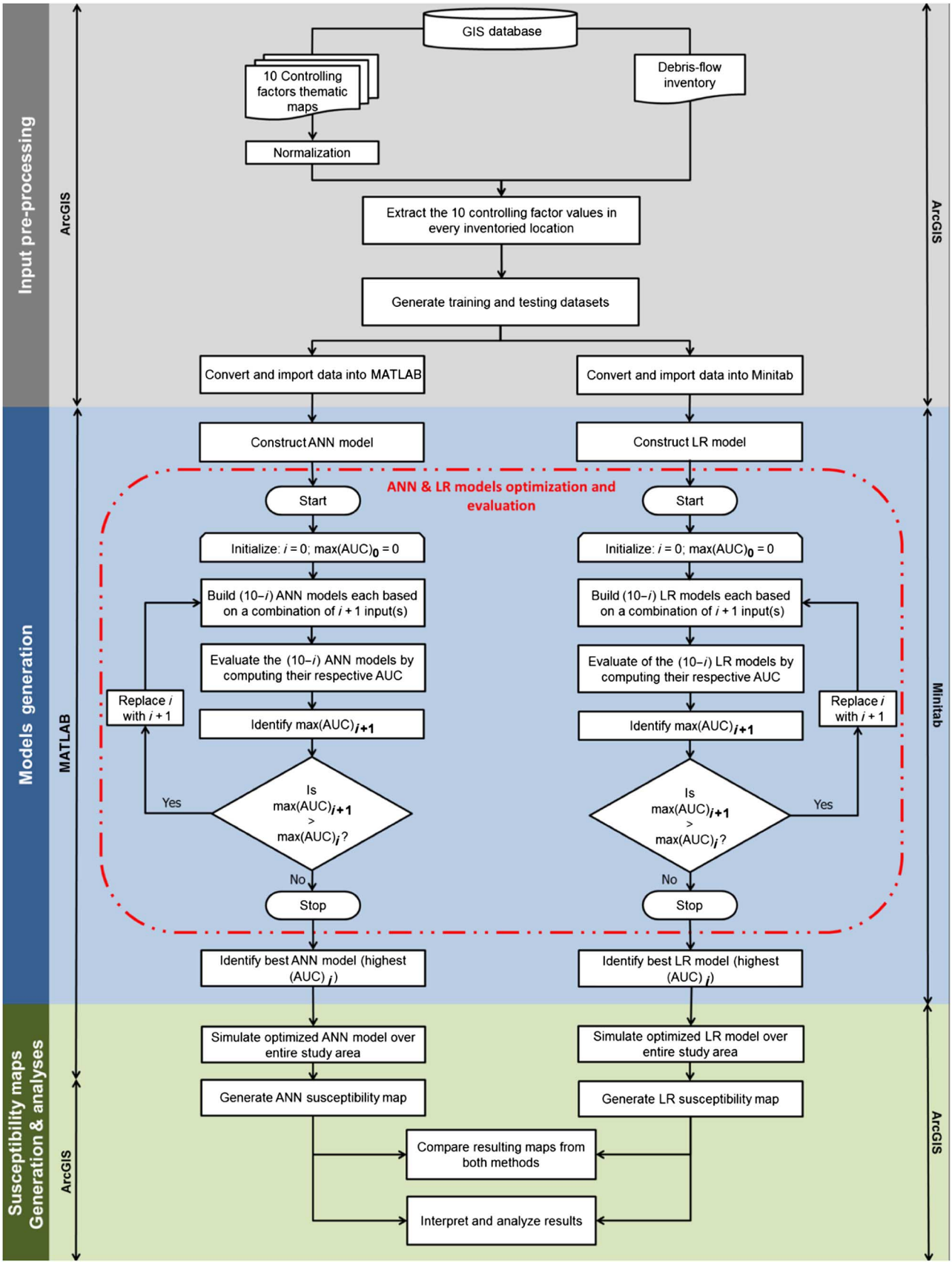


Fig. 5. Flowchart illustrating the modeling process.

1) Input Preprocessing for ANN and LR Models: Each controlling variable is traditionally segmented into classes based on the vision of the expert before inputting it into the model [25], [27], [67], [68]. These divisions, if applied, could introduce bias in the generated susceptibility/hazard maps, reduce data amount, and increase dependence on user knowledge [17]. None of the input variables were assigned classes, with one exception (i.e., SWI) that is a categorical variable.

Each of the input variables was normalized to the -1, 1 range using the following equation:

$$X_{i,n} = \frac{X_i - \left(\frac{X_{Max} + X_{Min}}{2}\right)}{\left(\frac{X_{Max} - X_{Min}}{2}\right)} \tag{4}$$

where $X_{i,n}$ denotes the normalized value for X_i, X_i represents each data point, X_{min} is the minima among the dataset, and X_{max} is the maxima among the dataset. The normalization was performed because the inputs come in varying units and display large variations in range and magnitude; if such variations were not accounted for, they could potentially affect the model outputs [7], [30], [69]. Normalization reduces cluster proliferation and allows successful implementation of ANN pattern recognition modeling in varying background levels [70].

For each inventoried location, we extracted the values of the normalized input variables (Fig. 5).

The available samples were portioned randomly into two subsets: training and testing. The former comprised 80% of the available samples (8267 samples) and the latter the remaining 20% (2067 samples). We used the training data subset to determine the models topology and the testing subset to compute the model performance and to find the optimum input combinations. For optimum results, all patterns (statistical properties) that are contained in the available data need to be represented in both the training and testing subsets [71]. Fig. 6 demonstrates similarities between the sample populations of each of the training and testing subgroups and the entire dataset.

2) Overview and Construction of ANN Model: ANN makes use of nonlinear and complex learning and prediction algorithms to extract the complex relationships among the various factors controlling debris-flow occurrences [67], [72]. In this study, a pattern-recognition neural-network module of MATLAB R2013a was utilized to uncover debris-flow patterns in the study area and conduct susceptibility analyses.

The adopted pattern-recognition neural-network architecture is based on a structure known as the multilayer perceptron (MLP). Fig. 7 provides a simplistic flowchart of our constructed ANN using the selected 10 controlling factors. The constructed model architecture and parameters are discussed in the following paragraphs.

We applied the widely used back-propagation learning algorithm in pattern-recognition modeling [73]–[75]. We used a set of examples for which both the input variables (10 controlling factor values) and the correct output values (presence or absence of debris flows) are known, weights (w) and biases (β) were assigned to the connections in a random manner, and comparisons were made between a random portion of the training samples calculated and expected outputs.

We adopted the *S*-shaped sigmoid transfer function that is widely used for the assessment of complex and nonlinear phenomena such as many natural hazards [27], [68], [76]–[78]. The sigmoid function constrains the outputs of a network between 0 and 1, which suits our case study of debris-flow susceptibility analyses where the desired output is represented by the presence or absence (1 or 0) of a debris flow.

The number of neurons in the hidden layer was calculated using the equation proposed by reference [79] and used by many scholars to assess natural hazards susceptibility (e.g., [11], [25]):

$$N_h = 2 * N_i + 1 \tag{5}$$

where N_h is the number of hidden neurons and N_i is the number of input neurons that is equal to the number of input controlling factors. We constructed multiple ANN models corresponding to various input combinations following a forward stepwise approach (refer to Section III-D4). Equation (5) was applied to each input combination. The hidden neuron numbers ranged from 3 (in the case of 1 input) to 21 (in the case of 10 inputs). The model with the best performance was selected (refer to Section IV-A).

