Nat Hazards (2015) 75:1555–1576
DOI 10.1007/s11069-014-1378-y

ORIGINAL PAPER

Development and application of Shannon's entropy integrated information value model for landslide susceptibility assessment and zonation in Sikkim Himalayas in India

L. P. Sharma · Nilanchal Patel · M. K. Ghose · P. Debnath

Received: 6 January 2013 / Accepted: 9 August 2014 / Published online: 3 September 2014

Abstract The state of Sikkim in India has many steep slopes and has been susceptible to landslides. Since 1968 there have been innumerable losses of lives and properties due to landslides. There is an urgent need for advance assessment of degrees of vulnerability and delineation of the most vulnerable zone for shifting of the population and infrastructure to a safer zone. The identification and formulation of most suitable and acceptable method for such assessment is still nascent and research based. In this study, an attempt has been made to integrate the concept of Shannon's entropy with the information value-based statistical model to evaluate the landslide susceptibility in the study area and assess the improvement made through the integration of Shannon's entropy by comparing the results with the landslide susceptibility determined from the information value-based statistical model alone. Initially, the thematic layers pertaining to all the causative parameters were overlaid with the help of geographical information system that resulted in the formation of 78,256 numbers of polygons for each one of which landslide susceptibility was determined. For each polygon, the total landslide information value (TLIV) was computed as the summation of the landslide information values determined for the individual sub-categories present within the respective polygons. Again for each polygon, the Shannon's entropy value of the individual parameters was multiplied with the summation of the landslide information values of all the sub-categories present within the respective parameters. The product values computed for the different causative parameters were summed up to

L. P. Sharma (✉)
National Informatics Centre, Tashiling Secretariat, Sikkim, India
e-mail: lp.sharma@nic.in

N. Patel
Department of Remote Sensing, Birla Institute of Technology Mesra, Ranchi, India

M. K. Ghose
Department of Computer Science, Sikkim Manipal Institute of Technology, Mazitar, Sikkim, India

P. Debnath
College of Horticulture and Forestry, Central Agriculture University, Pasighat, Arunachal Pradesh, India

determine the total landslide information value with entropy (TLIV_e). Finally, the entire study area was categorized into five zones of landslides susceptibility based on the TLIV and TLIV_e, respectively. The prediction accuracy of the landslides determined based on the landslide susceptibility derived from TLIV_e was found to be significantly high (91 %) as compared to that derived from TLIV (85 %) indicating the potential contribution of Shannon's entropy in the improved delineation of the landslide susceptibility zones.

Keywords Landslides · Susceptibility · Information value · Shannon's entropy

1 Introduction

Landslides have posed perennial problems in the Himalayas. Susceptibility analysis and zonation are very important and significant aspects of the disaster mitigation system through which the areas with different degrees of landslide susceptibility are segregated into various zones. With the enhanced capability of the computing devices in terms of storage, speed and flexibility, the geo-spatial technology has been potentially utilized to achieve improved understanding and characterization of the various causative parameters of landslides. The primary goal in landslide susceptibility analysis is to quantify the influence of each landslide causative factor as a piece of information contributing to the final predictive information. Landslide susceptibility analysis have been performed by many researchers and scientists all across the globe during the last two decades (Shu-Quin and Unwin 1992; Pachauri and Pant 1992; Van Westen et al. 1997; Pachauri et al. 1998; Uromeihy and Mahdavifar 2000; Patanakanog 2001; Sakellariou and Ferentinou 2001; Sivakumar and Mukesh 2002; Dhakal and Sidle 2002; Carro et al. 2003; George et al. 2007; Lee 2007; Jadda et al. 2009; Pradhan and Ahmed 2010). Several statistical models have also been developed and employed for performing such analysis. Logistic regression model was used for landslide mapping by Atkinson and Massari (1998), Lee (2004, 2005) and Ramakrishna et al. (2005) along with Frequency Ratio method (Lee and Sambath 2006; Lee and Pradhan 2006; Yilmaz 2009a; Avinash and Ashamanjari 2010) and information value method (Ramakrishna et al. 2005). Fuzzy Algebraic Function was used for landslide susceptibility modeling by Pistocchi et al. (2002), Ercanoglu and Gokceoglu (2004) and Lee (2007). Artificial neural network was used for landslide susceptibility study and modeling by Lee et al. (2003a, b, 2004, Yilmaz 2009b), Pradhan and Lee (2010) and many others. Knowledge-driven raster analysis for landslide study has been conducted by Gupta et al. (2009). The information value method, however, stands pivotal to all the methods.

The quantified information value for each of the causative factors and their sub-categories summed together will give a composite landslide susceptibility index. Higher the index value more vulnerable is the area to landslides. A thorough understanding of susceptibility is possible if the study area is divided into a number of elements based on the variation of the parameters that are considered, and the composite index value is computed for each of such elements. The occurrence of varying composite index values within the study area will enable the analyst to categories the polygon elements on the basis of them. To implement this idea, information value theory, a widely used statistical model is applied by many researchers in landslide susceptibility study (Yin and Yan 1988; Ramakrishna et al. 2005). The present investigation was carried out in the Rumtek Samdung area of Sikkim, India involving fourteen causative parameters identified in the area that were subcategorized into forty-eight

subclasses. In the same study area, the landslide vulnerability assessment and zonation was performed earlier using the same causative parameters and applying several methods. These include tri-variate approach (Sharma et al. 2009a), landslide susceptibility assessment based on soil characteristics (Sharma et al. 2009b, 2012b), ranking of causative parameters (Sharma et al. 2011), experts' weight and Shannon's entropy (Sharma et al. 2012a), landslide density of the causative parameters (Sharma et al. 2012c) and likelihood ratio approaches (Sharma et al. (2012d), Fuzzy Logic (Sharma et al. 2013).The present method of study introduces a new approach to perform the task of landslide susceptibility assessment. This approach integrates the concept of Shannon's entropy with the information value theory, to compute the landslide information values of the different polygons. The landslide susceptibility zonation determined through this new approach was compared with that derived from the application of the information value model alone in order to determine whether there occurs improvement through the new approach.

