Available online at www.sciencedirect.com
SCIENCE DIRECT
ELSEVIER
Engineering Geology 85 (2006) 347–366
ENGINEERING GEOLOGY
www.elsevier.com/locate/enggeo

A comparative study of conventional, ANN black box, fuzzy and combined neural and fuzzy weighting procedures for landslide susceptibility zonation in Darjeeling Himalayas

D.P. Kanungo [a,*], M.K. Arora [b], S. Sarkar [a], R.P. Gupta [c]

[a] *Geotechnical Engineering Division, Central Building Research Institute, Roorkee 247 667, India*
[b] *Department of Civil Engineering, Indian Institute of Technology Roorkee, Roorkee 247 667, India*
[c] *Department of Earth Sciences, Indian Institute of Technology Roorkee, Roorkee 247 667, India*

Received 12 September 2005; received in revised form 29 January 2006; accepted 13 March 2006
Available online 22 May 2006

Abstract

Landslides are one of the most destructive phenomena of nature that cause damage to both property and life every year, and therefore, landslide susceptibility zonation (LSZ) is necessary for planning future developmental activities. In this paper, apart from conventional weighting system, objective weight assignment procedures based on techniques such as artificial neural network (ANN), fuzzy set theory and combined neural and fuzzy set theory have been assessed for preparation of LSZ maps in a part of the Darjeeling Himalayas. Relevant thematic layers pertaining to the causative factors have been generated using remote sensing data, field surveys and Geographic Information System (GIS) tools. In conventional weighting system, weights and ratings to the causative factors and their categories are assigned based on the experience and knowledge of experts about the subject and the study area to prepare the LSZ map (designated here as Map I). In the context of objective weight assignments, initially the ANN as the black box approach has been used to directly produce an LSZ map (Map II). In this approach, however, the weights assigned are hidden to the analyst. Next, the fuzzy set theory has then been implemented to determine the membership values for each category of the thematic layer using the cosine amplitude method (similarity method). These memberships are used as ratings for each category of the thematic layer. Assuming weights of each thematic layer as one (or constant), these ratings of the categories are used for the generation of another LSZ map (Map III). Subsequently, a novel weight assignment procedure based on ANN is implemented to assign the weights to each thematic layer objectively. Finally, weights of each thematic layer are combined with fuzzy set derived ratings to produce another LSZ map (Map IV). The maps I–IV have been evaluated statistically based on field data of existing landslides. Amongst all the procedures, the LSZ map based on combined neural and fuzzy weighting (i.e., Map IV) has been found to be significantly better than others, as in this case only 2.3% of the total area is found to be categorized as very high susceptibility zone and contains 30.1% of the existing landslide area.

Keywords: Landslide susceptibility zonation; GIS; Remote sensing; ANN; Fuzzy; Combined neural and fuzzy

1. Introduction

Landslides are amongst the most damaging natural hazards in the hilly regions. The study of landslides has drawn global attention mainly due to increasing

* Corresponding author. Tel.: +91 1332 283418; fax: +91 1332 272272.
E-mail addresses: dpk_cbri@yahoo.com (D.P. Kanungo), manojfce@iitr.ernet.in (M.K. Arora), shantanu_cbri@yahoo.co.in (S. Sarkar), rpgesfes@iitr.ernet.in (R.P. Gupta).

0013-7952/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.enggeo.2006.03.004

awareness of its socio-economic impacts and also increasing pressure of urbanization on the mountain environment (Aleotti and Chowdhury, 1999). Landslides have represented 4.89% of the natural disasters that occurred worldwide during the years 1990 to 2005 (www.em-dat.net). According to Schuster (1996) and Ercanoglu and Gokceoglu (2004), this trend is expected to continue in future due to increased unplanned urbanization and development, continued deforestation and increased regional precipitation in landslide prone areas due to changing climatic patterns. Hence, the identification of landslide-prone areas is essential for safer strategic planning of future developmental activities in the region. Therefore, Landslide Susceptibility Zonation (LSZ) of an area becomes important whereby the area may be divided into near-homogeneous domains and ranked according to degrees of potential hazard due to mass movements (Varnes, 1984). The area may thus be categorized as very high susceptibility (VHS), high susceptibility (HS), moderate susceptibility (MS), low susceptibility (LS) and very low susceptibility (VLS) zones to produce an LSZ map.

Landslide susceptibility zonation studies in the Himalayas have conventionally been carried out based on manual interpretation of a variety of thematic maps and their superimposition (Anbalagan, 1992; Pachauri and Pant, 1992; Gupta et al., 1993; Sarkar et al., 1995; Mehrotra et al., 1996; Virdi et al., 1997). However, this approach is time consuming, laborious and uneconomical with data collected over long time intervals. In recent times, due to the availability of a wide range of remote sensing data together with data from other sources in digital form and their analysis using GIS, it has now become possible to prepare different thematic layers corresponding to the causative factors that are responsible for the occurrence of landslides in a region (Gupta and Joshi, 1990; van Westen, 1994; Nagarajan et al., 1998; Gupta, 2003). The integration of these thematic layers with weights assigned according to their relative importance in a GIS environment leads to the generation of an LSZ map (Gupta et al., 1999; Saha et al., 2002; Sarkar and Kanungo, 2004; Saha et al., 2005). However, in the studies cited above, the weights were assigned on the basis of the experience of the experts about the subject and the area. The weighting system was thus highly subjective and might therefore contain some implicit biases towards the assumptions made.

For minimizing the subjectivity and bias in the weight assignment process, quantitative methods, namely, statistical analysis, deterministic analysis, probabilistic models, distribution-free approaches and landslide frequency analysis may be utilized. During the last 5 years, bivariate statistical models (Lin and Tung, 2003; He et al., 2003; Suzen and Doyuran, 2004; Saha et al., 2005; etc.), multivariate methods (Dhakal et al., 2000; Clerici et al., 2002; etc.) and probabilistic prediction models (Chi et al., 2002a; Lee et al., 2002a,b; Lan et al., 2004; etc.) have been implemented for LSZ studies. Apart from these methods, some work on distribution-free approaches such as fuzzy set based methods (Chi et al., 2002b; Gorsevski et al., 2003; Tangestani, 2003; Metternicht and Gonzalez, 2005; Ercanoglu and Gokceoglu, 2004), artificial neural network (ANN) models (Arora et al., 2004; Gomez and Kavzoglu, 2005; Yesilnacar and Topal, 2005) and neuro-fuzzy models (Elias and Bandis, 2000; Lee et al., 2004; Kanungo et al., 2005) have recently been attempted for LSZ studies. Due to some success of neural networks and fuzzy set theories in these studies, an attempt has been made here to develop an objective procedure that takes into account the advantages of both neural networks and fuzzy set theory for landslide susceptibility zonation in a part of Darjeeling Himalayas.

2. Weight assignment procedures

The basic pre-requisite for landslide susceptibility zonation studies is the determination of weight and rating values representing the relative importance of factors and their categories respectively for landslide occurrence. These weights and ratings can be determined based on the subjective expert opinions as well as based on an objective analysis. Four different procedures have been implemented here to determine the weights and ratings in order to generate LSZ maps. This section highlights the salient features of these procedures.

2.1. Conventional weighting procedure

In this weighting scheme, factors and their categories are assigned numerical values based on the experience of experts on the subject and about the study area. The numerical values (generally at an ordinal scale from 0 to 9) assigned to factors are termed as weights and those assigned to categories of factors are termed as ratings (Lee et al., 2004). Higher is the numerical value of weight or rating, greater is its influence on the occurrence of landslide.

2.2. Artificial neural network procedure

ANN, which is a useful technique for regression and classification problems, has been successfully applied in

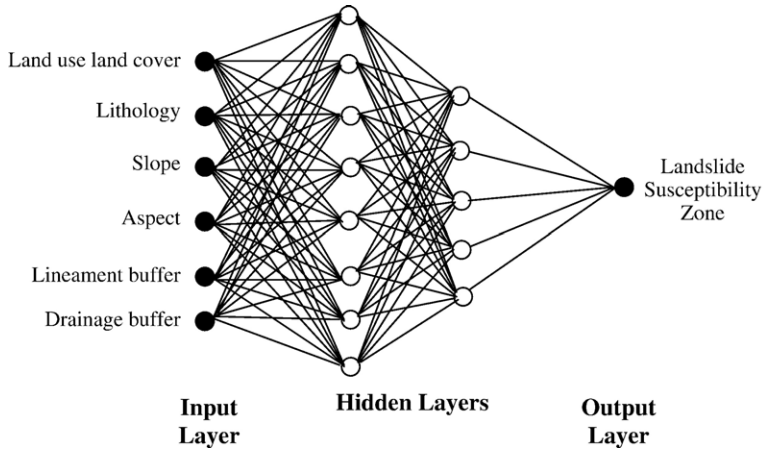

Fig. 1. A schematic diagram of artificial neural network for LSZ using ANN black box procedure.

other fields, and promises to be suitable for the delineation of areas prone to landslide activity. It has been found that ANNs have several advantages for LSZ mapping, as these are non-linear and thus have the capability to analyse complex data patterns. Also, ANN can process data at varied measurement scales such as continuous, ordinal and categorical data, a scenario which is often encountered in LSZ mapping.

An ANN comprises of a number of neurons that work in parallel to transform input data into output classes. A feed-forward multilayer network is generally adopted, which consists of three layers namely input, output and hidden layers in between these two (Paola and Schowengerdt, 1995). Each layer in a network contains sufficient number of neurons depending on the specific application. The neurons in a layer are connected to the neurons in the next successive layer and each connection carries a weight (Atkinson and Tatnall, 1997). The input layer receives the data from different sources (e.g., thematic layers). Hence, the number of neurons in the input layer depends on the number of input data sources. The hidden and output layers actively process the data. The number of hidden layers and their neurons are often determined by trial and error (Gong, 1996). The number of neurons in output layers is fixed by the application and is represented by the class being mapped (e.g., LSZ classes in the present case). Each hidden neuron responds to the weighted inputs it receives from the connected neurons from the preceding input layer (Lee et al., 2004). Once the combined effect on each hidden neuron is determined, the activation at this neuron is determined via a transfer function (Yesilnacar and Topal, 2005). Any differentiable nonlinear function can be used as a transfer function, but a sigmoid function is generally used though there are many other functions (Schalkoff, 1997). The sigmoid function constrains the outputs of a network between 0 and 1.