3) Overview and Construction of LR Model: The LR model is a logarithmic model that describes a multivariate relationship between one or more predictor variables and a response variable. In the case of debris-flows research, the goal of LR is to find the best-fitting model to describe the relationship between debris-flow occurrences and a set of controlling variables [40], [80]. In our study, the response variable is binary. A binary logistic regression (BLR) can be expressed as follows:

$$\mathrm{Ln}\left(\frac{p}{1-p}\right) = c + a_1x_1 + a_2x_2 + \cdots + a_nx_n \tag{6}$$

$$p = \frac{\mathrm{e}^{c+a_1x_1+a_2x_2+\cdots+a_nx_n}}{1+\mathrm{e}^{c+a_1x_1+a_2x_2+\cdots+a_nx_n}} \tag{7}$$

where p is the probability of a debris-flow event occurring; c denotes the intercept of the equation; $x_1, x_2, \ldots, x_n$ denote the controlling variables; and $a_1, a_2, \ldots, a_n$ denote the coefficients that need to be estimated for the controlling variables using a maximum-likelihood optimization procedure [2], [81].

4) Model Optimization and Evaluation: The input selection task is generally not given enough attention in environmental modeling, and the model inputs are often selected on an *ad hoc* basis or using *a priori* knowledge of the targeted system. This can result in the inclusion of too few or too many model inputs, either of which is undesirable and may affect the model robustness [30], [71], [82]. We adopted a stepwise approach where inputs are systematically added, and every combination of inputs is then built into a model to be evaluated through ROC testing using the area under the ROC curve (AUC). The AUC is a widely used measure of the performance of a predictive rule [25], [27], [67], [83]–[85] (Fig. 5). The ROC is a curve connecting the sensitivity and the 1-specificity.

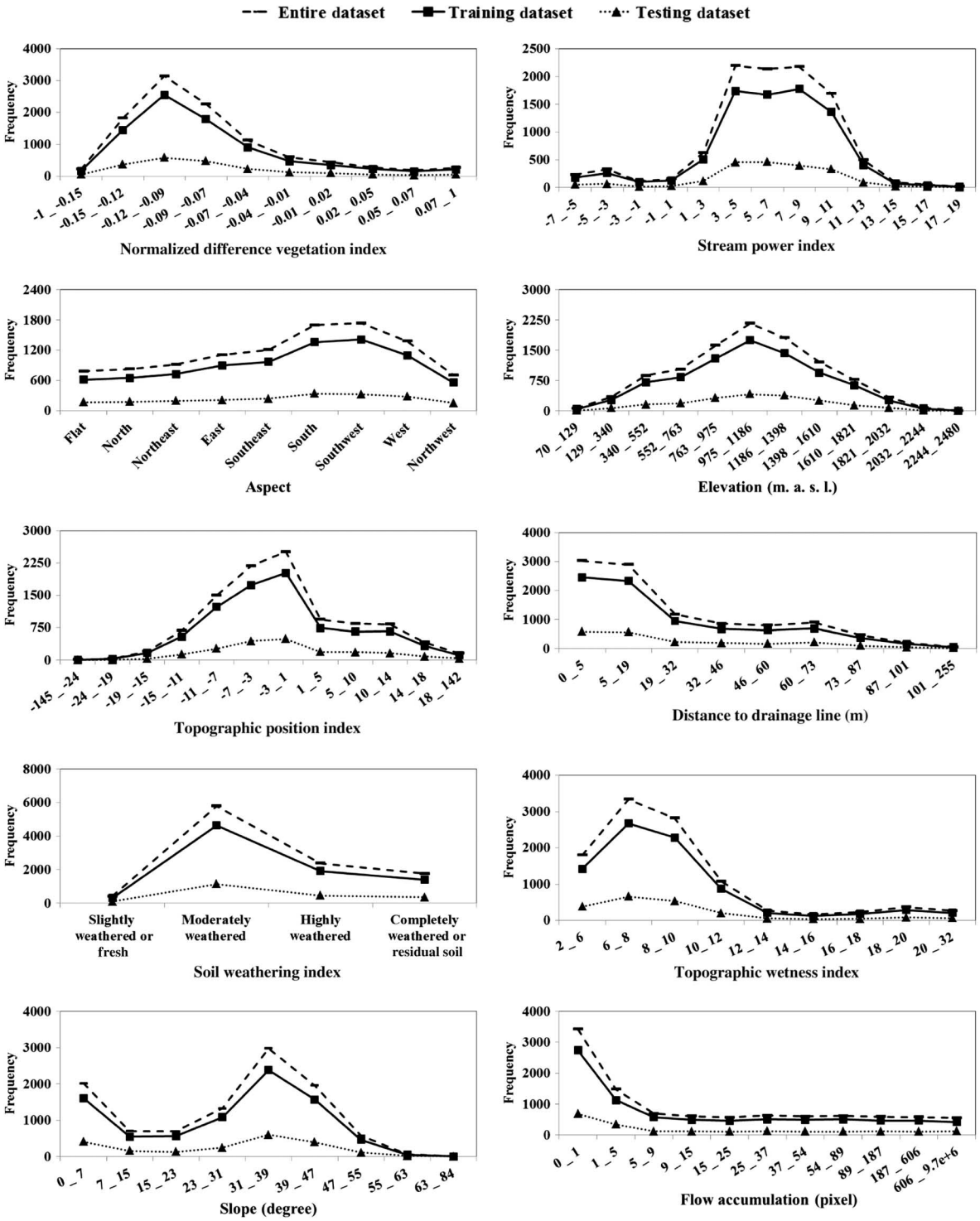


Fig. 6. Histograms for the entire dataset for the controlling factors and for the training and testing data subsets.

Sensitivity is the fraction of positive occurrences of debris flows that are correctly predicted, whereas 1-specificity is the fraction of incorrectly predicted cases that did not occur. The AUC characterizes the quality of a forecast system by describing the system's ability to anticipate the occurrence or nonoccurrence of predefined events correctly. A perfect model would have an AUC of 1.0 (100%); an AUC of 0.5 (50%) means the prediction is no better than a random guess [55].

We adopted a forward stepwise approach [71], [86], [87], in order to: 1) understand the most significant factors controlling the occurrence of debris flows and determine their order of importance for this environmental phenomenon; and 2) find the optimal combination of inputs that lead to the best predictive model. We added variables to the model (LR or ANN) one at a time, ran the model, and examined model outputs in comparison to test data. Initially, we ran the model 10 times using one of the 10 variables at a time, compared model