The objective of this study was to evaluate more robust method for landslide susceptibility assessment through accurate delineation of the study area into various zones with varying degrees of landslide susceptibility and try to improve the performance of earlier established information value model with the help of Shannon's entropy. The result of this study will give new insight for researchers and geo-scientist involved in geo-statistical modeling of trends, susceptibility and suitability, toward the integration of Shannon's entropy into the existing and earlier established methods for more accurate results.

2 Study area

The study area selected for this study the Rumtek Samdung area comprises of a sloppy stretch of around 26 villages. The area is surrounded of Gangtok town in the east, south district of Sikkim in the west, river Singtam in the south and a reserve forest area and some part of north district in the north (Fig. 1). At the central part of the area lies the famous *Tinjuray* hill which is a dense forest famous for availability of wide ranges of wild animals and attracts many adventure trackers. However, this reserve forest area is excluded for the study as no history of past landslides was found in the area. Within the study area, the rock types found were mostly a week combination of chlorite, phyllite and schist (Fig. 2), where as a small area was also characterized by a more stable Lingtse gneiss group of rocks. The slope in the area varied from 15 to 90 % where as in between the big terraces also lies the flat area that gives almost 0 % of slope. The land use land cover of the area is mostly cultivated, barren land, mixed forest or dense forest. Drainage lines are available every 3–4 km and most of them are heavily inundated during the monsoon. Two important road lines, the first one passing form Rumtek to Upper Samdung through Sang village and second one passing through Ranipool to Lower Samdung through Singtam, link the villages in the upper and lower belts of the hill slopes. However, lots of link roads are being newly constructed with unmanaged and muddy slopes on both sides of the road that adds the landslide susceptibility of the area. Many permanent landslides are seen in the study area (Fig. 3). Some of the landslides are stabilized after the first occurrence; where as many of them have become the permanent landslide spots.

2.1 Methodology

The spatial data required for the study were collected from various stake holders. Table 1 lists the various spatial data used in this study with their source and description. The digital

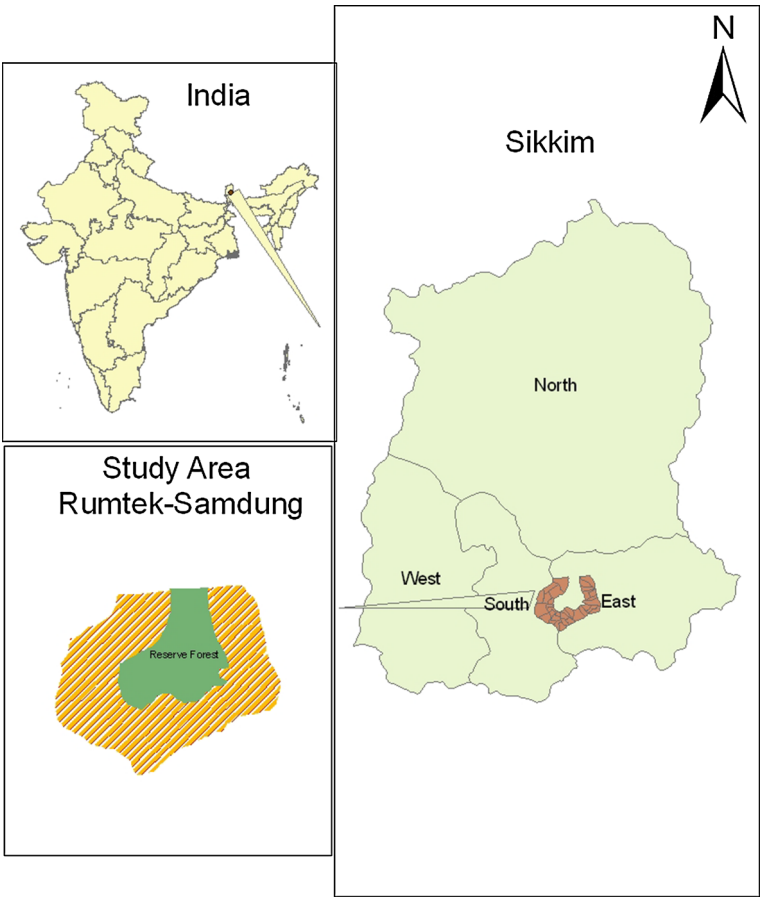

Fig. 1 Study area

elevation model (DEM) was developed from the contour maps and the GPS collected elevation data. The slope map and aspect map were derived from the DEM. The landslide spots, the road lines, drainage lines and land use and land covers were identified from the satellite images (Fig. 4). Other important landslide causative parameters were identified and their spatial layers were integrated into common spatial reference in geographical information system framework. Initially, the thematic layers corresponding to fourteen causative parameters were overlaid into one single layer using the overlay tool of the ArcGIS that yielded a final layer comprised of 78,256 polygons with each of the polygon associated with attributes containing information about the presence of all the fourteen causative parameters and their forty-eight sub-categories. Information value associated with each causative sub-category is determined as the logarithm of the ratio of the relative frequency of the landslides present within the sub-category to the relative frequency of the landslides present in all the sub-categories, i.e., in the entire study area. Information model is a statistical method based on the probability theory for spatial prediction of occurrence of an event from the parameter and event relationship. It has proved to be a very useful