An important characteristic of a neural network is its capability to learn from the data being processed. The network weights are adjusted in the training process, which can be executed through a number of learning algorithms based on backpropagation learning (Ripley, 1996; Haykin, 1999; Zhou, 1999; Lee et al., 2004; Gomez and Kavzoglu, 2005; Yesilnacar and Topal, 2005). The most widely used backpropagation algorithms are gradient descent and gradient descent with momentum. These are often too slow for the solution of practical problems. The faster algorithms use standard numerical optimizers such as conjugate gradient, quasi-Newton and Levenberg–Marquardt approach. In this study, Levenberg–Marquardt algorithm (implemented as TRAINLM in MATLAB software) has been used for training the neural network. The details of this algorithm can be found in Hagan and Menhaj (1994) and Hagan et al. (1996). Unlike gradient descent algorithms, it does not consider learning rate and momentum factor as its parameters. However, the main scalar parameter involved in this algorithm is mu (μ), which is modified in an adaptive fashion after giving an initial random value.

In back propagation learning, the difference (i.e., error) between the neural network outputs and target outputs is back propagated through the neural network and is minimized by updating interconnection weights between the layers (Arora et al., 2004; Lee et al., 2004). The process of back propagating the error is repeated iteratively until the error is minimized to an acceptable value and the adjusted weights are obtained, which are then used to determine the network outputs. The performance of the network depends on the accuracy obtained over a set of testing dataset. If the network is

trained and tested to an acceptable accuracy, then the adjusted weights are used to determine the outputs of each unknown data sample. This approach has been called as ANN black box approach.

In this study, a feed forward multi-layer ANN with one input layer, two hidden layers and one output layer has been considered to produce an LSZ map. The input layer contains 6 neurons each representing a causative factor that contributes to the occurrence of the landslide. The output layer contains a single neuron representing one of the five LSZ classes (VLS, LS, MS, HS and VHS) for a given set of input values for a pixel.

By varying the number of neurons in both the hidden layers, the neural networks are run several times to identify the most appropriate neural network architecture based on training and testing accuracies. A schematic diagram of the best neural network architecture is given in Fig. 1. The whole dataset of the study area is then processed with the most accurately trained and tested network to generate the LSZ map of the study area. The neural network processing has been implemented in Neural Network Tool Box of MATLAB Software.

2.3. Fuzzy set based procedure

Fuzzy relation concept defined by Zadeh (1973) is based on the theory of fuzzy sets. A fuzzy set can be explained as a set containing elements that have varying degrees of membership in the set (Ross, 1995). Fuzzy relations play an important part in fuzzy modeling and are based on the philosophy that everything is related to some extent (Dubois and Prade, 1980). In this paper, one of the well known similarity methods, cosine amplitude method (Ross, 1995; Ercanoglu and Gokceoglu, 2004), has been used to determine the relationship between the landslide occurrence and the categories responsible for such activity. The membership degrees of categories of each factor are determined by establishing the strength of the relationship (r_{ij}) between the existing landslides and the categories.

Let n be the number of categories of a thematic layer represented as an array $X=\{x_1, x_2, \ldots, x_n\}$, each of its elements, x_i, is a vector of pixels p (i.e., the size of the image in the present context) and can be expressed as,

$$x_i = \{x_{i1}, x_{i2}, \ldots, x_{ip}\} \tag{1}$$

Each element of a relation, r_{ij}, results from a pairwise comparison of a category of a thematic layer i (i.e., layer corresponding to a causative factor) with a category of thematic layer j (i.e., landslide distribution layer), say x_i and x_j containing elements x_{ik} and x_{jk} respectively. In the cosine amplitude method, for example, r_{ij} (membership grades) between categories of a thematic layer and that of the landslide distribution layer are computed by the following equation with its values ranging from 0 to 1 ($0 \le r_{ij} \le 1$).

$$r_{ij} = \frac{\left[\sum_{k=1}^{p} x_{ik} x_{jk}\right]}{\sqrt{\left(\sum_{k=1}^{p} x_{ik}^2\right)\left(\sum_{k=1}^{p} x_{jk}^2\right)}} \tag{2}$$

The r_{ij} value for a category can be defined as the ratio of total number of landslide pixels in the category to the square root of the multiplication of total number of pixels in that category and the total number of landslide pixels in the area. Values of r_{ij} close to 0 indicate dissimilarity, whereas values close to 1 indicate the similarity between the two datasets. Eq. (2) leads to $(n-1)$ r_{ij} images corresponding to each category of the thematic layers under consideration. These images show r_{ij} values at the pixels belonging to the category in question, whereas rest of the pixels indicate 0 values. The corresponding r_{ij} images for various categories of a thematic layer are composited together to generate an r_{ij} image for that thematic layer and is represented as R_l, where l varies from 1 to t thematic layers belonging to each causative factor (e.g., 6 thematic layers in the present case). The fuzzy processing has been implemented in MS Excel spreadsheet and ArcView GIS software.

2.4. Combined Neural and fuzzy procedure

In another procedure, the ANN connection weights may be used to characterize the input data sources (e.g., the causative factors) in terms of ranks or weights. In this process, the connection weight matrices for input–hidden, hidden–hidden and hidden–output layers are obtained for a two-hidden layer network. Simple matrix multiplications of these weight matrices give rise to the final weight matrix corresponding to the factors (Olden et al., 2004). For example, if a network of 6/14/8/1 architecture (representing 6 neurons in the input layer, 14 neurons in the 1st hidden layer, 8 neurons in the 2nd hidden layer and one neuron in the output layer) is considered, connection weight matrices of 6×14, 14×8 and 8×1 are obtained. The product of 6×14 and 14×8 matrices gives a resultant matrix of 6×8. Subsequently, the product of 6×8 and 8×1 matrices gives an output

[1] Input(I) – Hidden A (HA) Connection Weights

	HA_1	HA_2	HA_3	HA_4	HA_5	HA_6	HA_7	HA_8	HA_9	HA_{10}	HA_{11}	HA_{12}	HA_{13}	HA_{14}
I_1	20.45	-12.98	3.52	-0.98	23.70	9.68	-1.92	-22.60	12.37	-3.26	5.35	12.36	16.72	9.68
I_2	-49.57	-16.84	51.54	87.70	-20.19	52.71	-79.29	-24.83	43.90	-91.70	-61.46	-46.21	-43.25	15.49
I_3	-31.63	-48.57	-52.19	47.39	-11.27	52.44	-5.31	-12.58	44.33	3.84	-46.30	24.22	44.39	-66.34
I_4	12.83	-9.89	18.44	0.15	21.10	-3.93	-16.33	-24.43	-20.15	14.86	-15.81	3.08	-23.33	13.79
I_5	-2.28	9.72	6.48	3.14	-0.06	8.78	8.53	3.95	5.98	1.89	-3.46	-4.09	-4.07	-0.30
I_6	13.46	-12.42	14.70	5.39	-11.54	9.72	2.81	-3.41	10.30	6.19	17.86	25.66	1.28	20.66

[2] Hidden A (HA) – Hidden B (HB) Connection Weights

	HB_1	HB_2	HB_3	HB_4	HB_5	HB_6	HB_7	HB_8
HA_1	-1.41	2.15	-3.63	-1.70	2.96	1.06	-1.71	-2.60
HA_2	3.10	-1.31	0.69	-1.21	-0.90	0.60	-1.92	2.65
HA_3	1.42	-0.88	-0.07	-1.85	-0.90	-0.20	0.19	2.53
HA_4	-1.17	-0.25	-0.66	1.76	2.04	1.55	-0.72	0.60
HA_5	2.56	-2.74	0.08	-5.36	-0.76	-1.32	1.86	-2.10
HA_6	1.98	1.50	-0.45	3.35	-1.65	1.91	-3.33	1.71
HA_7	-1.54	2.76	2.04	-3.07	0.60	-1.84	-2.96	2.43
HA_8	1.84	-1.46	1.27	0.90	-2.89	-1.83	0.76	-0.08
HA_9	-2.55	-0.45	-0.96	-1.43	-3.94	-0.21	-3.50	-0.20
HA_{10}	1.42	-1.84	-2.44	0.59	0.42	-4.59	-3.45	-1.41
HA_{11}	0.01	-2.68	0.48	1.37	1.55	-0.52	-0.13	-0.71
HA_{12}	-0.30	-1.71	-1.84	2.01	1.34	0.27	-1.09	-0.04
HA_{13}	-2.62	-1.19	1.75	2.32	-1.97	-2.60	1.65	-2.55
HA_{14}	4.22	0.96	-1.32	-2.72	-2.99	-2.59	0.61	-2.55

[3] Connection Weight Products of [1] & [2]

	HB_1	HB_2	HB_3	HB_4	HB_5	HB_6	HB_7	HB_8
I_1	-64.53	-10.19	-125.47	-111.32	10.13	-11.87	-21.14	-184.95
I_2	66.70	259.49	29.64	234.11	-377.06	884.93	225.86	446.47
I_3	-686.94	81.93	48.73	741.85	-52.38	229.35	-242.08	24.89
I_4	193.99	11.19	-204.81	-211.31	151.55	13.25	70.24	-83.25
I_5	48.23	21.67	19.55	-36.32	-60.27	3.15	-96.88	95.38
I_6	-4.84	0.50	-154.42	69.02	13.00	-42.62	-138.08	-59.90

[4] Hidden B – Output Connection Weights

	HB_1	HB_2	HB_3	HB_4	HB_5	HB_6	HB_7	HB_8
Output	3.43	-1.61	-3.87	4.83	2.71	-2.14	-3.67	-3.25

[5] Connection Weight Products of [3] & [4]

Y / X	HB_1	HB_2	HB_3	HB_4	HB_5	HB_6	HB_7	HB_8
I_1	-221.34	16.41	485.57	-537.68	27.45	25.40	77.58	601.09
I_2	228.78	-417.78	-114.71	1130.75	-1021.83	-1893.75	-828.91	-1451.03
I_3	-2356.2	-131.91	-188.59	3583.14	-141.95	-490.81	888.43	-80.89
I_4	665.39	-18.02	792.61	-1020.63	410.70	-28.36	-257.78	270.56
I_5	165.43	-34.89	-75.66	-175.43	-163.33	-6.74	355.55	-309.99
I_6	-16.60	-0.81	597.61	333.37	35.23	91.21	506.75	194.68

$$I_X = \left| \sum_{Y=1}^{8} HB_{XY} \right|$$

	Weight	Rank	Normalized Weight
I_1	474.48	5	0.544
I_2	4368.48	1	5.007
I_3	1081.22	3	1.239
I_4	814.47	4	0.933
I_5	245.06	6	0.281
I_6	1741.44	2	1.996

Fig. 2. Steps for computations of connection weight matrices of ANN to characterize input data layers in terms of ranks and weights (in the present study, I_1—Land use land cover, I_2—Lithology, I_3—Slope, I_4—Aspect, I_5—Drainage buffer and I_6—Lineament buffer); Note: In step [5], X: 1–6; Y: 1–8.