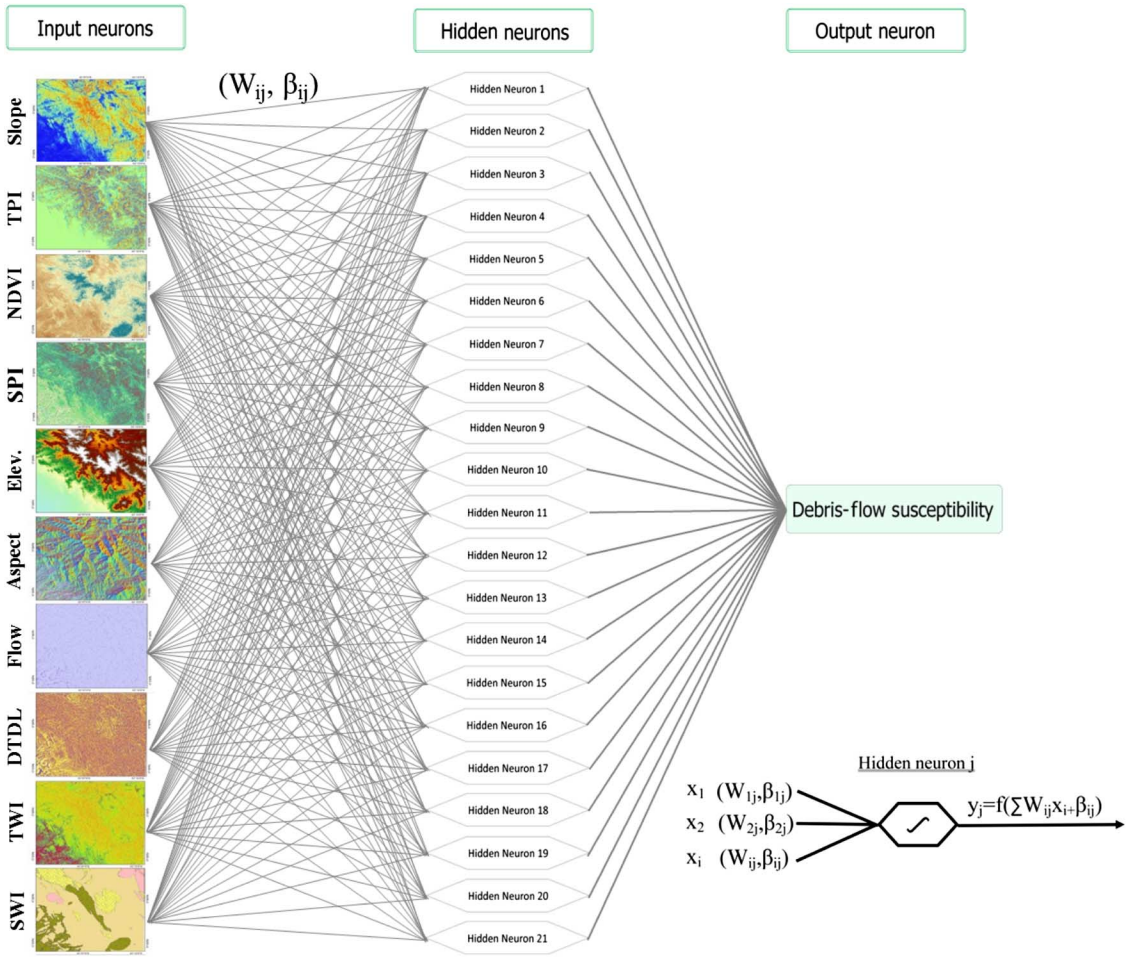


Fig. 7. Schematic diagram for the constructed pattern-recognition ANN. W_{ij} and β_{ij} refer to the weight and bias terms, respectively assigned between input neuron i and hidden neuron j during the learning procedure; f is the transfer function, x_i is the input from input neuron i, and y_j is the output corresponding to the hidden neuron j.

predictions to test data, and selected the variable that correctly predicted the largest number of test data. Using TPI alone, the LR and ANN prediction performances were 71.4% and 76.2%, respectively. Next, we constructed and ran nine models, this time using two variables: the TPI that was selected earlier, and one of the remaining nine variables at a time. We then selected the variable whose addition produced the largest increment in the model prediction performance. This variable was slope; its addition increased the obtained AUC values from 76.2% to 91.1% for the ANN and from 71.4% to 90.3% for the LR. The process of adding controlling variables was repeated until we found that the addition of remaining variables no longer improved or decreased the overall model performance (Fig. 5).

E. Generation of Susceptibility Maps (Step V)

The optimum and validated LR and ANN models were used to derive a susceptibility measure for each pixel in the study area.

1) The obtained coefficients corresponding to the LR model with the highest performance were plugged into (7) and applied to generate a final LR susceptibility map for the entire study area using the Spatial Analyst—Raster Calculator toolset of ArcGIS 10.1 software.
2) The trained ANN model with the highest performance (from Section III-D4) was used to generate the ANN final susceptibility map for the entire study area. Since the back-propagation ANN training algorithm presents difficulties when trying to retrieve the internal processes and coefficients of the procedure [78], the trained model was used to estimate a susceptibility measure for each pixel in the study area using the MATLAB R2013a software. The final susceptibility-derived matrix was plotted in ArcGIS 10.1 software for visualization and analyses.

The resulting susceptibility measures from both methods are in the form of debris-flow occurrence probabilities ranging from 0 to 1. The final maps delineate debris flows with a confidence level of 95%, corresponding to an occurrence probability equal to or higher than 0.95.

IV. Discussion and Findings

Our findings could be grouped in a number of categories: those related to the identification of the most significant controlling factors for debris-flow development in the study area (Section IV-A), and others related to the observations and information extracted from the susceptibility maps (Section IV-B). Our findings also have general applications (Section IV-C).

TABLE II
RESULTS OF THE VALIDATION TESTS FOR THE 45 ANN MODELS, REPRESENTED BY RESPECTIVE AUC VALUES

	1-variable models	2-variable models	3-variable models	4-variable models	5-variable models	6-variable models
TPI	**0.762**	–	–	–	–	–
NDVI	0.500	0.730	0.930	**0.967**	–	–
Flow	0.500	0.746	0.913	0.940	**0.961**	–
Elevation	0.500	0.776	0.927	0.955	0.940	0.954
Aspect	0.500	0.789	0.930	0.953	0.941	0.914
SWI	0.500	0.772	0.870	0.954	0.952	0.941
TWI	0.500	0.814	0.926	0.934	0.945	0.946
Slope	0.502	**0.911**	–	–	–	–
DTDL	0.500	0.770	**0.946**	–	–	–
SPI	0.609	0.817	0.940	0.926	0.933	0.956

TABLE III
RESULTS OF THE VALIDATION TESTS FOR THE 40 LR MODELS, REPRESENTED BY RESPECTIVE AUC VALUES

	1-variable models	2-variable models	3-variable models	4-variable models	5-variable models
TPI	**0.714**	–	–	–	–
NDVI	0.500	0.714	0.909	**0.963**	–
Flow	0.500	0.738	0.903	0.954	0.955
Elevation	0.500	0.743	0.904	0.959	0.958
Aspect	0.500	0.719	0.909	0.955	0.957
SWI	0.500	0.721	0.904	0.954	0.957
TWI	0.500	0.770	0.926	0.954	0.954
Slope	0.528	**0.903**	–	–	–
DTDL	0.500	0.708	**0.954**	–	–
SPI	0.567	0.664	0.909	0.954	0.955

A. Identification of Significant Controlling Factors and Model Validation Results

As described above, the forward stepwise models were applied up to a point where the inclusion of additional variables no longer produced significant improvements or decreased model accuracy. Forty-five ANN models and 40 LR models were constructed to reach that point. The validation test results (represented by AUC values) for the 45 ANN models are summarized in Table II and those for the 40 LR models are given in Table III. For illustration, Fig. 8(a) compares the ROC curves for the 1- and 5-variable (optimum) ANN-based models and Fig. 8(b) compares 1- and 4-variable (optimum) LR-based models.

Inspection of Tables II and III and Fig. 8 shows that for the ANN method, an accuracy of 96.1% in model prediction was achieved using five variables (TPI, slope, DTDL, NDVI, and flow accumulation), which correspond to 11 hidden neurons. An accuracy of 96.3% was achieved for the LR method using four variables (TPI, slope, DTDL, and NDVI). In both cases, inclusion of additional variables decreased model accuracy. Not only were the most significant four factors (TPI, slope, DTDL, and NDVI) identified by both approaches, but also they were ranked in the same order. The two models both have an AUC higher than 0.96, which is considered by references [27] and [88] to indicate excellent discrimination. The correspondence between the findings (type and rank of controlling variables) from both methods gives us confidence in the obtained results and suggests that the constructed ANN and LR models could be used to delineate areas prone to debris flows in the entire study area (Section IV-B). We found that the flow accumulation was important for optimum result in ANN but not important for LR, possibly because ANN picks all types of data distributions and patterns, even the smallest and less obvious fluctuations [89], [90]. However, it is noteworthy that the flow-accumulation contribution to the overall accuracy is only 1.1%.

B. Delineation of Areas Prone to Debris Flows

The generated LR prediction model is expressed by (7), where $X_1 = \text{TPI}$, $X_2 = \text{slope}$, $X_3 = \text{DTDL}$, $X_4 = \text{NDVI}$, $c = -15.2183$, $a_1 = -67.9145$, $a_2 = 14.1240$, $a_3 = -23.2804$, and $a_4 = -4.84220$. The two most commonly used measures of the goodness of fit in LR, namely the Pearson's chi-squared and Homster–Lemshow tests [91], showed perfect results (Pearson: 1; Homster–Lemshow: 1).