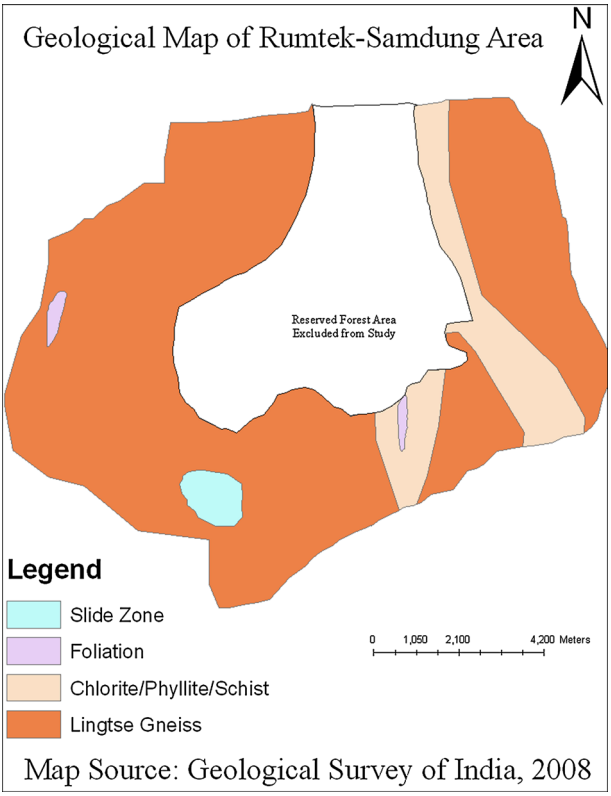

Fig. 2 Geological map of the study area

method for landslide susceptibility analysis and has been successfully used by several researchers (Yin and Yan 1988, Ramakrishna et al. 2005).

The information value I_i for a sub-category i can be expressed as:

$$I_i = \log \frac{S_i/N_i}{S/N} \tag{1}$$

where S_i = number of polygons with the landslide events in sub-category i, N_i = total number of polygons in the sub-category i, S = number of polygons with landslide events, N = total number of polygons

The ratio in the numerator (S_i/N_i) represents the landslide density in each sub-category and the ratio in the denominator (S/N) represents the landslide density in the study area; in each ratio the unit of measurement is the polygon).

Total information value (TLIV_j) in any of the jth polygon is given by

$$I_j = \sum_{i=1}^{M} X_{ji} I_i \tag{2}$$

where X_{ji} is the value of the parameter i which is equal to either 0 or 1 as defined earlier depending upon its presence or absence in the jth polygon.

Fig. 3 Landslides of Sirwani village

$j = 1,2,\ldots N$ where N is the total number of polygons in the study area and
$i = 1,2,3,\ldots M$ where M is the total number of parameters considered.

The above model was used to determine the total landslide information Value (TLIV) for each of the polygons that were created through the overlay of the fourteen spatial layers. Greater the TLIV, more vulnerable is the polygon to landslide. The entire study area was categorized into five zones of landslide susceptibility viz. least vulnerable zone, moderately vulnerable zone and the most vulnerable zone considering the TLIVs based on which a landslide susceptibility zonation map was prepared. The methodology adopted is depicted with a flow chart diagram in Fig. 5.

2.2 Calculation of landslide information values of parameters

Landslide information value of a parameter or sub-category informs about its degree of influence in causing the landslides. Landslide information values for all the sub-categories of the different parameters considered were computed using Eq. 1. The information values of the pedologic parameters are computed and tabulated in Table 1 where as for those of the non-pedologic parameters are computed and tabulated in Table 2. It is to be reiterated here that more the information value associated with the sub-categories of the parameter higher is its susceptibility to landslides. It can be seen from the tables that the sub-categories of soil parameters such as moderate deep, coarse loamy texture, shallow, somewhat excessively drained, high hydraulic conductivity, low stoniness and severe erosion have comparatively high information values indicating their higher susceptibility and conduciveness to landslides. Looking at the slope variable, 15–30 % of slope and above 60 % of slope have reported higher information values. Looking at the aspect, west, south and southeast are showing higher information values. Hills facing south and

Table 1 Computation of information values of pedologic parameters

Soil parameters	Parameter sub-categories	Variable	No. of polygons (Ni)	No. of polygons with landslides (Si)	Si/Ni	Information value associated with the sub-categories type [log (Si/Ni)/(S/N)]
Soil depth	Moderate deep (75–100 cm)	*X*1	21,565	199	0.009228	0.070
	Moderate shallow (50–75 cm)	*X*2	944	0	0	0.000
	Shallow (<50 cm)	*X*3	55,747	415	0.007444	−0.023
Drainage characteristics	Well drained	*X*4	32,284	200	0.006195	−0.103
	Some what excessively drained	*X*5	45,960	414	0.009008	0.060
	Excessively drained	*X*6	12	0	0	0.000
Texture	Fine loamy	*X*7	41,028	320	0.0078	−0.003
	Loamy	*X*8	25,863	187	0.00723	−0.035
	Coarse loamy	*X*9	11,365	107	0.009415	0.079
Hydraulic conductivity	Low	*X*10	41,028	320	0.0078	−0.003
	Moderate	*X*11	25,863	187	0.00723	−0.035
	High	*X*12	11,365	107	0.009415	0.079
Stoniness	High	*X*13	0	0	0	0.000
	Moderate	*X*14	16,116	123	0.007632	−0.012
	Low	*X*15	62,140	491	0.007902	0.003
Erosion	Low	*X*16	0	0	0	0.000
	Moderate	*X*17	53,523	395	0.00738	−0.027
	Severe	*X*18	24,733	219	0.008855	0.053
Surface texture	Fine loamy	*X*19	0	0	0	0.000
	Loamy	*X*20	55,546	419	0.007543	−0.017
	Coarse loamy	*X*21	22,710	195	0.008587	0.039