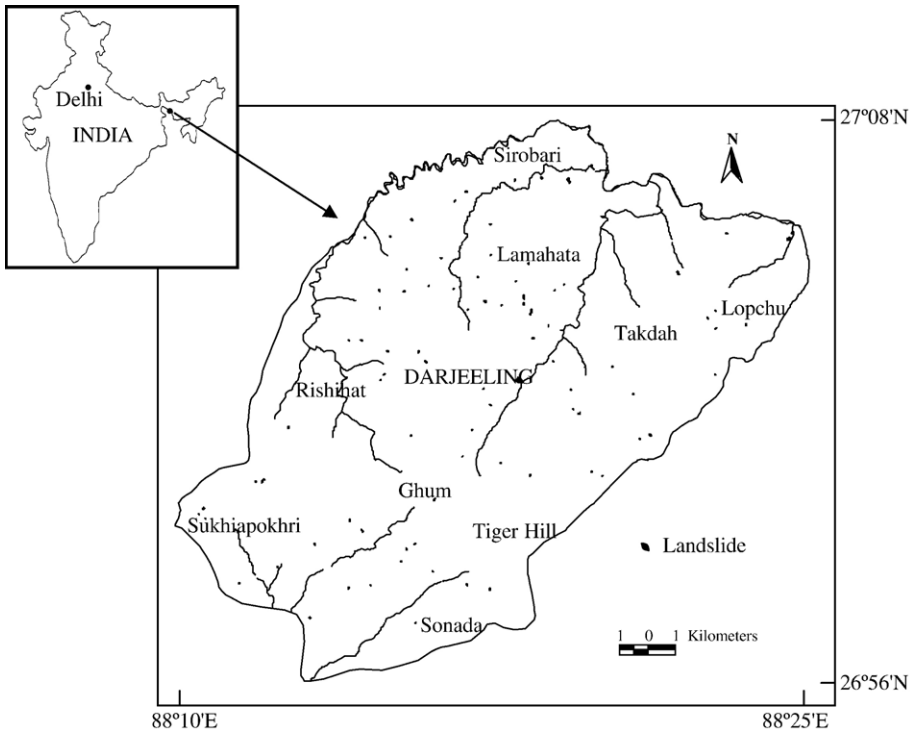

Fig. 3. The study area with landslide distribution in Darjeeling Himalayas.

matrix of 6×1 which corresponds to the weights of 6 factors. The absolute values of these weights are considered in the present work to rank the factors meaning thereby that the factor with maximum absolute weight is assigned as rank 1 and the factor with the minimum absolute weight as rank 6. This has been illustrated in Fig. 2.

3. Study area

The Darjeeling Himalayas, encompassing a total area of 3000 km^2 rise abruptly from the alluvial plains of West Bengal and attain a maximum elevation of about 2600 m. The area lies between Sikkim in the North, Bhutan in the east and Nepal in the west. The southern foot hill region is characterized by East–West trending highly dissected platform of terrace deposits. The southerly flowing river Tista approximately divides the Darjeeling Himalayas into two parts, the eastern and the western parts occupied by Kalimpong and Darjeeling hills respectively. The Tista River however, does not fall within the study area. The study area encompasses Darjeeling hill which lies between latitude 26°56′–27°8′N and longitude 88°10′–88°25′E and covers an area of about 254 km^2 (Fig. 3). The main habitat areas are Darjeeling, Sonada and Sukhiapokhri.

The study area is highly dissected by many ridges and valleys. The maximum elevation of 2584 m occurs at the Tiger hill. The area is dominated by slopes ranging between 15° and 35° while steep slopes of >35° occupy smaller area. In general, the gentle slopes of 0–15° were found on the ridges and at places in the region of lower relief also.

The Darjeeling Himalayas lie within the Lesser and Sub Himalayan belts. The tectonic units in the area occur in inverted order of stratigraphic superposition. Various rock groups have been named locally (Acharya, 1989). The Daling group of rocks comprises of low-grade metamorphic rocks and includes slate, phyllite, schist, quartzite, greywacke and epidiorite. The Darjeeling Group consists primarily of foliated gneisses. Rocks of the Paro Sub-group, which have similar characteristics to the Darjeeling Group, are present at lower elevations.

The annual rainfall in the area is of the order of 3000 mm to 6000 mm. The rainfall pattern is highly seasonal with a maximum rainfall during the monsoon season from June to October. The main land use practice in the study area is tea plantation. The agriculture land is mostly developed around the habitat areas. In general, the area is dominated by thick forest particularly in the eastern part.

4. Thematic data layer preparation

Various thematic data layers corresponding to causative factors namely lithology, slope, aspect, lineaments, land use land cover and drainage have been prepared. These factors fall under the category of preparatory factors, which make the slope susceptible to movement without actually initiating it and thus, are considered responsible for the occurrence of landslides in the region for which pertinent data can be collected from available resources as well as from the field. The triggering factors such as rainfall and earthquakes, set off the movement by shifting the slope from a marginally stable to an actively unstable state. Further, the attributes of the ground (internal factors) in terms of landslide susceptibility have been considered here. Rainfall and earthquakes are external factors and temporal phenomena. Also, past data on these external factors in relation to landslide occurrence are not available. Therefore, these factors have not been included in this study. A thematic layer corresponding to the landslide distribution map has also been prepared to establish a spatial correlation between existing landslides and the causative factors, which will be helpful for the preparation and evaluation of LSZ maps of the area using different weighting procedures.

The IRS-1C-LISS-III (acquired on 22nd March, 2000) and 1D-PAN (acquired on 3rd April, 2000) data along with Survey of India topographic maps at 1:25,000 and 1:50,000 scale, and the geological map at 1:250,000 scale published by Acharya (1989) are the main data sources used to generate these thematic data layers. Extensive field data have been collected during the years 2001 to 2003 to collect information on existing landslide distribution to assist in creation of training and testing datasets, finding out fuzzy membership values and validation of LSZ maps. The months of March and April were preferred for field data collection as these coincided with the date of remote sensing data acquisition. All the thematic data layers have been resampled to match the nominal spatial resolution (i.e., 25 m) of IRS-LISS-III multispectral image.

4.1. Landslide distribution map

The mapping of existing landslides is essential to study the relationship between the landslide distribution and the causative factors. As, it is not possible to map each and every landslide through field surveys in such a terrain, a comprehensive mapping of landslide has been undertaken through remote sensing image interpretation. The identification of landslides on remote sensing image is based on the spectral characteristics, shape, contrast and the morphological expression. In general, there is a distinct spectral contrast between landslides and the background area. High spatial resolution IRS-1C-PAN and PAN-sharpened LISS-III images have been used for landslide mapping. On the PAN image, landslides appear as features of very light tones due to rock debris without any vegetation on the slope. After enhancing the contrast of the PAN image, landslides occurring in barren areas can also be identified. A few old landslides are identified on the basis of their shape, landform and drainage. On the PAN-sharpened LISS-III image, the landslides appear as bright-white features (due to high reflectance) that are easily distinguished from other features. Further, landslides are characterized by fan shape, sharp lines of break in topography and sometimes a local drainage anomaly. Often, the toe part of the slide gives rise to a debris flow channel. Many of the landslides identified on remote sensing images have also been verified in the field.

A total of 101 landslides of varying dimensions (180 m^2 to 27400 m^2) have been identified from remote sensing images and field surveys. A majority of landslides have areal extent of 500 m^2–2000 m^2. Most of the observed landslides are rock slides. However, in some cases, complex types of failure are also possible.

The landslides thus identified have been digitized as polygons in separate vector layers one each for remote sensing derived and field mapped, which are then merged into single landslide layer. This layer has been converted to a rasterized landslide distribution map (Fig. 3) for further analysis.

4.2. Digital elevation model and its derivatives

The Digital Elevation Model (DEM) is an excellent source to derive topographic attributes responsible for landslide activity in a region. Therefore, a DEM at a spatial resolution corresponding to pixel size 25 m × 25 m has been generated by digitization of contours on the topographic maps. The DEM is subsequently used to derive the slope and aspect data layers. Slope angle is one of the key factors in inducing slope instability. The slope data layer consists of 5 classes with 10° interval as per the slope classification used in other studies (Anbalagan, 1992; Gupta et al., 1999; Dhakal et al., 2000). Aspect is defined as the direction of maximum slope of the terrain surface and has an indirect influence on slope instability. In general, the south facing slopes have lesser vegetation density as compared to the north facing slopes and hence, the erosional activity is relatively more in the former case (Sinha et al., 1975). The aspect data layer derived here

represents nine classes, namely, N, NE, E, SE, S, SW, W, NW and flat as per the classification given in other studies (Sarkar and Kanungo, 2004; Saha et al., 2005).

4.3. Lithology

Different rock types (or lithology) have varied composition and structure, which contribute to the strength of the material. The stronger rocks give more resistance to the driving forces as compared to the weaker rocks, and hence are less prone to landslides and vice versa. The lithology data layer has been prepared by digitizing the polygons from the co-registered geological map of Sikkim–Darjeeling area (Acharya, 1989) in a vector layer. Necessary modifications have also been incorporated in this vector layer after field verification. This lithology data layer is later rasterized at 25 m spatial resolution. The six rock types present in this data layer are Darjeeling gneiss, Paro gneiss, Lingtse granite gneiss, feldspathic greywacke, and quartzites of the Paro sub-group and the Reyang group.

4.4. Lineaments

Lineaments are the structural features which describe the zone/plane of weakness, fractures and faults along which landslide susceptibility is higher. It has generally been observed that the probability of landslide occurrence increases at sites close to lineaments, which not only affect the surface material structures but also make contribution to terrain permeability causing slope instability. Lineaments have been interpreted from the PAN and LISS-III images. The individual bands of LISS-III image are enhanced using linear contrast stretching followed with a 3×3 edge filters to highlight high frequency features. Subsequently, all the four bands are layer stacked to produce the edge-enhanced image which has been used for visual interpretation of lineaments. The lineaments have been interpreted based on the tonal contrast, structural alignments and rectilinear trends of morphological features and linear stream courses that are conspicuous by their abrupt changes in the course. There is no major thrust/fault reported in the study area, but mega lineaments have been identified. The interpreted lineaments have been digitized on-screen and subsequently rasterized to produce the lineament data layer. Initially, buffer zones at 250 m intervals were created. These buffer zones were spatially cross-correlated with the landslide pixels in the area and it was observed that 98% of landslide pixels occurred in 1st two buffer zones (up to 500 m). Hence, it was decided to consider four buffer zones at 125 m intervals up to 500 m and another buffer zone beyond 500 m to establish the influence of lineaments on landslide occurrence. Thus, a lineament buffer layer consisting of five classes such as 0–125 m, 125–250 m, 250–375 m, 375–500 m and >500 m has been prepared.