Inspection of the outputs of the LR model indicated that the TPI, the DTDL, and the NDVI variables are inversely correlated with the debris-flow occurrence probability, whereas the slope is positively correlated. These patterns are in agreement with our conceptual model; debris flows are more prone to develop in sparsely vegetated areas that have low root pressure and low cohesion; in gullies and concavities; in mid-slope ranges; and close to drainage lines.

We compared the ANN and LR susceptibility maps [Fig. 9(a) and (d)] in ArcGIS by generating a difference image. We observed good correspondence (91.5%) between the two susceptibility images. The remaining 8.5% are all pixels located at the peripheries of the identified debris flows. The two methods picked the same individual debris-flows, but we found minor mismatches at their borders. This high correspondence between the outcomes of the two methods reinforces the validity of the extracted debris-flow susceptibility results in the region and highlights the potential for using such maps for urbanization planning and civil protection services. In this respect, we cite a few applications for the study area.

The susceptibility maps indicate that over 13% of the total pixels were identified as being prone to debris flows. Spatial correlations of the distribution of these picture elements with that of the roads and buildings (from remote-sensing-based land use maps) reveal that the delineated debris flows intersect with the roads at 1492 locations. In addition, 278 buildings (3% of the total buildings in the study area) are built within the debris-flow prone areas. Those intersections and buildings will be the first to experience damages in the case of a triggering event (i.e., extreme precipitation or earthquake). The susceptibility maps could potentially be used to identify: 1) hazardous zones where building permits should be denied; and 2) locations for the construction of mitigation designs such as retaining and catchment structures that can be built at the edge of the roads to minimize the damage at those

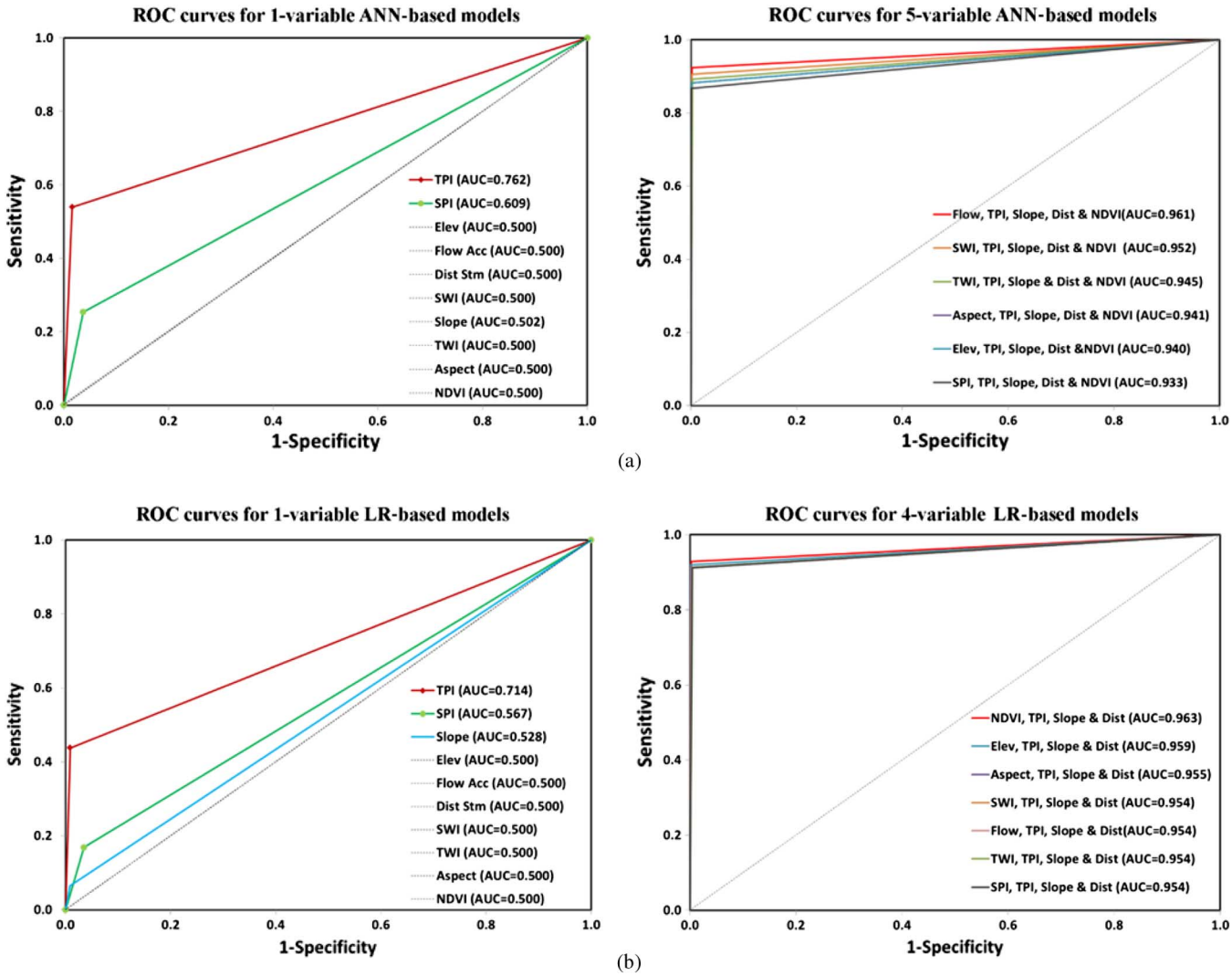


Fig. 8. (a) Comparative ROC curves in the case of 1-variable ANN-based models and 5-variable (optimum) ANN-based models. (b) Comparative ROC curves in the case of 1-variable LR-based models and 4-variable (optimum) LR-based models.

locations. Inspection of the susceptibility maps reveals several locations that are susceptible to debris-flow development, yet our inspection of high-resolution satellite images did not show evidence for the existence of active debris-flows in these areas. Such areas could potentially be the location of future debris flows.

Inspection of the susceptibility maps reveals dense debris flows in areas that coincide with the distribution of the highly elevated Baish group ridge [Fig. 9(a) and (d); area I]. This is likely related to the fact that the lithological group is largely composed of weathered metabasalt that is susceptible to erosion and mobilization by fluvial processes [60]. These fluvial processes are enhanced at higher elevations where precipitation is generally more pronounced [92]. The susceptibility maps also showed that debris flows develop preferentially (36% of the total delineated debris flows) on the west- and southwest-facing slopes. These slope directions are more exposed (compared to all eight remaining aspect directions) to the Indian monsoons, the main source of precipitation in the study area. We suggest that the regional predisposing factors (i.e., elevation, geology, and aspect) exert more control on the distribution of debris flows on a regional scale, whereas the local factors (TPI, slope, NDVI, and DTDL) are better parameters for a detailed delineation of debris flows.

C. Regional and Global Implications

We presented susceptibility analyses that were conducted through the integration of readily available remote-sensing data, GIS techniques, and data mining with ANN and LR modeling approaches. Application of the developed methodologies in the Red Sea hills and elsewhere worldwide will lead to significant improvements in assessing and understanding the nature and scope of ground failure problems and will contribute to building safe and sustainable communities around the world. These methods are not a substitute for traditional approaches that rely on extensive field data, but they are particularly useful in inaccessible areas and those lacking adequate monitoring systems. Remote-sensing datasets are nearly the only source of consistent and high-resolution observations over remote and rugged mountainous regions where debris flows typically occur.