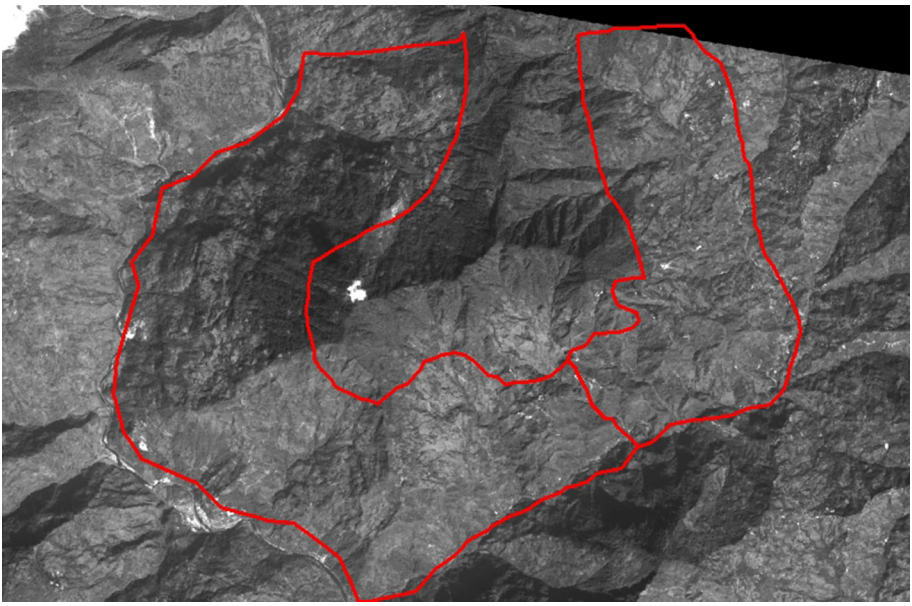

Fig. 4 Study area in Cartosat panchromatic image

southeast showing higher information values and being more susceptible to landslides was also reported earlier by Sarkar et al. (2008). In general, south facing slopes are said to have less vegetation density and more of erosional activities under other constant influencing parameters (Sinha et al. 1975).

2.3 Computation of TLIV

After overlaying spatial map layers pertaining to all the causative factors, the final resultant layer named as rsf.shp was a map containing 78,256 polygons. For each of the polygons, 48 variables named as $\underline{X}1$, $X2$, $X3$, etc. were introduced to denote the presence or absence of the respective parameters in the polygon. For example, if in a polygon, the soil depth was of moderate deep category then $X1 = 1$ else $X1 = 0$. Similarly if soil depth was of moderate shallow category then $X2 = 1$ else $X2 = 0$. In the same manner, the presence or absence of a parameter type was denoted by the values of $X1$, $X2$, $X3$,…,$X48$ as listed in the third column of Tables 1 and 2. Finally, for each of the polygons in the study area, TLIV, which is the total information value provided by all the influencing parameters existing in the polygon was calculated using Eq. 2. Hence higher the TLIV value, higher is the landslide susceptibility of the polygon. The TLIV calculated for all the 78,256 polygons ranged from -3.676 to -1.898 with a mean of -2.5833, median of -2.5643 and a standard deviation of 0.2747. The TLIV values were then grouped into five classes with varying order of their expected landslide susceptibility using the natural breaks (Jenks) method in ArcGIS. The polygons with TLIV ranging from -3.676 to -2.97 covering 5.56 sq. km of area were classified in the least vulnerable zone, the polygons with TLIV ranging from -2.969 to -2.715 covering 20.22 sq. km. area were classified in the low vulnerable zone, the polygons with TLIV ranging from -2.7149 to -2.508 covering an area of 26.38 sq. km in the moderately vulnerable zone, the polygons with TLIV ranging between