4.5. Drainage

Many of the landslides in hilly areas occur due to the erosional activity associated with drainage. Therefore, a drainage data layer has been prepared by digitizing the drainages from the topographic maps in a vector layer. Later, this layer has been overlaid on IRS-LISS-III image for updating the drainages. This was felt necessary as most of the 1st order drainages, which were not present on the topographic maps, could be interpreted from the LISS-III image, which also showed change in the course of the river and other major drainages at some places. The ordering of the drainage has been performed on the basis of Strahler's classification scheme (Strahler, 1964). Drainages up to 6th order have been observed in the study area. Subsequently, the vector layer has been rasterized at 25 m spatial resolution. Initially, 25 m buffer zones on either side of the drainages for all the drainage orders were created. It was observed from the spatial correlation of landslide distribution in these buffer zones that majority of landslide pixels occurred in the 1st and 2nd order drainage buffers only. Therefore, 25 m buffer zones around these drainages only have been considered to create a drainage buffer layer for further analysis.

4.6. Land use land cover

Land use land cover is also a key factor responsible for landslide occurrences. The incidence of landslide is inversely related to the vegetation density. Hence, barren slopes are more prone to landslide activity as compared to the forest area. Eight dominant land use land cover classes namely thick forest, sparse forest, tea plantation, agriculture, barren, built up, water bodies and river sand have been considered similar to other studies (Sarkar and Kanungo, 2004; Saha et al., 2005). A very small portion of the study area is covered by cloud and its shadow in the LISS-III image. Initially, this portion was masked. The four spectral bands of LISS-III image, Digital Elevation Model (DEM) and Normalized Difference Vegetation Index (NDVI) image have been considered to prepare a land use land cover map in a multi-source classification process using the most widely adopted maximum likelihood classifier. The map has been prepared at an overall classification accuracy of 94.7%. Subsequently, masked portion of the land use land cover

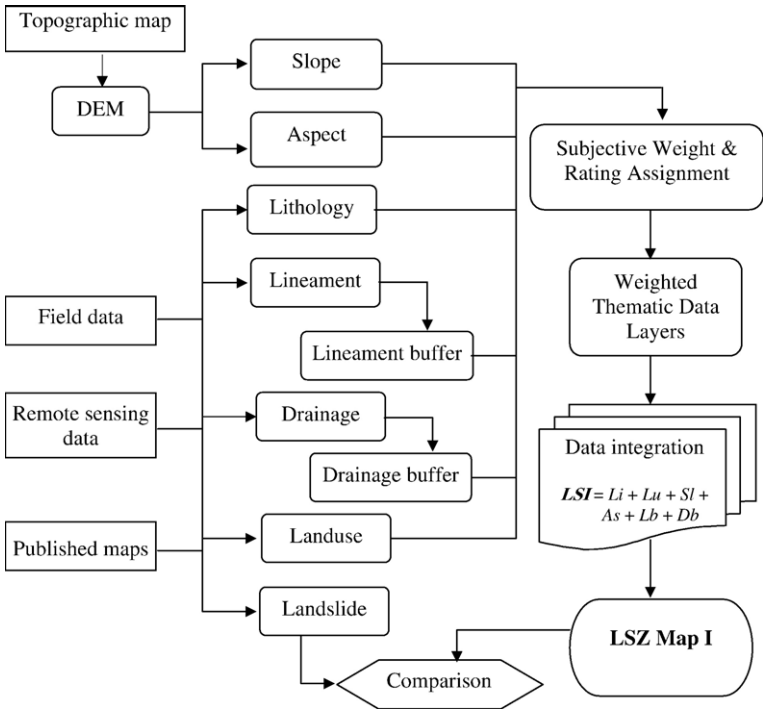

Fig. 4. Flow diagram showing different stages of preparing LSZ map using conventional weighting procedure.

map thus prepared has been filled with the land use land cover information obtained from field surveys to generate the final land use land cover layer.

5. Implementation of various procedures for LSZ

The LSZ mapping was performed in GIS environment to categorize each and every pixel of the dataset to one of the landslide susceptibility zones. GIS tool allows for the storage and manipulation of information concerning different factors as distinct layers and thus provide an excellent tool for LSZ mapping. The thematic data layers were prepared in GIS platform and the data were stored as attributes for further analysis. The weights and ratings of the thematic layers and their categories respectively were determined using four different weighting procedures as mentioned earlier. All the four procedures were implemented in ArcView GIS software to generate LSZ maps.

5.1. LSZ using conventional weighting procedure

5.1.1. Implementation

The conventional weighting procedure involves assignment of weights and ratings to the thematic layers and their categories respectively based on the knowledge of the study area and the experience on the subject. Different steps of this procedure for LSZ are given in Fig. 4. Such weighting scheme was used by Sarkar and Kanungo (2004) with a different combination of thematic layers and their categories.

In the study area, it was observed that most of the landslides were associated with drainage channels and hence the maximum weight was assigned to the drainage layer. Also, maximum rating of 9 was assigned to 1st order drainage buffer category as most of the landslides initiate from the lower order drainages. The next important factor considered was the lineament. Here, the maximum rating was given to the 0–125 m lineament buffer category as the nearness to the lineaments controls the occurrence of the landslide. Since the steeper slopes are more prone to landslide, the slope classes were given ratings in the descending order. The competent rocks such as quartzite, greywacke are less susceptible to landslides than the gneisses as more number of landslides are observed in gneissic rock in the field. Hence, the ratings to lithology categories were assigned accordingly. Occurrence of landslides also depends on the type of the land use land cover. Barren slopes are more susceptible to erosion as compared to areas with thick forest and hence, maximum rating was assigned to the barren slopes and minimum to the thick forest. The slope aspect has an indirect influence on slope instability. Based upon the landslide distribution, south and east

Table 1
Weights and ratings for thematic layers and their categories (conventional weighting procedure)

Thematic layers	Categories	Weights	Ratings
Drainage buffer	1. 1st order	9	9
	2. 2nd order		5
Lineament buffer	1. 0–125 m	8	9
	2. 125–250 m		7
	3. 250–375 m		5
	4. 375–500 m		3
	5. >500 m		1
Slope	1. 0–15°	7	1
	2. 15–25°		3
	3. 25–35°		5
	4. 35–45°		7
	5. >45°		9
Lithology	1. Darjeeling gneiss	6	7
	2. Feldspathic greywacke		3
	3. Paro gneiss		5
	4. Lingse granite gneiss		9
	5. Paro quartzite		1
	6. Reyang quartzite		1
Land use land cover	1. Agriculture land	4	5
	2. Tea plantation		3
	3. Thick forest		1
	4. Sparse forest		7
	5. Barren land		9
	6. Habitation		2
	7. Water body		0
	8. River sand		0
Aspect	1. Flat	1	0
	2. North		1
	3. Northeast		4
	4. East		7
	5. Southeast		8
	5. South		9
	7. Southwest		6
	8. West		3
	9. Northwest		2

facing slopes were considered more prone to landslides than the other slopes (Dhakal et al., 2000). Considering these facts and field observations, ratings for slope aspect categories were assigned accordingly. The weights and ratings thus assigned to each thematic layer and their categories are given in Table 1.

The weighted thematic data layers were generated by algebraically multiplying the weight of the layer with the ratings of the corresponding categories of each layer. In the present case, six weighted thematic data layers namely lithology, land use land cover, slope, aspect, lineament buffer and drainage buffer were produced. These layers were laid over one another and arithmetically added according to the following equation to generate a Landslide Susceptibility Index (LSI) map in GIS,

$$\mathrm{LSI} = \mathrm{Li} + \mathrm{Lu} + \mathrm{Sl} + \mathrm{As} + \mathrm{Lb} + \mathrm{Db} \quad (3)$$

where Li, Lu, Sl, As, Lb and Db are abbreviations for the weighted thematic layers for lithology, land use land cover, slope, aspect, lineament buffer and drainage buffer respectively.

The LSI values range from 21 to 310, which were then categorized to produce landslide susceptibility classes. A judicious way for such classification is to search for abrupt changes in values (Davis, 1986). The classification procedure reported by Sarkar and Kanungo (2004) was followed. The class boundaries were drawn at LSI values of 68, 137, 176 and 236 to obtain five susceptibility zones. The LSZ map (referred here as Map I), thus obtained is shown in Fig. 5a. The area covered by five different landslide susceptibility classes and the area of landslides occupied per class have also been determined (Table 2).

5.1.2. Analysis of the output LSZ Map I

The five susceptibility zones in LSZ Map I were distributed all over the study area. The map did not show any definite pattern for the distribution of susceptibility zones. It can be inferred from Table 2 that 33.3% of HS and VHS areas together could predict 58.7% of existing landslide area. It was again observed from this map that the VHS and HS zones represented mostly the 1st and 2nd order drainage buffer areas. This happened as the drainage layer was assigned the maximum weight in this conventional procedure.