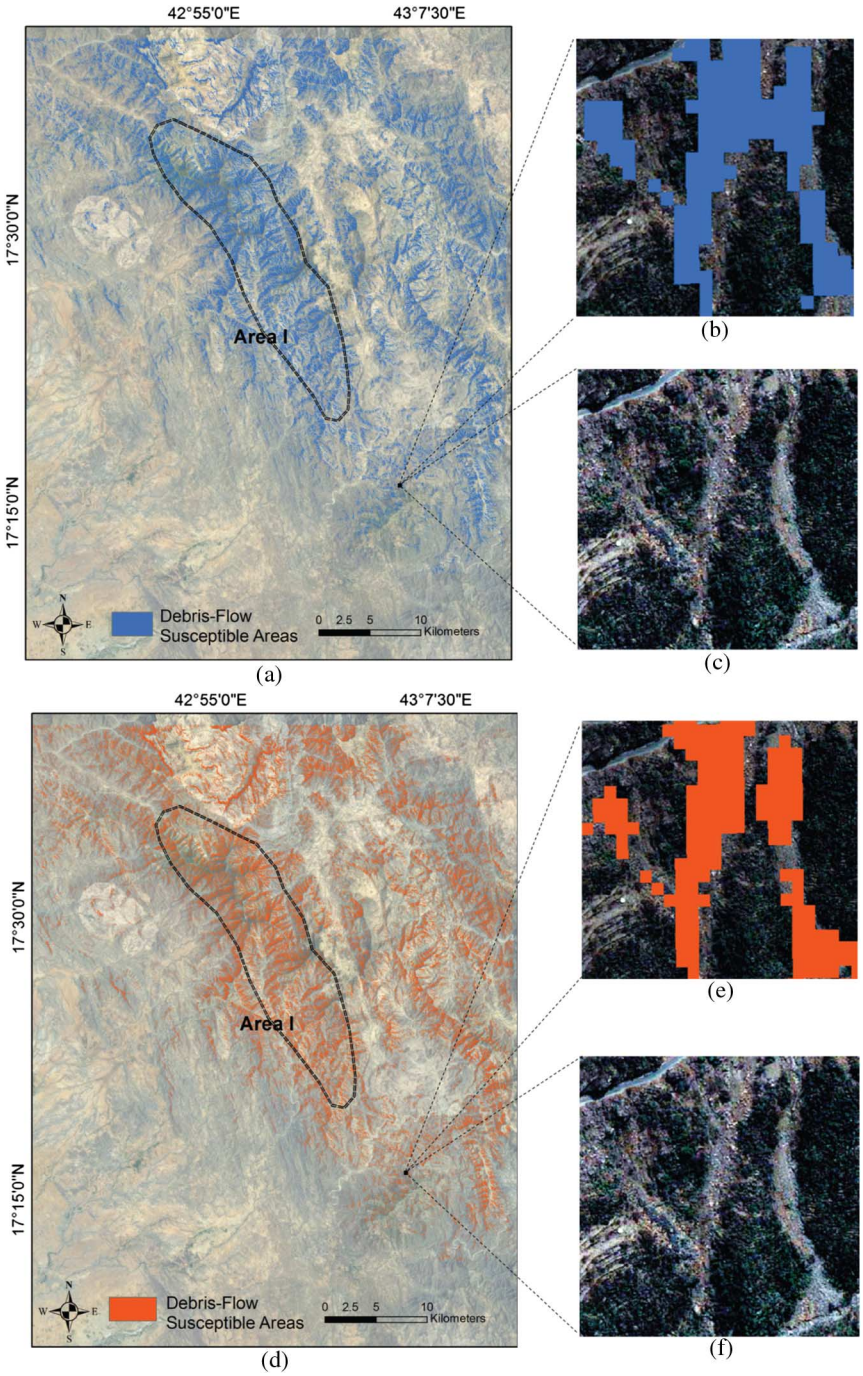


Fig. 9. (a) Generated debris-flow susceptibility map using ANN. (b) Enlargement of the boxed area in Fig. 9a. (c) Same as Fig. 9b, but with the modeled debris flows omitted. Note the correspondence between the modeled (Fig. 9b: blue areas) and observed (Fig. 9c: bright areas) debris flows. (d) Generated debris-flow susceptibility map using LR. (e) Enlargement of the boxed area in Fig. 9d. (f) Same as Fig. 9e, but with the modeled debris flows omitted. Note the correspondence between the modeled (Fig. 9e: orange areas) and observed (Fig. 9f: bright areas) debris flows. Area I displays dense debris-flow distribution. This area coincides with the highly elevated and erodible Baish lithological group.

ACKNOWLEDGMENT

The authors would like to thank the Saudi Geological Survey for providing scientific, financial, and logistical support for this project. The authors also acknowledge the European Space Agency (ESA) for providing Envisat and ERS datasets through the accepted Cat-1 proposal #8102. The authors would also like to thank Mr. C. Manche for help in preparing the manuscript.

APPENDIX I

Detailed List Of Used Datasets

1) Orbview-3 mosaic (1 m spatial resolution).
2) GeoEye mosaic (0.5 m spatial resolution).
3) Nine Landsat TM and ETM false-color composites of band and ratio images (spatial resolution: 15 m and/or 30 m; acquisition period: 1994–2007).

4) Satellite pour l'observation de la terre (SPOT) 5 false-color composite (spatial resolution: 2.5 and 5 m).
5) Temporal normalized difference vegetation index (NDVI) maps computed from nine Landsat ETM and TM (spatial resolution: 15 m and/or 30 m; acquisition period: 1994–2007).
6) Advanced spaceborne thermal emission and reflection radiometer (ASTER) false-color composites (spatial resolution: 15 m).
7) Eighteen European remote-sensing satellite (ERS) and environmental satellite (Envisat) radar images and derived backscatter coefficient images (spatial resolution: 25 m; acquisition period: 1996–2009).
8) Distribution of faults/fractures extracted from geologic maps, radar, Landsat ETM/TM, and ASTER images.
9) Geologic maps (scale: 1/100 000, 1/250 000, and 1/4 000 000) provided by the Saudi Geological Survey (SGS).
10) Lithologic maps derived from geological maps and observations extracted from remote-sensing data.
11) Soil weathering index (SWI) map derived mainly from the lithological maps (spatial resolution: 10 m).
12) Digital elevation model (DEM) from Shuttle radar topography mission (SRTM) (spatial resolution: 90 m).
13) DEM from ASTER imagery (spatial resolution: 30 m).
14) Topographic maps (contour interval: 100, 50, 30, 15, 10, and 5 m) and global positioning system (GPS) points provided by SGS.
15) DEM derived from SGS topographic contours and GPS points using triangulated irregular network data structures and calibrated using SRTM and ASTER DEMs (spatial resolution: 10 m).
16) Three-dimensional surface model based on the 10 m spatial resolution DEM.
17) Slope angle map derived from DEM (spatial resolution: 10 m).
18) Slope aspect map derived from DEM (spatial resolution: 10 m).
19) DEM hillshade map derived from DEM (spatial resolution: 10 m).
20) Topographic position index (TPI) map derived using the Jenness algorithm (neighboring radius: 20, 30, 40, 50, 100, 200, 300, 400, and 500 m; spatial resolution: 10 m).
21) Flow accumulation map derived from DEM (spatial resolution: 10 m).
22) Stream distribution map derived from DEM (spatial resolution: 10 m).
23) Climatic stations distribution map.
24) Distance to drainage line (DTDL) map derived from the generated stream network (spatial resolution: 10 m).
25) Stream power index (SPI) map (spatial resolution: 10 m).
26) Topographic wetness index (TWI) map (spatial resolution: 10 m).
27) Road distribution map digitized from high-spatial-resolution imagery (GeoEye).
28) Building distribution map digitized from high-spatial-resolution imagery (GeoEye).

REFERENCES

[1] G. Chevalier, V. Medina, M. Hürlimann, and A. Bateman, "Debris-flow susceptibility analysis using fluvio-morphological parameters and data mining: Application to the Central-Eastern Pyrenees," *Nat. Hazards*, vol. 67, pp. 1–26, 2013.

[2] W. Xu *et al.*, "Debris flow prediction models based on environmental factors and susceptible subarea classification in Sichuan, China," *Nat. Hazards*, vol. 67, pp. 869–878, Jun. 2013.