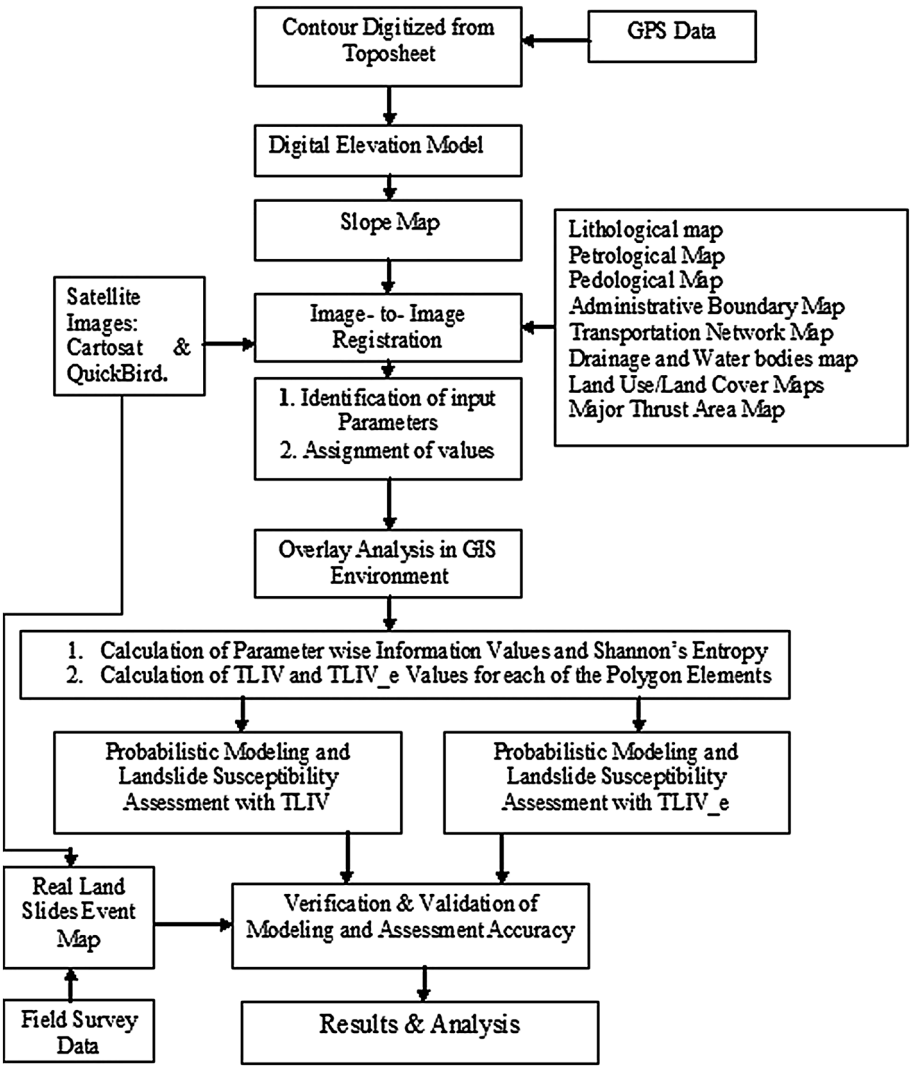

Fig. 5 Methodology flowchart

−2.5079 and −2.311 covering 24.28 sq. km. of area were classified in the highly vulnerable zone and polygons with TLIV ranging from −2.3109 to −1.898 covering 17.48 sq. km. of area were classified in the most vulnerable zone as depicted in Table 3.

2.4 Validation of TLIV-based zonation and the susceptibility analysis

Classification of polygon elements vis-à-vis zonation of the study area is done based on the statistically computed landslide information values of polygons that were further based on the landslide information values supplied by each of the sub-categories and ranges of parameters. It is important to compare the result of the zonation with the actual landslides.

Table 2 Computation of information values for non-pedologic parameters

Soil parameters	Parameter sub-categories	*X* Variable	No. of polygons (*Ni*)	No. of polygons with landslides (*Si*)	*Si*/*Ni*	Information value associated with the sub-categories type [log (*Si*/*Ni*)/(*S*/*N*)]
Geology	Lingtse gneiss	*X*22	10,302	76	0.007377	−0.027
	Chungthang sub-group	*X*23	0	0	0	0.000
	Chlorite phyllite/schist, sericite	*X*24	67,954	538	0.007917	−0.413
Foliation	Yes	*X*25	565	2	0.00354	−0.763
	No		77,691	612	0.007877	−0.415
Road in 40 m distance	Yes	*X*26	12,457	91	0.007305	−0.448
	No		65,769	523	0.007952	−0.411
Slope (%)	0–15	*X*27	11,505	91	0.00791	−0.413
	15–30	*X*28	13,022	115	0.008831	−0.366
	30–45	*X*29	17,993	141	0.007836	−0.417
	45–60	*X*30	14,802	95	0.006418	−0.504
	>60	*X*31	20,934	172	0.008216	−0.397
Land cover	Dense forest	*X*32	35,753	228	0.006377	−0.507
	Open forest	*X*33	2,258	31	0.013729	−0.174
	Scrub land	*X*34	0	0	0	0.000
	Cultivable/barren land	*X*35	40,245	355	0.008821	−0.366
Aspect	Flat	*X*36	9,368	82	0.008753	−0.369
	North	*X*37	7,371	41	0.005562	−0.566
	Northeast	*X*38	5,115	21	0.004106	−0.698
	East	*X*39	7,117	18	0.002529	−0.909
	Southeast	*X*40	9,567	73	0.00763	−0.429
	South	*X*41	11,628	116	0.009976	−0.313
	Southwest	*X*42	11,280	101	0.008954	−0.360
	West	*X*43	9,456	105	0.011104	−0.266
	Northwest	*X*44	7,354	57	0.007751	−0.422
Drainage in 30 m distance	Yes	*X*45	23,721	163	0.006872	−0.475
	No		54,535	451	0.00827	−0.394