5.2. LSZ using ANN Black Box procedure

5.2.1. Implementation

The flow diagram indicating the different steps followed in this procedure is shown in Fig. 6. Initially, the field data on landslides (i.e., existing landslide distribution) were used to process the neural network. But, due to less number of landslide pixels, the neural network accuracies were found to be very low. Therefore, similar to earlier studies (Arora et al., 2004), the map obtained from conventional approach was considered as reference map, to derive representative sample sizes. Two independent sets of training and testing data were randomly selected with their land slide susceptibility class known from the LSZ map I. Each dataset consisted of 500 mutually exclusive pixels corresponding to each landslide susceptibility zone (Foody and Arora, 1997). The training dataset was used to train various network architectures while the testing dataset was used to control the overtraining of the network and to evaluate the accuracy of the networks. The input values for neural network processing correspond to the attributes of the category of a thematic layer (as mentioned in 2nd

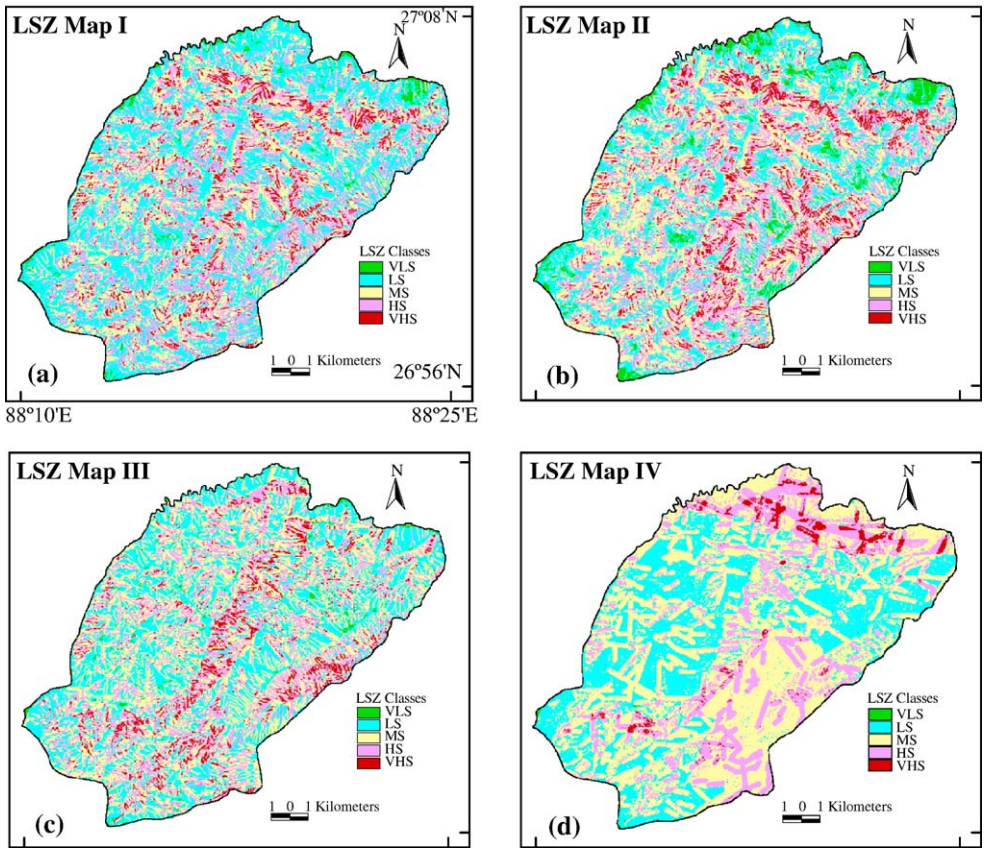

Fig. 5. LSZ maps using four different procedures. (a) Conventional weighting procedure. (b) ANN black box procedure. (c) Fuzzy set based procedure. (d) Combined neural and fuzzy procedure.

column of Table 2), which were normalized with respect to the highest value within each thematic layer. The normalized attribute values of different landslide susceptibility zones (VLS, LS, MS, HS and VHS) in the desired output in order of increasing susceptibility correspond to 0.2, 0.4, 0.6, 0.8 and 1.0 respectively.

A total of 39 neural network architectures were created by varying the number of neurons in both the hidden layers. The training process was initiated by assigning arbitrary initial connection weights, which were constantly updated until an acceptable accuracy over training data was reached. These adjusted weights obtained from the trained network were subsequently used to process the testing dataset to examine the generalization capability of the network.

The performance of the networks was evaluated by determining both the training and testing data accuracies

Table 2
Distribution of landslide susceptibility zones and landslides in different LSZ maps

Landslide susceptibility zones	LSZ Map I		LSZ Map II		LSZ Map III		LSZ Map IV	
	Area (%)	Landslide area occupied per class (%)	Area (%)	Landslide area occupied per class (%)	Area (%)	Landslide area occupied per class (%)	Area (%)	Landslide area occupied per class (%)
Very high susceptibility zone	6.5	10.6	7.7	10.3	6.1	41.0	2.3	30.1
High susceptibility zone	26.8	48.1	26.9	40.4	22.7	25.1	20.2	31.9
Moderate susceptibility zone	30.2	26.5	35.4	36.1	39.4	25.9	48.4	26.5
Low susceptibility zone	34.9	14.5	25.8	12.7	30.4	8.0	28.8	11.5
Very low susceptibility zone	1.6	0.3	4.2	0.5	1.4	0.0	0.3	0.0

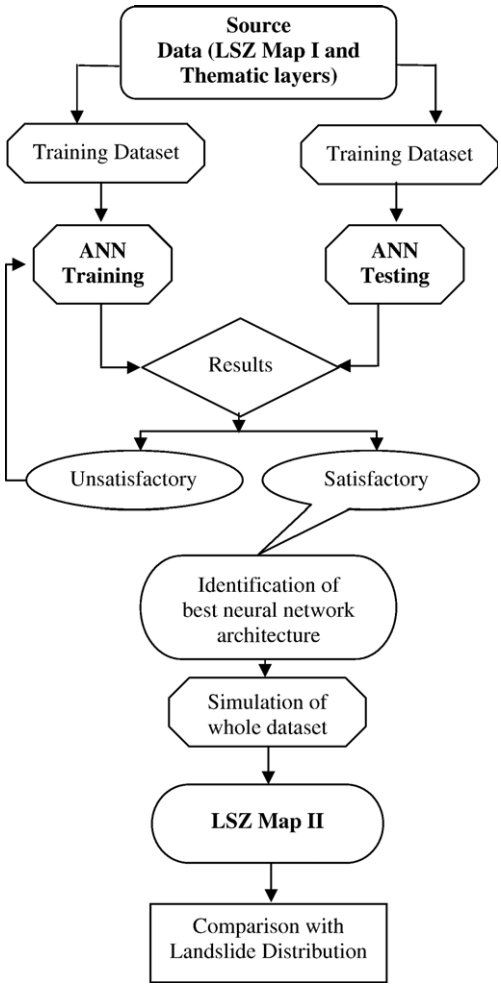

Fig. 6. Flow diagram showing different steps of ANN black box procedure.

in terms of correlation coefficient, root mean squared error (RMSE) (Freund, 1992) and also by the percent correct or overall classification accuracy (Congalton, 1991). The training and testing accuracies for some networks are given in Table 3. A variation in both training and testing data accuracies can be seen as the neural network architectures change. This infers that there exists an optimal network architecture for a given dataset. In the present case, the network architecture 6/13/7/1 with training data accuracy (correlation coefficient of 0.918, root mean square error 0.112 and 74.4% correct) and the testing data accuracy (correlation coefficient of 0.896, root mean square error 0.126 and 72.6% correct) was the most appropriate one. Therefore, weights obtained from this network were subsequently used to obtain the network output of each pixel. The output values ranged from 0.062 to 0.993, which were categorized into one of the five landslide susceptibility zones (Table 4) to produce the LSZ map in GIS. The LSZ map (referred here as Map II) thus produced is shown in Fig. 5b. The area covered by different landslide susceptibility zones and the area of landslides occupied per class were also determined (Table 2).

5.2.2. Analysis of the output LSZ Map II

It can be inferred from Table 2 that in case of LSZ Map II, an area of 34.6% belonging to VHS and HS zones put together could predict only 50.7% landslide area. Further, the map does not depict any definite pattern, as the susceptibility zones are distributed all over the area. Moreover, the drainage buffer layer has a major control on this LSZ map also. This may be due to the fact that the LSZ Map I was used as the reference map for training the network to produce an ANN black box based LSZ map (Map II). Hence, there is lot of similarity between LSZ Maps I and II.

Table 3
Training and testing data accuracies of ANN black box procedure (bold indicates the best acceptable architecture in this study)

NN architecture	Correlation coefficient		RMSE		Accuracy (%)		
	Training	Testing	Training	Testing	Training	Testing	Diff.
6–4–2–1	0.877	0.870	0.136	0.139	64.9	63.2	1.7
6–5–2–1	0.886	0.875	0.131	0.137	67.4	65.0	2.4
6–6–4–1	0.892	0.881	0.128	0.134	68.0	65.3	2.7
6–7–5–1	0.906	0.893	0.120	0.127	71.0	69.3	1.7
6–8–5–1	0.908	0.891	0.119	0.128	70.8	68.7	2.1
6–9–5–1	0.911	0.893	0.117	0.127	72.1	68.6	3.5
6–10–4–1	0.912	0.893	0.116	0.128	71.4	69.8	1.6
6–11–3–1	0.915	0.897	0.114	0.125	73.9	70.0	3.9
6–12–4–1	0.912	0.892	0.116	0.128	72.9	70.9	2.0
6–13–5–1	0.915	0.899	0.114	0.124	73.8	70.3	3.5
6–13–7–1	**0.918**	**0.896**	**0.112**	**0.126**	**74.4**	**72.6**	**1.8**
6–13–9–1	0.918	0.893	0.112	0.128	73.2	68.0	5.2
6–14–3–1	0.912	0.892	0.116	0.128	71.9	70.1	1.8
6–15–4–1	0.919	0.891	0.112	0.128	74.2	69.2	5.0
6–15–6–1	0.915	0.894	0.114	0.127	73.8	69.5	4.3
6–15–8–1	0.914	0.894	0.115	0.127	72.9	69.6	3.3
6–16–3–1	0.917	0.890	0.113	0.129	72.7	69.0	3.7
6–16–7–1	0.920	0.889	0.111	0.130	73.8	67.2	6.6

5.3. LSZ using fuzzy set based procedure

5.3.1. Implementation

In the fuzzy set based procedure, ratings of each category of a given thematic layer were determined on the basis of cosine amplitude similarity method, which were then integrated in GIS by considering the weight of each thematic layer as one (or constant) to generate the LSZ map.

The cosine amplitude method as described earlier was adopted to determine the ratings of the categories of factors. The landslide distribution map and different categories of thematic layers taken one at a time were considered as two datasets for the computation of rating or strength of relationship (r_{ij}). In the landslide distribution layer, pixels belonging to landslides were assigned as 1 and the rest were assigned as 0. Similarly, a value of 1 was assigned to a pixel belonging to a particular category of a thematic layer and a value of 0 to the rest. Hence, in total 36 data layers in binary form were generated, which contained 35 layers of categories of thematic layers (Table 5) and one layer of landslide distribution. These layers were used for the determination of r_{ij} in GIS so as to generate 35 images denoting r_{ij} values. The r_{ij} values thus obtained are listed in Table 5.

The corresponding r_{ij} images for various categories of a thematic layer were combined together to generate an R_l image for that thematic layer. The integration of these R_l images for various thematic layers can be performed in several ways to compute LSI values. The simplest approach is to add this arithmetically which is similar to any other conventional GIS integration process. Alternatively, to bring fuzziness in the integration process also, the use of fuzzy algebraic sum, fuzzy algebraic product and fuzzy gamma operator can be put forth. In view of this, the LSI values were computed in two different ways: (a) using arithmetic integration and (b) using fuzzy gamma operator. The performance of the two methods was also examined. It was found that the arithmetic overlay approach of thematic layer integration yielded better results than the fuzzy gamma operator for this dataset. Therefore, this approach was considered further.