[3] J. Chiu, C. Dow, C. Lin, J. Lin, and H. Hsieh, "A watershed-based debris flow early warning system using sensor web enabling techniques in heterogeneous environments," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 6, pp. 1729–1739, Dec. 2012.

[4] H. Lee, A. Banerjee, Y. Fang, B. Lee, and C. King, "Design of a multifunctional wireless sensor for *in-situ* monitoring of debris flows," *IEEE Trans. Instrum. Meas.*, vol. 59, no. 11, pp. 2958–2967, Nov. 2010.

[5] Y. Hong, R. F. Adler, and G. Huffman, "An experimental global prediction system for rainfall-triggered landslides using satellite remote sensing and geospatial datasets," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 6, pp. 1671–1680, Jun. 2007.

[6] S. Chiang, K. Chang, A. C. Mondini, B. Tsai, and C. Chen, "Simulation of event-based landslides and debris flows at watershed level," *Geomorphology*, vol. 138, pp. 306–318, 2012.

[7] W. Liang, D. Zhuang, D. Jiang, J. Pan, and H. Ren, "Assessment of debris flow hazards using a Bayesian network," *Geomorphology*, vol. 171, pp. 94–100, Oct. 2012.

[8] R. K. Dahal *et al.*, "GIS-based weights-of-evidence modelling of rainfall-induced landslides in small catchments for landslide susceptibility mapping," *Environ. Geol.*, vol. 54, pp. 311–324, 2008.

[9] F. Dai, C. Lee, and Y. Y. Ngai, "Landslide risk assessment and management: An overview," *Eng. Geol.*, vol. 64, pp. 65–87, 2002.

[10] F. Dai, C. Lee, J. Li, and Z. Xu, "Assessment of landslide susceptibility on the natural terrain of Lantau Island, Hong Kong," *Environ. Geol.*, vol. 40, pp. 381–391, 2001.

[11] I. Yilmaz, "Landslide susceptibility mapping using frequency ratio, logistic regression, artificial neural networks and their comparison: A case study from Kat landslides (Tokat–Turkey)," *Comput. Geosci.*, vol. 35, pp. 1125–1138, 2009.

[12] K. Muthu, M. Petrou, C. Tarantino, and P. Blonda, "Landslide possibility mapping using fuzzy approaches," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 4, pp. 1253–1265, Apr. 2008.

[13] R. Pack, D. Tarboton, and C. Goodwin, "The SINMAP approach to terrain stability mapping," in *Proc. 8th Congr. Int. Assoc. Eng. Geol.*, Vancouver, BC, Canada, 1998, pp. 21-25.

[14] P. E. Miller *et al.*, "A remote sensing approach for landslide hazard assessment on engineered slopes," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 4, pp. 1048–1056, Apr. 2012.

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

[16] A. Yalcin, "GIS-based landslide susceptibility mapping using analytical hierarchy process and bivariate statistics in Ardesen (Turkey): Comparisons of results and confirmations," *Catena*, vol. 72, pp. 1–12, 2008.

[17] M. L. Suzen and V. Doyuran, "Data driven bivariate landslide susceptibility assessment using geographical information systems: A method and application to Asarsuyu catchment, Turkey," *Eng. Geol.*, vol. 71, pp. 303–321, Feb. 2004.

[18] J. L. Naranjo, C. J. van Westen, and R. Soeters, "Evaluating the use of training areas in bivariate statistical landslide hazard analysis-A case study in Colombia," *ITC J.*, vol. 3, pp. 292–300, 1994.

[19] J. Choi, H. Oh, H. Lee, C. Lee, and S. Lee, "Combining landslide susceptibility maps obtained from frequency ratio, logistic regression, and artificial neural network models using ASTER images and GIS," *Eng. Geol.*, vol. 124, pp. 12–23, 2012.

[20] E. A. Sezer, B. Pradhan, and C. Gokceoglu, "Manifestation of an adaptive neuro-fuzzy model on landslide susceptibility mapping: Klang valley, Malaysia," *Expert Syst. Appl.*, vol. 38, pp. 8208–8219, 2011.

[21] B. Pradhan, E. A. Sezer, C. Gokceoglu, and M. F. Buchroithner, "Landslide susceptibility mapping by neuro-fuzzy approach in a landslide-prone area (Cameron Highlands, Malaysia)," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 12, pp. 4164–4177, Dec. 2010.

[22] D. Kawabata and J. Bandibas, "Landslide susceptibility mapping using geological data, a DEM from ASTER images and an artificial neural network (ANN)," *Geomorphology*, vol. 113, pp. 97–109, 2009.

[23] A. Akgun, E. A. Sezer, H. A. Nefeslioglu, C. Gokceoglu, and B. Pradhan, "An easy-to-use MATLAB program (MamLand) for the assessment of landslide susceptibility using a Mamdani fuzzy algorithm," *Comput. Geosci.*, vol. 38, pp. 23–34, 2012.

[24] C. Xu, X. Xu, Q. Yao, and Y. Wang, "GIS-based bivariate statistical modelling for earthquake-triggered landslides susceptibility mapping related to the 2008 Wenchuan earthquake, China," *Quart. J. Eng. Geol. Hydrogeol.*, vol. 46, pp. 221–236, 2013.

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

[26] Á. M. Felicísimo, A. Cuartero, J. Remondo, and E. Quirós, "Mapping landslide susceptibility with logistic regression, multiple adaptive regression splines, classification and regression trees, and maximum entropy methods: A comparative study," *Landslides*, vol. 10, pp. 1–15, 2013.

[27] S. Park, C. Choi, B. Kim, and J. Kim, "Landslide susceptibility mapping using frequency ratio, analytic hierarchy process, logistic regression, and artificial neural network methods at the Inje area, Korea," *Environ. Earth Sci.*, vol. 68, pp. 1443–1464, 2013.

[28] A. Akgun, "A comparison of landslide susceptibility maps produced by logistic regression, multi-criteria decision, and likelihood ratio methods: A case study at İzmir, Turkey," *Landslides*, vol. 9, pp. 93–106, 2012.

[29] H. Quan and B. Lee, "GIS-based landslide susceptibility mapping using analytic hierarchy process and artificial neural network in Jeju (Korea)," *KSCE J. Civil Eng.*, vol. 16, pp. 1258–1266, 2012.

[30] C. Xu, X. Xu, F. Dai, and A. K. Saraf, "Comparison of different models for susceptibility mapping of earthquake triggered landslides related with the 2008 Wenchuan earthquake in China," *Comput. Geosci.*, vol. 46, pp. 317–329, 2012.

[31] National Meteorology & Environment Center, *Surface Climatological Report*. Saudi Arabia: Ministry of Defense & Environment Protection, National Meteorology & Environment Center, 2013.

[32] M. S. Alyamani and Z. Sen, "Regional variation of monthly rainfall amounts in the Kingdom of Saudi Arabia," *J. King Abdulaziz Univ. Earth Sci.*, vol. 6, pp. 113–133, 1992.

[33] D. B. Stoeser and V. E. Camp, "Pan-African microplate accretion of the Arabian shield," *Geol. Soc. Amer. Bull.*, vol. 96, pp. 817–826, 1985.

[34] W. R. Greenwood, *Late Proterozoic Island-Arc Complexes and Tectonic Belts in the Southern Part of the Arabian Shield*. Riyadh, Saudi Arabia: Ministry of Petroleum and Mineral Resources, 1982.

[35] E. Wolfenden, C. Ebinger, G. Yirgu, A. Deino, and D. Ayalew, "Evolution of the northern Main Ethiopian rift: Birth of a triple junction," *Earth Planet. Sci. Lett.*, vol. 224, pp. 213–228, 2004.

[36] R. Agar, "The Najd fault system revisited: a two-way strike-slip orogen in the Saudi Arabian shield," *J. Struct. Geol.*, vol. 9, pp. 41–48, 1987.