The actual landslide map available in the form of polygon was overlaid earlier during the process of overlay analysis. Hence rsf.shp; the final overlaid spatial dataset containing 78,265 polygons contained a total of 614 polygons containing the instances of landslides. In order to validate the zonation results, the numbers of such polygons were counted in each of the susceptibility zones. Number of such polygons counted and the landslide density computed for each of the susceptibility zones is depicted in Table 4. As depicted in the table, the least vulnerable zone contained a total of 13 (2.12 %) polygons, low vulnerable zone contained 65 (10.59 %) polygons, moderately vulnerable zone contained 134 (21.82 %) polygons, highly vulnerable zone contained a total of 223 (36.32 %) polygons and the most vulnerable zone contained a total of 179 (29.15 %) polygons with the instances of past landslides. Figure 6 depicts the percentage of area and landslides in each of the susceptibility zones. As depicted from the figure, the percentage of landslide is less up to the moderately vulnerable zone in comparison with the percentage of area where as in the high and the most vulnerable zone the percentage of landslides is more than the percentage of area which also correlates and agrees with the fact that the higher vulnerable areas will have more number of occurrences of landslides as compared to the less vulnerable areas. Higher percentage of landslides in higher susceptibility zones is also depicted by line graphs in Fig. 7.

In order to further assess the relevance of zonation method and the result, landslide density is computed for each of the susceptibility zones as a ratio of number of landslide containing polygons to the total area in each of the zones. The computed landslide densities for each of the zones are tabulated in last column of Table 9. As depicted in the table, the landslide density constantly increases from the least value (2.34) at the least vulnerable zone to the highest value (10.24) at the most vulnerable zone. Figure 8 depicts the landslide densities in different zones with the help of a bar chart.

Hence, the prevalence of higher percentage of landslides and higher landslide densities in the higher susceptibility zones indicate the reliability of the zonation method. The zonation map prepared for the study area based on this zonation method is depicted in Fig. 9. The total numbers of polygons with landslides in the three highest vulnerable zone, i.e., moderate, high and most vulnerable zones accounted for 536 (87 %) out of a total of 614. Hence, the prediction accuracy derived from the information value model for performing the task of landslide prediction for the Rumtek Samdung study area is found to be 87 %.

2.5 Application of Shannon's entropy on information value model

Shannon's entropy (1948) is measure of uncertainty associated with a random variable defined as the information content in any system. The average uncertainty defined as Shannon's entropy and denoted with function H_N for an event or value is given by the following expression:

$$H_N = -\sum_{i=1}^{n} P_i \log P_i \tag{3}$$

where P_i is the probability of occurrence of the ith event or value; for example, in case of soil, i represents the three different types of soil textures viz. fine loamy, loamy and coarse loamy.

Table 3 TLIV-based susceptibility classification

Ranges of TLIV	Area (Sq Km)	No. of polygons	Type of zones
−3.676 to −2.97	5.56 (5.92 %)	7,500 (9.58 %)	Least vulnerable zone
−2.969 to −2.715	20.22 (21.53 %)	16,598 (21.21 %)	Low vulnerable zone
−2.7149 to −2.508	26.38 (28.09 %)	20,338 (25.99 %)	Moderately vulnerable zone
−2.5079 to −2.311	24.28 (25.85 %)	20,997 (26.83 %)	Highly vulnerable zone
−2.3109 to −1.898	17.48 (18.61 %)	12,823 (16.39 %)	Most vulnerable zone
Total	93.92 (100 %)	78,256 (100 %)	

H_N denotes the Shannon's entropy for thirteen control parameters; H_1 for soil depth, H_2 for soil drainage behavior and so on as shown in Table 5.

Shannon's entropy was earlier introduced and applied in landslides susceptibility assessment to balance the influence of variation and prevalence of parameters within the study area based on the knowledge-driven weights assigned to the sub-categories of the different parameters (Sharma et al. 2012a). It is once again applied here in order to refine the computed information value of the various sub-categories of the different parameters with respect to their spatial variation and prevalence on the field. The total landslide information value with entropy (TLIV_e) for each of the polygons lying within the study area can be computed by further improving Eq. 2 in the following way:

$$\text{TLIV\_e} = \sum_{i=1}^{M} X_{ji} I_i H_i \tag{4}$$

where X_{ji} and I_i have the same meaning as given in Eq. (2) and H_i is the Shannon's entropy of the ith parameter. Shannon's entropy computed for different causative parameters are shown in Table 5.

The TLIV_e computed for 78,256 polygons present in the final resultant layer of the data set ranged from −5.4289 to −1.815 with a mean of −3.0427, median of −2.8811 with a standard deviation of .6495. Then the values of TLIV_e computed for the different polygons were grouped into five categories of landslide susceptibility applying the natural breaks (Jenks) method available with the ArcGIS. The different zones are named as the least, low, moderate, high and most vulnerable zones in the increasing order of their susceptibility to landslides. Higher the value of TLIV_e within a polygon, higher is its susceptibility to landslides. The area and number of polygons lying within the different susceptibility zones are shown in Table 6.

2.6 Verification and validation of TLIV_e-based zonation

For the verification and validation of the final landslide susceptibility map generated on the basis of the TLIV_e, the various statistics pertaining to the occurrences of landslides in the different susceptibility zones were compared with their actual occurrences in the respective zones. Out of 614 polygons with the instances of landslide events of the past, it was found that 14 (3.42 %) polygons with landslide events are falling in the least vulnerable zone, 18 (5.54 %) polygons with landslides are in the low susceptibility zone, 71 (18.73 %) polygons are in the moderately vulnerable zone, 155 (33.55 %) polygons with landslides are in the highly vulnerable zone and 356 (38.76 %) of polygons with landslides are in the most vulnerable zone, respectively, as depicted in Table 6. The percentage of area, number of