In the arithmetic overlay approach, the LSI for each pixel of the study area was obtained using the following equation,

$$\mathrm{LSI} = \sum_{l=1}^{t} (R_l) \tag{4}$$

The LSI values were found to lie in the range from 0.014 to 0.252. The observed mean (μ_o) and standard

Table 4
Classification scheme for neural network output values for LSZ mapping (ANN black box procedure)

Range of values	Landslide susceptibility zone
0–0.3	Very low susceptibility zone
0.3–0.5	Low susceptibility zone
0.5–0.7	Moderate susceptibility zone
0.7–0.9	High susceptibility zone
>0.9	Very high susceptibility zone

Table 5
Fuzzy ratings for different categories of factors

Thematic layers corresponding to factors	Categories	Number of pixels	Number of landslide pixels	Fuzzy rating (r_{ij})
Land use land cover	Agriculture land	35,692	85	0.0488
	Tea plantation	142,541	84	0.0243
	Thick forest	72,685	38	0.0229
	Sparse forest	131,088	65	0.0223
	Barren land	14,237	58	0.0638
	Habitation	10,341	9	0.0295
	Water	970	0	0
	River sand	1005	0	0
Lithology	Darjeeling gneiss	73,371	77	0.0324
	Feldspathic greywacke	45,938	61	0.0364
	Paro gneiss	247,242	158	0.0253
	Lingtse granite gneiss	20,926	15	0.0268
	Paro quartzite	12,154	14	0.0339
	Reyang quartzite	8089	14	0.0416
Slope	0–15°	51,380	23	0.0212
	15–25°	146,974	117	0.0282
	25–35°	144,495	131	0.0301
	35–45°	50,246	58	0.0340
	>45°	14,329	10	0.0264
Aspect	Flat	2072	0	0
	N	59,880	22	0.0192
	NE	45,077	32	0.0266
	E	52,868	73	0.0372
	SE	45,689	77	0.0411
	S	37,630	49	0.0361
	SW	29,860	20	0.0259
	W	55,132	26	0.0217
	NW	79,148	40	0.0225
Drainage buffer	25 m along 1st order drainage	116,168	102	0.0296
	25 m along 2nd order drainage	27,690	44	0.0399
Lineament buffer	0–125 m	146,761	243	0.0407
	125–250 m	108,929	35	0.0179
	250–375 m	72,380	36	0.0223
	375–500 m	41,360	17	0.0203
	>500 m	38,317	8	0.0144

deviation (σ_o) from the probability distribution curve of these LSI values are 0.150 and 0.024 respectively. The LSI values were divided into five distinct classes (susceptibility zones) with boundaries at ($\mu_o - 1.5\ m\sigma_o$), ($\mu_o - 0.5\ m\sigma_o$), ($\mu_o + 0.5\ m\sigma_o$) and ($\mu_o + 1.5\ m\sigma_o$) where m is a positive, non-zero value (Saha et al., 2005) which controls in fixing the most appropriate boundaries within the LSI range for landslide susceptibility classes. This classification scheme was adopted to fix the boundaries of classes statistically and also to avoid the subjectivity in arbitrarily selecting the boundaries of classes as was done in the conventional procedure.

Several LSZ maps of the study area were prepared for different values of m. The cumulative percentage of landslide occurrences in various susceptibility zones ordered from VHS to VLS were plotted against the cumulative percentage of area of the susceptibility zones for these LSZ maps. These curves, defined as success rate curves (Chung and Fabbri, 1999; Lu and An, 1999; Lee et al., 2002b), were used to select the appropriate value of m to decide the suitability of a LSZ map. Five representative success rate curves corresponding to $m=0.8$, 1.0, 1.1, 1.2 and 1.4 are shown in Fig. 7a. The suitability of any LSZ map can be judged by the fact that more percentage of landslides must occur in VHS zone as compared to other zones. It can be seen from Fig. 7a that for 10% of the area in VHS zone, the curves corresponding to $m=0.8$, 1.0, 1.1, 1.2 and 1.4 show the landslide occurrences of 42.7%, 47.7%, 48.6%, 47.3% and 41.8% respectively. Hence, for the first 10% area, the curve corresponding to $m=1.1$ has the highest success rate. Based on this analysis, the LSZ map corresponding to $m=1.1$ appears to be the most appropriate one for the study area. Accordingly, the landslide susceptibility class

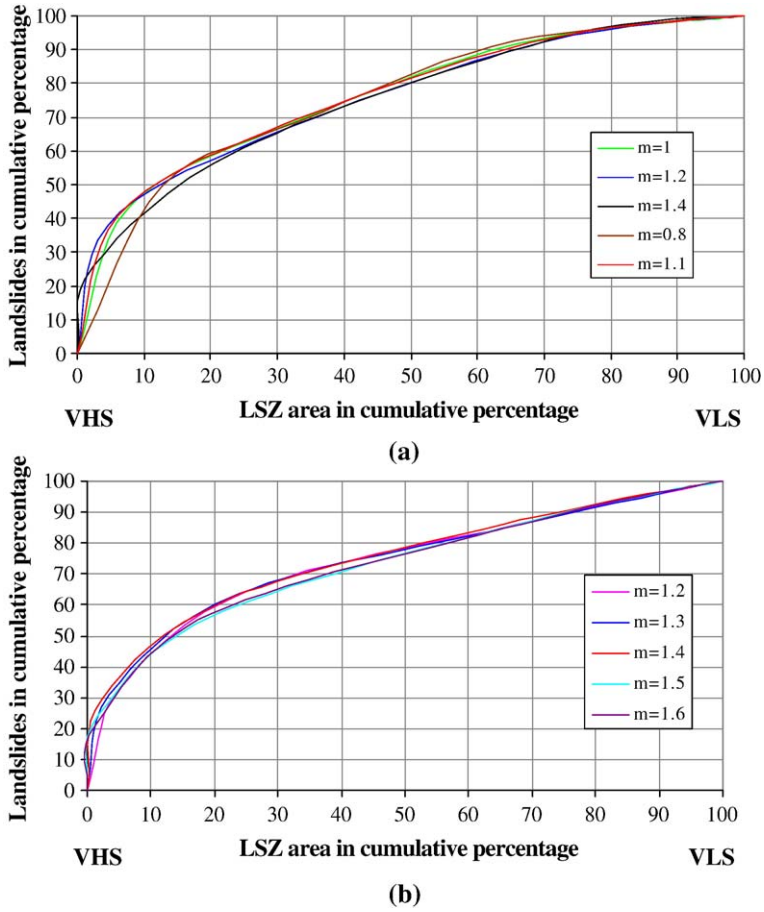

Fig. 7. Success rate curves to select the appropriate LSZ map. (a) Fuzzy set based procedure. (b) Combined neural and fuzzy procedure.

boundaries were fixed at LSI values of 0.110, 0.136, 0.163 and 0.189. The LSZ map (referred here as Map III) thus produced is shown in Fig. 5c. The area covered by different landslide susceptibility zones and the area of landslides occupied per class are also given in Table 2.

5.3.2. Analysis of output LSZ Map III

The visual inspection of the LSZ Map III depicts an overall NNE–SSW zonation trend in the area. It has been observed that the southeast and east facing slopes are more prone to landslides than other slopes. Hence, it can be stated that there is a topographic control over this LSZ map. Further, the spatial correlation between the landslide distribution and the LSZ map shows that 41.0% of landslide area has predicted over 6.1% area of VHS zone. It can also be stated that 28.8% of the total area occupied by HS and VHS zones are able to predict 66.1% of the total landslide area (Table 2).

5.4. LSZ using combined neural and fuzzy procedure

5.4.1. Implementation

The combined neural and fuzzy approach involves three main stages:

1) determination of weights of thematic layers through ANN connection-weight approach
2) determination of ratings for categories of thematic layers using cosine amplitude method
3) integration of ratings and weights using GIS to arrive at the LSZ map.

The methodology for LSZ using this procedure is shown in Fig. 8.

A feed forward back-propagation multi-layer ANN with one input layer, two hidden layers and one output layer was considered to determine the weights of the causative factors. The number of neurons in the input layer equals the number of input thematic layers. The

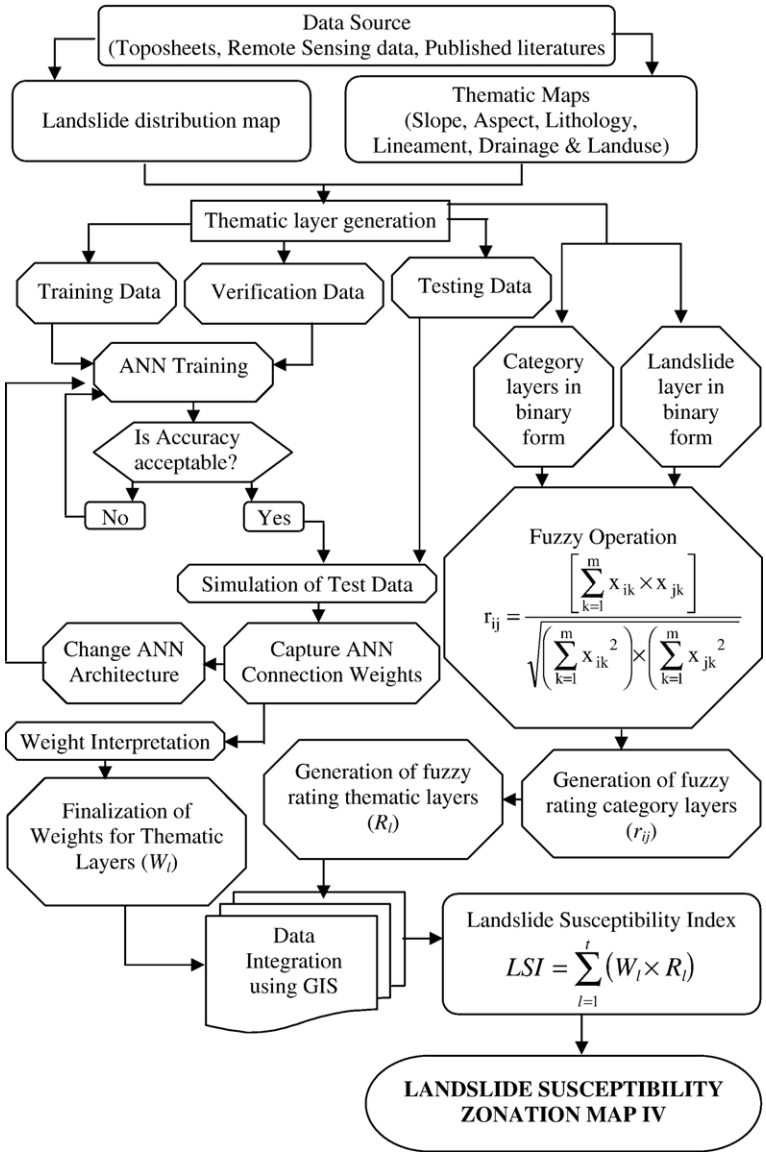

Fig. 8. Flow diagram showing the combined neural and fuzzy procedure for LSZ mapping.

data at each neuron of the input layer correspond to the weighted normalized rating or r_{ij} of the corresponding category (i.e., last column of Table 5). The output layer consists of a single neuron representing presence or absence of landslide for a pixel. Hence, the output value is either 0 or 1. The number of neurons in the hidden layers is varied by running the networks several times to achieve the desired training and testing data accuracies.