[37] R. Coleman, *Reconnaissance Geology of the Khamis Mushayt Quadrangle*. Riyadh, Saudi Arabia: Saudi Arabian Directorate General Mineral Resources Geologic Map GM-5, Scale, 1973, vol. 1.

[38] Earth Science Remote Sensing Facility (ESRS). (2014). Online Map Databases, Saudi [Online]. Available: http://www.esrs.wmich.edu/webmap

[39] D. J. Varnes, *Landslide Hazard Zonation: A Review of Principles and Practice*. Paris, France: International Association Engineering Geology, Commission on Landslides and Other Mass Movements on Slopes, 1984.

[40] K. C. Devkota *et al.*, "Landslide susceptibility mapping using certainty factor, index of entropy and logistic regression models in GIS and their comparison at Mugling-Narayanghat road section in Nepal Himalaya," *Nat. Hazards*, vol. 65, pp. 135–165, 2013.

[41] K. Tsutsui *et al.*, "Detection and volume estimation of large-scale landslides based on elevation-change analysis using DEMs extracted from high-resolution satellite stereo imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 6, pp. 1681–1696, Jun. 2007.

[42] S. Voigt *et al.*, "Satellite image analysis for disaster and crisis-management support," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 6, pp. 1520–1528, Jun. 2007.

[43] Saudi Geological Survey (SGS), *Landslide Field Investigations: Mobtaat Road*. Jeddah, Saudi Arabia: Saudi Geological Survey, 2011.

[44] Saudi Geological Survey (SGS), *Landslide Field Investigations: 20th Road*. Jeddah, Saudi Arabia: Saudi Geological Survey, 2011.

[45] Saudi Geological Survey (SGS), *Landslide Field Investigations: Doha Road*. Jeddah, Saudi Arabia: Saudi Geological Survey, 2011.

[46] Saudi Geological Survey (SGS), *Landslide Field Investigations: 12th Road*. Jeddah, Saudi Arabia: Saudi Geological Survey, 2011.

[47] Saudi Geological Survey (SGS), *Landslide Field Investigations: Serba Road*. Jeddah, Saudi Arabia: Saudi Geological Survey, 2011.

[48] Saudi Geological Survey (SGS), *Landslide Field Investigations: 8th Road*. Jeddah, Saudi Arabia: Saudi Geological Survey, 2011.

[49] W. Yang, M. Wang, and P. Shi, "Using MODIS NDVI time series to identify geographic patterns of landslides in vegetated regions," *IEEE Geosci Remote Sens. Lett.*, vol. 10, no. 4, pp. 707–710, Jul. 2013.

[50] F. Chang, Y. Chiang, and W. Lee, "Investigating the impact of the Chi-Chi earthquake on the occurrence of debris flows using artificial neural networks," *Hydrol. Process.*, vol. 23, pp. 2728–2736, 2009.

[51] R. C. Sidle, and H. Ochiai, *Landslides: Processes, Prediction, and Land use*. Washington, DC, USA: American Geophysical Union, 2006.

[52] J. Rau, L. Chen, J. Liu, and T. Wu, "Dynamics monitoring and disaster assessment for watershed management using time-series satellite images," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, pp. 1641–1649, 2007.

[53] J. Vauhkonen *et al.*, "Classification of spruce and pine trees using active hyperspectral LiDAR," *IEEE Geosci Remote Sens. Lett.*, vol. *10*, no. 5, pp. 1138–1141, Sep. 2013.

[54] C. J. Tucker, "Red and photographic infrared linear combinations for monitoring vegetation," *Remote Sens. Environ.*, vol. 8, pp. 127–150, 1979.

[55] P. Vorpahl, H. Elsenbeer, M. Märker, and B. Schröder, "How can statistical models help to determine driving factors of landslides?" *Ecol. Model.*, vol. 239, pp. 27–39, 2012.

[56] J. Jenness, *Topographic Position Index (tpi_jen. Avx) Extension for ArcView 3.x, v.1.3 a. Jenness Enterprises*, 2006 [Online]. Available: http://www.jennessent.com/arcview/tpi/htm

[57] I. D. Moore and R. B. Grayson, "Terrain-based catchment partitioning and runoff prediction using vector elevation data," *Water Resour. Res.*, vol. 27, pp. 1177–1191, 1991.

[58] S. Jenson and J. Domingue, "Extracting topographic structure from digital elevation data for geographic information system analysis," *Photogramm. Eng. Remote Sensing*, vol. 54, pp. 1593–1600, 1988.

[59] J. Garbrecht and J. Campbell, *An Automated Digital Landscape Analysis Tool for Topographic Evaluation, Drainage Identification, Watershed Segmentation and Subcatchment Parameterization: TOPAZ User Manual*. El Reno, OK, USA: USDA, Agricultural Res. Service, Grazinglands Res. Lab., 1997.

[60] G. Fairer, F. J. Fuller, and S. Arabia, *Geology of the Wadi Baysh Quadrangle, Sheet 17F, Kingdom of Saudi Arabia*. Riyadh, Saudi Arabia: Ministry of Petroleum and Mineral Resources, Deputy Ministry for Mineral Resources, 1983.

[61] P. Guyer, *Introduction to Identification and Classification of Soil and Rock*. Stony Point, New York, USA: Continuing Education and Development Inc., pp. 1–37, 2010.

[62] E. T. Brown, *Rock characterization, testing & monitoring: ISRM suggested methods*. Oxford, U.K.: Commission on Testing methods, International Society for Rock Mechanics by Pergamon Press, 1981.

[63] M. Sultan, R. E. Arvidson, N. C. Sturchio, and E. A. Guinness, "Lithologic mapping in arid regions with Landsat thematic mapper data: Meatiq dome, Egypt," *Geol. Soc. Amer. Bull.*, vol. 99, pp. 748–762, 1987.

[64] M. J. Abrams, R. P. Ashley, L. C. Rowan, A. F. Goetz, and A. B. Kahle, "Mapping of hydrothermal alteration in the Cuprite mining district, Nevada, using aircraft scanner images for the spectral region 0.46 to 2.36 μm," *Geology*, vol. 5, pp. 713–718, 1977.

[65] R. Sörensen, U. Zinko, and J. Seibert, "On the calculation of the topographic wetness index: Evaluation of different methods based on field observations," *Hydrol. Earth Syst. Sci. Discuss.*, vol. 10, pp. 101–112, 2006.

[66] G. P. Zhang, "Avoiding pitfalls in neural network research," *IEEE Trans. Syst., Man, Cybern. C. Appl. Rev.*, vol. 37, pp. 3–16, 2007.

[67] S. Pascale *et al.*, "Landslide susceptibility mapping using artificial neural network in the urban area of senise and san costantino albanese (Basilicata, Southern Italy)," in *Computational Science and Its Applications-ICCSA 2013*. New York, NY, USA: Springer, 2013, pp. 473-488.