Table 4 Computation of landslide densities in TLIV zonation

Type of zones	No. of polygons with landslides	Landslide density	Prediction accuracy
Least vulnerable	13 (2.12 %)	2.34	87 %
Low vulnerable	65 (10.59 %)	3.21	
Moderate vulnerable	134 (21.82 %)	5.08	
High vulnerable	223 (36.32 %)	9.18	
Most vulnerable	179 (29.15 %)	10.24	
Total	614 (100 %)	6.54	

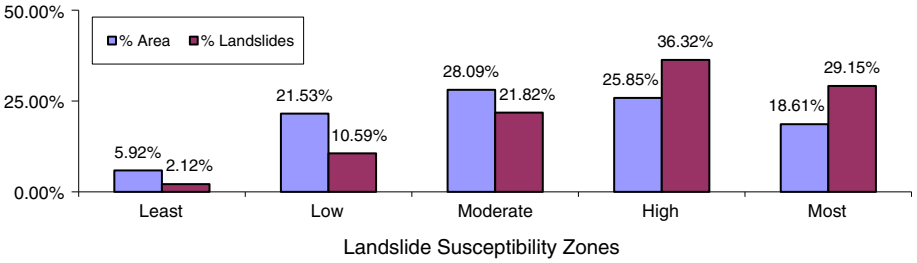

Fig. 6 *Bar chart* for area and landslides in TLIV-based zonation

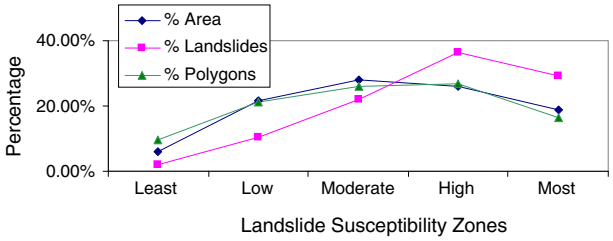

Fig. 7 *Line graph* for percentage of area, polygons and landslides in TLIV-based zonation

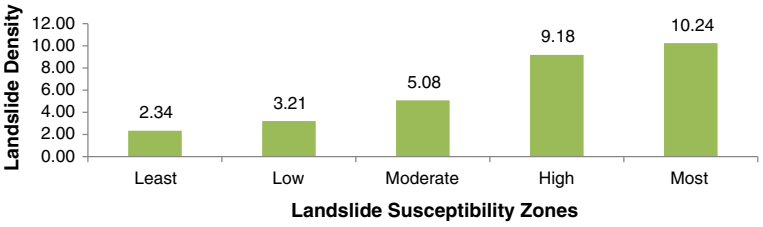

Fig. 8 *Bar graph* of landslide densities

polygons and landslides lying within each susceptibility zone depicted by bar chart in Figs. 10 and 11 as line graph more clearly depicts the rise of intensity of landslide from lower to higher susceptibility zones. Analysis of the line graphs exhibit constant but