One set each of training, verification and testing data were randomly generated from the study area. The datasets consist of 226 pixels each, out of which 113 pixels were landslide pixels and rest 113 pixels were no landslide pixels. All the pixels in the datasets were mutually exclusive (Foody and Arora, 1997). The training dataset was used to train different network architectures while the verification dataset was used simultaneously with the training dataset to control the overtraining of the network. The testing dataset was used to evaluate the accuracy of the networks. Similar to ANN black box approach, the well known back-propagation learning algorithm was used to train the neural networks. 100 neural network architectures were designed, trained and tested. The training process was

initiated with arbitrary initial connection weights, which were constantly updated until an acceptable accuracy was reached. The training accuracy observed for the networks was of the order of 75% to 90%.

The final adjusted weights of the trained network were used to derive outputs of the testing data to evaluate the performance of the network. The testing accuracy observed for the networks was of the order of 60% to 70%. The adjusted weights of input–hidden, hidden–hidden and hidden–output connections for each network were captured for further analysis. Simple matrix multiplication was performed on these weight matrices to obtain a 6×1 weight matrix for each network which represents the weights of six causative factors (thematic layers) in this study. These causative factors were ranked according to the corresponding absolute weights for each network which means the higher the value of absolute weight, the more crucial the factor is for the occurrence of landslide. Considering all the 100 networks, the rank of a factor was decided based on the rank observed by the maximum number of networks (majority rule). Out of the 100 networks, 41 networks categorized lithology as rank 1 (most important), 31 networks categorized lineament as rank 2, 30 networks categorized slope as rank 3, 27 networks categorized aspect as rank 4, 33 networks categorized land use land cover as rank 5 and 49 networks categorized drainage as rank 6 (least important). These results are summarized in Table 6. Subsequently, the weighted normalized average of the weights of these networks at a scale of 0–10 for a particular factor was calculated and assigned as the weight of that factor (W_l) for the preparation of LSZ map. The weights thus obtained through ANN for all the factors are listed in Table 7.

It has been observed that the network with architecture 6/14/8/1 has been found to be the best for this dataset as it produced the same ranking pattern as mentioned above. The normalized weights obtained through this network at a scale from 0 to 10 for lithology, lineament buffer, slope, aspect, land use land cover and drainage buffer are 5.007, 1.996, 1.239, 0.933, 0.544 and 0.281 respectively (Fig. 2). These weights are almost at par with the weights obtained through majority rule (Table 7) which have been finally considered for producing the LSZ map.

The integration of 6 thematic layers representing the ratings for the categories (R_l) of the layers (obtained from fuzzy logic) and weights for the layers (W_l) (obtained from ANN) was performed by using simple arithmetic overlay operation in GIS. Hence, this procedure has been named here as combined neural and fuzzy weighting procedure. The LSI for each pixel of the study area was thus obtained by using the following equation.

$$\mathrm{LSI} = \sum_{l=1}^{t} (W_l \times R_l) \tag{5}$$

The LSI values were found to lie in the range from 0.030 to 0.408. The success rate curve approach was used to classify the LSI values into five different susceptibility zones to produce the LSZ map. Five representative success rate curves corresponding to $m = 1.2$, 1.3, 1.4, 1.5 and 1.6 are shown in Fig. 7b. It can be observed that for 10% of the area in VHS zone the curves corresponding to $m = 1.2$, 1.3, 1.4, 1.5 and 1.6 show the landslide occurrences of 43.9%, 45.6%, 46.7%, 43.3% and 43.9% respectively. Hence, for the first 10% area, the curve corresponding to $m = 1.4$ has the highest success rate. Based on this analysis, the LSZ map corresponding to $m = 1.4$ appears to be the most appropriate one for the study area. Accordingly, the boundaries of landslide susceptibility zones were fixed at LSI values of 0.208, 0.253, 0.299 and 0.344. The LSZ map (referred here as Map IV) thus produced is given in Fig. 5d. The area covered by different landslide susceptibility zones and the area of landslides occupied per class are also given in Table 2.

5.4.2. Analysis of output LSZ Map IV

In LSZ Map IV, only 2.3% of the total study area was occupied by the VHS zone and 30.1% of landslide area was predicted over this zone. It was also inferred that 22.5% of the total area occupied by VHS and HS zones could predict 62% of landslide area (Table 2). Further, this LSZ map has shown preferential distribution of higher landslide susceptibility zones along structural discontinuities (lineaments), which should indeed be the case. The buffer zones of lineaments have clearly indicated the VHS and HS zones in the north and southeast parts of the area. Therefore, it indicates the "ghost-effect" of lineaments on LSZ map as stated by Saha et al. (2005). Also, the Darjeeling gneiss rock type in southeastern part, feldspathic greywacke and Reyang quartzite in the northern part of the study area have clearly indicated moderate to very high susceptibility zones. Most of the lineaments up to 125 m buffer zone in these rock types have indicated high and very high susceptibility zones. Hence, it depicts the importance of lithology (i.e., rock types) as well as lineaments on the LSZ.

6. Comparative analysis and discussion

The LSZ maps were prepared using four different weighting procedures in a GIS-based approach. The

Table 6
Ranks of factors based on majority rule in combined neural and fuzzy weighting procedure (the number represents the number of artificial neural networks out of 100 networks categorizing a factor with respect to a particular rank and the rank corresponding to the maximum number of neural networks for a factor represents the final rank of that factor)

Factors	Number of networks						Final rank (majority rule)
	Rank 1	Rank 2	Rank 3	Rank 4	Rank 5	Rank 6	
Land use land cover	1	8	10	22	33	26	5
Lithology	41	21	12	10	10	6	1
Slope	23	24	30	9	9	5	3
Aspect	13	15	22	27	12	11	4
Drainage buffer	0	1	2	17	31	49	6
Lineament buffer	22	31	24	15	5	3	2

comparative analysis of different LSZ maps has been described below.

The pattern of percent area distribution of susceptibility classes in different LSZ maps prepared in this study appears to be quite similar to that obtained in other LSZ studies in the Himalayan regions (Gupta et al., 1999; Arora et al., 2004; Sarkar and Kanungo, 2004; Saha et al., 2005). However, in the LSZ Map I prepared using conventional weighting procedure, the LS zone occupied the maximum percent area (34.9%) in comparison to the MS zone which occupied 30.2% area.

Further, the VHS zone in the LSZ Map IV occupied 2.3% of the total study area, whereas in all other LSZ maps the area occupied by VHS is more than 6% of total area. Subsequently, the landslide distribution map was spatially cross-checked with all the four LSZ maps. The landslide distribution in the VHS and HS zones of LSZ maps (Table 2) indicate that the LSZ maps produced by fuzzy and combined neural and fuzzy procedures could predict more landslides in these zones as compared to other two LSZ maps.

Moreover, it can also be observed that the LSZ map produced by combined neural and fuzzy procedure shows preferential distribution of higher landslide susceptibility zones along structural discontinuities (lineaments) as compared to other LSZ maps, which may indeed be the case. Overall, the buffers of lineaments have left traces on the LSZ map. Owing to the landslide susceptibility of the terrain, the lineaments ought to leave some traces (termed as "ghost effect" in Saha et al., 2005) on the LSZ map.

Furthermore, the Darjeeling gneiss rock type in southeastern part, feldspathic greywacke and Reyang quartzite in the northern part of the study area have clearly indicated MS to VHS zones. Most of the lineaments up to 125 m buffer zone in these rock types have indicated HS and VHS zones. Hence, it depicts the importance of lithology (i.e., rock types) as well as lineaments on LSZ.

On the basis of these results, it can be concluded that the LSZ map derived from combined neural and fuzzy weighting procedure appears to be the best amongst all the weighting procedures and may thus be a useful way of assigning weights to the factors in an objective manner thereby minimizing the subjectivity.

7. Conclusions

In this study, four different weighting procedures viz. conventional based on subjective weighting, ANN black box, fuzzy logic and combined neural and fuzzy were applied for LSZ mapping in part of Darjeeling Himalayas and a comparative evaluation was carried out. The combined neural and fuzzy weighting integration produced the most accurate LSZ map. This may be attributed to the following reasons:

1) It represents an objective approach where weights for factors are determined through ANN connection weight approach and ratings of the categories of factors are determined through cosine amplitude similarity method based on fuzzy relation concept.
2) The LSZ map reflects preferential distribution of higher landslide susceptibility zones along lineaments which may indeed be the case.

Table 7
Weights of thematic layers derived through ANN (combined neural and fuzzy weighting procedure)

Thematic layers	ANN derived weights
Lithology	4.807
Lineament buffer	2.113
Slope	1.318
Aspect	1.065
Land use land cover	0.495
Drainage buffer	0.202

3) It delineates a relatively small area (only 2.3% of total area) for VHS zone, which can be more meaningful for practical applications.

Therefore, the integration of different factors in GIS environment using the combined neural and fuzzy weighting procedure may serve as one of the key objective approaches in this direction because of the fact that it can narrow down the potential susceptibility zones in a meaningful way for planning future developmental activities and implementation of disaster management programmes in hilly terrains.

References

Acharya, S.K., 1989. The Daling Group, its nomenclature, tectonostratigraphy and structural grain: with notes on their possible equivalents. Geol. Surv. Indones., Spec. Publ. 22, 5–13.

Aleotti, P., Chowdhury, R., 1999. Landslide Hazard Assessment: Summary, Review and New Perspectives. Bull. Eng. Geol. Environ. 58, 21–44.

Anbalagan, R., 1992. Landslide Hazard Evaluation and Zonation Mapping in Mountainous Terrain. Eng. Geol. 32, 269–277.

Arora, M.K., Das Gupta, A.S., Gupta, R.P., 2004. An artificial neural network approach for landslide hazard zonation in the Bhagirathi (Ganga) Valley, Himalayas. Int. J. Remote Sens. 25 (3), 559–572.