[68] M. Zare, H. R. Pourghasemi, M. Vafakhah, and B. Pradhan, "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.*, pp. 1–16, 2012.
[69] Y. Li, G. Chen, C. Tang, G. Zhou, and L. Zheng, "Rainfall and earthquake-induced landslide susceptibility assessment using GIS and artificial neural network," *Nat. Hazards Earth Syst. Sci.*, vol. 12, pp. 2719–2729, 2012.
[70] W. S. Sarle, "Neural networks and statistical models," in *Proc. Int. Conf. 19th Annu. SAS Users Group*, 1994, pp. 1538–1550.
[71] H. R. Maier, A. Jain, G. C. Dandy, and K. P. Sudheer, "Methods used for the development of neural networks for the prediction of water resource variables in river systems: Current status and future directions," *Environ. Model. Softw.*, vol. 25, pp. 891–909, 2010.
[72] Y. Chen, P. Ni, J. Chen, and S. Hsieh, "The application of remote sensing technology to the interpretation of land use for rainfall-induced landslides based on genetic algorithms and artificial neural networks," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 2, no. 2, pp. 87–95, 2009.
[73] C. Melchiorre, M. Matteucci, A. Azzoni, and A. Zanchi, "Artificial neural networks and cluster analysis in landslide susceptibility zonation," *Geomorphology*, vol. 94, pp. 379–400, 2008.
[74] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, *Learning Internal Representations by Error Propagation*. California, USA: California Univ. San Diego, La Jolla inst for Cognitive science, 1985.
[75] C. M. Bishop, *Neural Networks for Pattern Recognition*. London, U.K.: Oxford Univ. Press, 1995.
[76] H. Gomez and T. Kavzoglu, "Assessment of shallow landslide susceptibility using artificial neural networks in Jabonosa River Basin, Venezuela," *Eng. Geol.*, vol. 78, pp. 11–27, 2005.
[77] D. Kanungo, M. Arora, S. Sarkar, and R. Gupta, "A comparative study of conventional, ANN black box, fuzzy and combined neural and fuzzy weighting procedures for landslide susceptibility zonation in Darjeeling Himalayas," *Eng. Geol.*, vol. 85, pp. 347–366, 2006.
[78] M. B. Kia *et al.*, "An artificial neural network model for flood simulation using GIS: Johor River Basin, Malaysia," *Environ. Earth Sci.*, vol. 67, pp. 251–264, 2012.
[79] R. Hecht-Nielsen, *Neurocomputing*. Reading, MA, USA: Addison-Wisley, 1990.
[80] C. Xu, X. Xu, F. Dai, and A. K. Saraf, "Comparison of different models for susceptibility mapping of earthquake triggered landslides related with the 2008 Wenchuan earthquake in China," *Comput. Geosci.*, vol. 46, pp. 317–329, 2012.
[81] M. Amini, K. C. Abbaspour, and C. A. Johnson, "A comparison of different rule-based statistical models for modeling geogenic groundwater contamination," *Environ. Model. Softw.*, vol. 25, pp. 1650–1657, 2010.
[82] H. R. Maier and G. C. Dandy, "Neural networks for the prediction and forecasting of water resources variables: A review of modelling issues and applications," *Environ. Model. Softw.*, vol. 15, pp. 101–124, 2000.
[83] H. A. Guvenir and M. Kurtcephe, "Ranking instances by maximizing the area under ROC curve," *IEEE Trans. Knowled. Data Eng.*, vol. 25, no. 10, pp. 2356–2366, Oct. 2013.
[84] K. Woods and K. W. Bowyer, "Generating ROC curves for artificial neural networks," *IEEE Trans. Med. Imag.*, vol. 16, no. 3, pp. 329–337, 1997.
[85] J. A. Swets, "Measuring the accuracy of diagnostic systems," *Science*, vol. 240, pp. 1285–1293, 1988.
[86] E. Theodoratou *et al.*, "Associations between dietary and lifestyle risk factors and colorectal cancer in the Scottish population," *Eur. J. Cancer Prev.*, vol. 23, pp. 8–17, Jan, 2014.
[87] C. Conoscenti *et al.*, "Gully erosion susceptibility assessment by means of GIS-based logistic regression: A case of Sicily (Italy)," *Geomorphology*, vol. 204, pp. 399–411, 2014.
[88] J. Fan, S. Upadhye, and A. Worster, "Understanding receiver operating characteristic (ROC) curves," *Cjem*, vol. 8, pp. 19–20, 2006.
[89] S. Haykin, *Neural Networks: A Comprehensive Foundation*. Englewood Cliffs, NJ, USA: Prentice Hall, 1994.
[90] R. S. Govindaraju and A. R. Rao, *Artificial Neural Networks in Hydrology*. New York, NY, USA: Springer, 2010.
[91] D. Hallett, *Goodness of Fit Tests in Logistic Regression*, Toronto, Canada: Univ. of Toronto, 1999, pp. 1–120.
[92] C. Fetter, *Applied Hydrogeology*, 4th ed. Englewood Cliffs, NJ, USA: Prentice Hall, 2001.

Racha Elkadiri was born in Tangier, Morocco, in November 1987. She received the "Ingenieur d'Etat" degree from Mohammadia School of Engineers (EMI), Mohammed V Agdal University, Rabat, Morocco, in 2010. She is currently pursuing the Ph.D. degree at Western Michigan University, Kalamazoo, MI, USA.

Her research interests include geo-engineering, remote sensing, GIS, hydrology, environmental modeling, and data mining applications.

Mohamed Sultan was born in Cairo, Egypt, in 1952. He received the B.Sc. and M.Sc. degrees from Ain Shams University, Cairo, Egypt, in 1974 and 1978, respectively, and the Ph.D. degree from Washington University in St. Louis, St. Louis, MO, USA, in 1984.

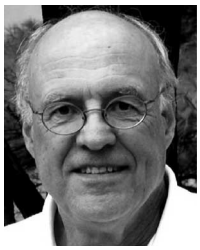
Currently, he is a Full Professor and Chairman of Department of Geosciences with Western Michigan University, Kalamazoo, MI, USA. He has published over 80 research articles in refereed scientific journals. He applies integrated and interdisciplinary research approaches that take advantage of available tools and disciplines (e.g., remote sensing, GIS, geochemistry, geochronology, hydrology, surface runoff and groundwater flow modeling) to address a wide range of complex geological and environmental problems.

Ahmed M. Youssef was born in Sohag, Egypt, in 1970. He received the B.Sc. and M.Sc. degrees from Sohag University, Sohag, Egypt, in 1992 and 1998, respectively, the M.S. degree in geotechnical engineering and the Ph.D. degree in geological engineering from Missouri University of Science and Technology, Rolla, MO, USA, in 2004.

Currently, he is an Associate Professor of Geological Hazards and Engineering with Sohag University, Sohag, Egypt. In addition, he is a Consultant with the Saudi Geological Survey, Jeddah, Saudi Arabia. He has published over 40 research articles in refereed scientific journals. His research interests include geotechnical engineering, geological hazards, site investigation, remote sensing and GIS applications in geotechnical, geological, and environmental problems.

Tamer Elbayoumi was born in Mansoura City, Egypt, in 1976. He received the B.Sc. degree in statistics and the M.Sc. degree in applied statistics from Mansoura University, Mansoura, Egypt, in 1997 and 2003, respectively, and the Ph.D. degree in statistics from Western Michigan University, Kalamazoo, MI, USA, in 2013.

Currently, he is a Visiting Assistant Professor with the Department of Statistics, Oklahoma State University, Stillwater, OK, USA. His research interests include spatial statistical analysis, nonparametric and robust statistics for linear models and time series analysis, and neural networks.



Ronald Chase was born in Marshfield Wisconsin, USA, in 1936. He received the M.S. and Ph.D. degrees from the University of Montana, Missoula, MT, USA, in 1961 and 1968, respectively.

He has been with the Marshall University from 1964 to 1967, the University of Colorado–Pueblo from 1967 to 1973, and Western Michigan University from 1973 to retirement in 2012. From 1996 until retirement, he was the Principal Investigator for multiple U.S. Army Research Office and U.S. Army Corps of Engineers studies and publications that emphasized the kinematics, causes, predictions, and mitigation strategies for mass movements.

Ali B. Bulkhi received the B.Sc. degree in engineering geology from King Abdul Aziz University, Jeddah, Saudi Arabia, in 2000.

Since 2000, he has been working with the Department of Geohazards and Engineering Geology, Saudi Geological Survey, Jeddah, Saudi Arabia.

Mohamed M. Al-Katheeri received the Master degree in engineering geology from King Abdul Aziz University, Jeddah, Saudi Arabia, in 2002.

Since 2000, he has been working with the Geohazards and Engineering Geology Department, Saudi Geological Survey, Jeddah, Saudi Arabia.