Atkinson, P.M., Tatnall, A.R.L., 1997. Neural networks in remote sensing. Int. J. Remote Sens. 18, 699–709.

Chi, K.-H., Park, N.-W., Lee, K., 2002a. Identification of Landslide Area using Remote Sensing Data and Quantitative Assessment of Landslide Hazard. Proc. IEEE Int. Geosciences and Remote Sensing Symp., 19 July, Toronto, Canada.

Chi, K.-H., Park, N.-W., Chung, C.-J., 2002b. Fuzzy logic integration for landslide hazard mapping using spatial data from Boeun, Korea. Proc. of Symp. on Geospatial Theory, Processing and Applications, Ottawa.

Chung, C.-J.F., Fabbri, A.G., 1999. Probabilistic prediction models for landslide hazard mapping. Photogramm. Eng. Remote Sensing 65, 1389–1399.

Clerici, A., Perego, S., Tellini, C., Vescovi, P., 2002. A procedure for landslide susceptibility zonation by the conditional analysis method. Geomorphology 48, 349–364.

Congalton, R.G., 1991. A review of assessing the accuracy of classifications of remotely sensed data. Remote Sens. Environ. 37, 35–46.

Davis, J.C., 1986. Statistics and Data Analysis in Geology. John Wiley & Sons, New York. 646 pp.

Dhakal, A.S., Amada, T., Aniya, M., 2000. Landslide hazard mapping and its evaluation using GIS: an investigation of sampling schemes for a grid-cell based quantitative method. Photogramm. Eng. Remote Sensing 66 (8), 981–989.

Dubois, D., Prade, H., 1980. Fuzzy Sets and Systems: Theory and Applications. Academic Press, New York.

Elias, P.B., Bandis, S.C., 2000. Neurofuzzy systems in landslide hazard assessment. Proc. 4th Int. Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Science, pp. 199–202.

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

Foody, G.M., Arora, M.K., 1997. An evaluation of some factors affecting the accuracy of classification by an artificial neural network. Int. J. Remote Sens. 18, 799–810.

Freund, J.E., 1992. Mathematical Statistics, Fifth Edition. Printice-Hall of India Pvt. Ltd., New Delhi, India.

Gomez, H., Kavzoglu, T., 2005. Assessment of shallow landslide susceptibility using artificial neural networks in Jabonosa River Basin, Venezuela. Eng. Geol. 78 (1–2), 11–27.

Gong, P., 1996. Integrated analysis of spatial data for multiple sources: using evidential reasoning and artificial neural network techniques for geological mapping. Photogramm. Eng. Remote Sensing 62, 513–523.

Gorsevski, P.V., Gessler, P.E., Jankowski, P., 2003. Integrating a fuzzy *k*-means classification and a bayesian approach for spatial prediction of landslide hazard. J. Geogr. Syst. 5, 223–251.

Gupta, R.P., 2003. Remote Sensing Geology, 2nd Edition. Springer-Verlag, Berlin Heidelberg, Germany.

Gupta, R.P., Joshi, B.C., 1990. Landslide Hazard Zonation using the GIS Approach—A case Study from the Ramganga Catchment, Himalayas. Eng. Geol. 28, 119–131.

Gupta, V., Sah, M.P., Virdi, N.S., Bartarya, S.K., 1993. Landslide Hazard Zonation in the Upper Satlej Valley, District Kinnaur, Himachal Pradesh. J. Himal. Geol. 4, 81–93.

Gupta, R.P., Saha, A.K., Arora, M.K., Kumar, A., 1999. Landslide Hazard Zonation in a part of Bhagirathy Valley, Garhwal Himalayas, using integrated Remote Sensing—GIS. J. Himal. Geol. 20 (2), 71–85.

Hagan, M.T., Menhaj, M., 1994. Training feedforward networks with the Marquardt algorithm. IEEE Trans. Neural Netw. 5 (6), 989–993.

Hagan, M.T., Demuth, H.B., Beale, M.H., 1996. Neural Network Design. PWS Publishing, Boston, MA.

Haykin, S., 1999. Neural Networks: A Comprehensive Foundation, Second edition. Prentice Hall, New Jersey.

He, Y.P., Xie, H., Cui, P., Wei, F.Q., Zhong, D.L., Gardner, J.S., 2003. GIS-based hazard mapping and zonation of debris flows in Xiaojiang Basin, Southwestern China. Environ. Geol. 45, 286–293.

Kanungo, D.P., Arora, M.K., Gupta, R.P., Sarkar, S., 2005. GIS-based landslide hazard zonation using neuro-fuzzy weighting. Proc. 2nd Ind. Int. Conf. on Artificial Intelligence (IICAI-05), Pune, pp. 1222–1237.

Lan, H.X., Zhou, C.H., Wang, L.J., Zhang, H.Y., Li, R.H., 2004. Landslide hazard spatial analysis and prediction using GIS in the Xiaojiang Watershed, Yunnan, China. Eng. Geol. 76, 109–128.

Lee, S., Choi, J., Chwae, U., Chang, B., 2002a. Landslide susceptibility analysis using weight of evidence. Proc. of IEEE Int. Geosciences and Remote Sensing Symposium, 19 July, Toronto, Canada.

Lee, S., Choi, J., Min, K., 2002b. Landslide susceptibility analysis and verification using the bayesian probability model. Environ. Geol. 43, 120–131.

Lee, S., Ryu, J., Won, J., Park, H., 2004. Determination and application of the weights for landslide susceptibility mapping using an artificial neural network. Eng. Geol. 71, 289–302.

Lin, M.-L., Tung, C.-C., 2003. A GIS-based Potential Analysis of the Landslides induced by the Chi-Chi Earthquake. Eng. Geol. 71, 63–77.

Lu, P.F., An, P., 1999. A metric for spatial data layers in favorability mapping for geological events. IEEE Trans. Geosci. Remote Sens. 37, 1194–1198.

Mehrotra, G.S., Sarkar, S., Kanungo, D.P., Mahadevaiah, K., 1996. Terrain analysis and spatial assessment of landslide hazards in parts of Sikkim Himalaya. Geol. Soc. India 47, 491–498.

Metternicht, G., Gonzalez, S., 2005. FUERO: foundations of a fuzzy exploratory model for soil erosion hazard prediction. Environ. Model. Softw. 20 (6), 715–728.

Nagarajan, R., Mukherjee, A., Roy, A., Khire, M.V., 1998. Temporal remote sensing data and GIS application in landslide hazard zonation of part of Western Ghat, India. Int. J. Remote Sens. 19, 573–585.

Olden, J.D., Joy, M.K., Death, R.G., 2004. An accyrate comparison of methods for quantifying variable importance in artificial neural networks using simulated data. Ecol. Model. 178, 389–397.

Pachauri, A.K., Pant, M., 1992. Landslide hazard mapping based on geological attributes. Eng. Geol. 32, 81–100.

Paola, J.D., Schowengerdt, R.A., 1995. A review and analysis of backpropagation neural networks for classification of remotely sensed multi-spectral imagery. Int. J. Remote Sens. 16, 3033–3058.

Ripley, B., 1996. Pattern Recognition and Neural Networks. Cambridge Univ. Press, Cambridge.

Ross, T.J., 1995. Fuzzy Logic with Engineering Applications. McGraw-Hill, New York.

Saha, A.K., Gupta, R.P., Arora, M.K., 2002. GIS-based landslide hazard zonation in a part of the Himalayas. Int. J. Remote Sens. 23, 357–369.

Saha, A.K., Gupta, R.P., Sarkar, I., Arora, M.K., Csaplovics, E., 2005. An approach for GIS-based statistical landslide susceptibility zonation—with a case study in the Himalayas. Landslides 2, 61–69.

Sarkar, S., Kanungo, D.P., 2004. An integrated approach for landslide susceptibility mapping using remote sensing and GIS. Photogramm. Eng. Remote Sensing 70 (5), 617–625.

Sarkar, S., Kanungo, D.P., Mehrotra, G.S., 1995. Landslide hazard zonation: a case study in Garhwal Himalaya, India Moun. Res. Dev. 15 (4), 301–309.

Schalkoff, R.J., 1997. Artificial Neural Networks. Wiley, New York.

Schuster, R., 1996. Socioeconomic significance of landslides. In: Turner, A.K., Schuster, R.L. (Eds.), Landslides: Investigation and Mitigation: Special Report, vol. 247. National Academic Press, Washington, DC, pp. 12–36.

Sinha, B.N., Varma, R.S., Paul, D.K., 1975. Landslides in Darjeeling district (West Bengal) and adjacent areas. Bull. Geol. Surv. India, Ser. B 36, 1–45.

Strahler, A.N., 1964. Quantitative geomorphology of basins and channel networks. In: Chow, V.T. (Ed.), Handbook of Applied Hydrology. Mcgraw Hill, New York.

Suzen, M.L., Doyuran, V., 2004. Data driven bivariate landslide susceptibility assessment using geographical information systems: a method and application to Asarsuyu Catchment, Turkey. Eng. Geol. 71, 303–321.

Tangestani, M.H., 2003. Landslide susceptibility mapping using the fuzzy gamma operation in a GIS, Kakan Catchment Area, Iran. Proc. of Map India Conference.

van Westen, C.J., 1994. GIS in landslide hazard zonation: a review, with examples from the Andes of Colombia. In: Price, M., Heywood, I. (Eds.), Mountain Environments and Geographic Information System. Taylor & Francis, Basingstoke, pp. 135–165.

Varnes, D.J., 1984. Landslide Hazard Zonation: A Review of Principles and Practice. UNESCO, Paris, pp. 1–63.

Virdi, N.S., Sah, M.P., Bartarya, S.K., 1997. Mass wasting, its manifestations, causes and control: some case histories from Himachal Himalaya. In: Agarwal, D.K., Krishna, A.P., Joshi, V., Kumar, K., Palni, M.S. (Eds.), Perspectives of Mountain Risk Engineering in the Himalayan Region. Gyanodaya Prakashan, Nainital, pp. 111–130.

www.em-dat.net EM-DAT: The OFDA/CRED International Disaster Database. Universite Catholique de Louvain, Brussels, Belgium.

Yesilnacar, E., Topal, T., 2005. Landslide susceptibility mapping: a comparison of logistic regression and neural networks methods in a medium scale study, Hendek region (Turkey). Eng. Geol. 79, 251–266.

Zadeh, L.A., 1973. Outline of a New Approach to the Analysis of Complex Systems and Decision Processes. IEEE Trans. Syst. Man Cybern SMC-3 (1), 28–46.

Zhou, W., 1999. Verification of the nonparametric characteristics of backpropagation neural networks for image classification. IEEE Trans. Geosci. Remote Sens. 37, 771–779.