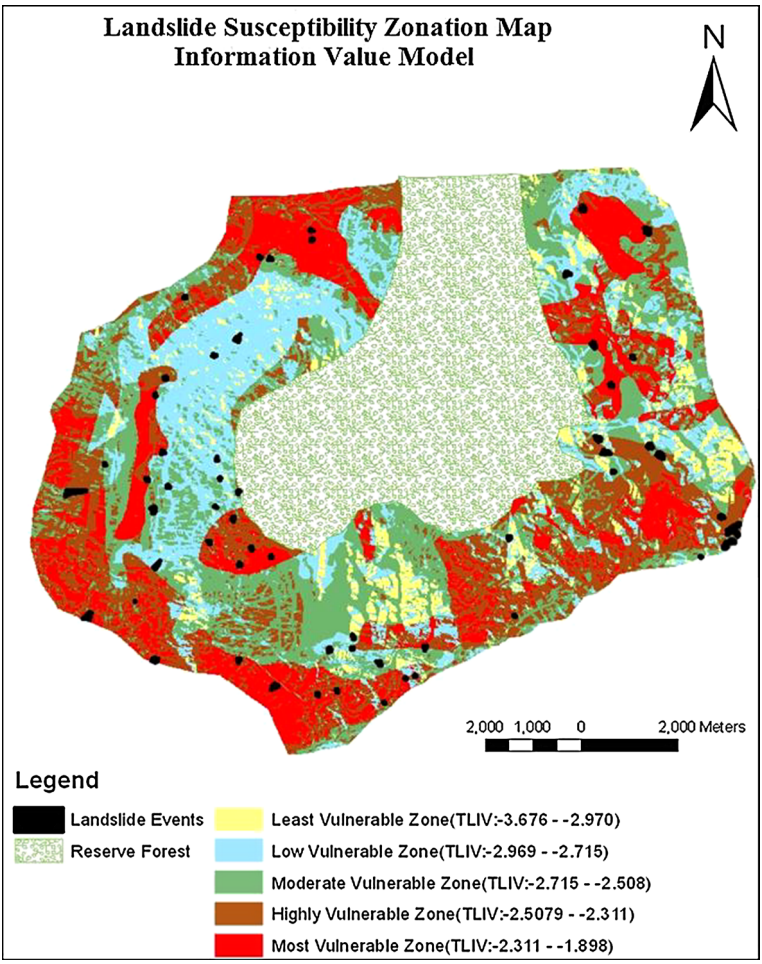

Fig. 9 Landslide susceptibility zonation with TLIV

significant increase in the frequency of the landslide events from the least susceptibility zone to the most vulnerable zone. The prevalence of considerable improvement in the occurrence of landslides events in the upper two susceptibility zone (72 %) as compared to their occurrence in the lower three susceptibility zones together, i.e., (28 %) indicates significant influence of the Shannon's entropy-based TLIV, i.e., TLIV_e on the susceptibility zonation. Landslide densities computed for the different susceptibility zones as the ratio of the number of polygons with landslides to the area covered by the respective susceptibility zones exhibit smooth but noticeable increase from the least vulnerable zone (3.31) to the most vulnerable zone (9.13) with the low, moderate and high vulnerable zones associated with 4.74, 6.18 and 5.77 landslide densities, respectively. The landslide densities computed for the different susceptibility zones are depicted in Table 7 and by a bar graph in Fig. 12. In order to validate the results of the landslide susceptibility determined from the Shannon's entropy-based TLIV, i.e., TLIV_e prediction accuracy for this technique has been computed as the percentage of the total number of polygons with landslides

Table 5 Calculation of Shannon's entropy for different parameters

Sl. No.		Parameter types	Frequency (No. of polygons, n)	Probability ($P = n/78{,}256$*)	$-(P^*\log_2 P)$	Entropy (H) ($\Sigma P^*\log_2 P$)
Soil parameters						
1	Soil depth	Moderate deep	21,565	0.275569924	0.51242493	0.937871879
		Moderate shallow	944	0.012062973	0.07688059	
		Shallow	55,747	0.712367103	0.34856635	
2	Drainage characteristics	Well drained	32,284	0.412543447		0.452887661
		Somewhat excessively drained	45,960	0.58730321	0.45094466	
		Excessively drained	12	0.000153343	0.001943	
3	Depth texture	Fine loamy	41,028	0.524279288	0.48841467	1.420572329
		Loamy	25,863	0.330492231	0.52789912	
		Coarse loamy	11,365	0.145228481	0.40425853	
4	Hydraulic conductivity	Low	41,028	0.524279288	0.48841467	1.420572329
		Moderate	25,863	0.330492231	0.52789912	
		High	11,365	0.145228481	0.40425853	
5	Stoniness	High	0			0.733649173
		Moderate	16,116	0.205939481	0.46948181	
		Low	62,140	0.794,060,519	0.26,416,736	
6	Erosional characteristics	Low	0	0		0.900036864
		Moderate	53,523	0.683947557	0.37483225	
		Severe	24,733	0.316052443	0.52520461	
7	Surface texture	Fine loamy	0	0		0.868981253
		Loamy	55,546	0.70979861	0.35100843	
		Coarse loamy	22,710	0.29020139	0.51797282	

Table 5 continued

Sl. No.		Parameter types	Frequency (No. of polygons, *n*)	Probability ($P = n/78{,}256$*)	$-(P^*\log_2 P)$	Entropy (H) ($\Sigma P^*\log_2 P$)
Non-soil parameters						
8	Lithology	Lingtse gneiss	10,302	0.131644858	0.38509767	0.56193202
		Chungthang sub-group	0	0		
		Chlorite phyllite/schist, sericite	67,954	0.868355142	0.17683435	
9	Foliation	Yes	565	0.007219894	0.05136093	0.06173934
		No	77,691	0.992780106	0.01037841	
10	40 m buffer of roads	Yes	12,457	0.159182682	0.42203224	0.632807732
		No	65,769	0.84043396	0.21077549	
11	Slope (%)	0–15	11,505	0.147017481	0.40664159	2.288087569
		15–30	13,022	0.166402576	0.43052512	
		30–45	17,993	0.229924862	0.48761674	
		45–60	14,802	0.189148436	0.45441195	
		>60	20,934	0.267506645	0.50889217	
12	Land cover	Dense forest	35,753	0.456872316	0.51632834	1.157309009
		Open forest	2,258	0.028854018	0.14759072	
		Scrub land	0	0		
		Cultivable/barren Land	40,245	0.514273666	0.49338995	
13	30 m buffer of drainage	Yes	23,721	0.303120527	0.52198462	0.885
		No	54,535	0.696879473	0.3630874	

Table 6 Susceptibility classification with TLIV_e

Ranges of TLIV	Area (Sq Km)	No of polygons	Type of zones
−5.4289 to −4.0635	6.35 (6.76 %)	8,204 (10.48 %)	Least vulnerable
−4.0635 to −3.403	7.18 (7.64 %)	9,943 (12.71 %)	Low vulnerable
−3.403 to −2.95	18.62 (19.83 %)	17,169 (21.94 %)	Moderate vulnerable
−2.95 to −2.56	35.69 (38 %)	24,502 (31.31 %)	High vulnerable
−2.56 to −1.815	26.08 (27.77 %)	18,438 (23.56 %)	Most vulnerable
Total	93.92 (100 %)	78,256 (100 %)	

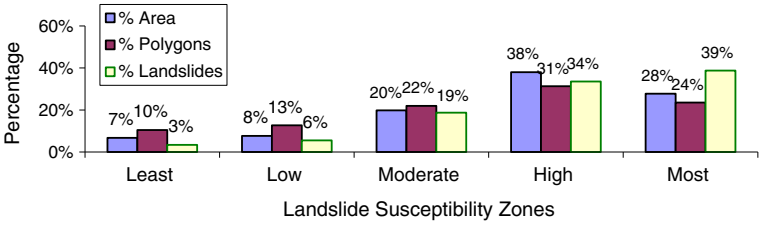

Fig. 10 *Bar chart* showing percentage of area and landslides in susceptibility zones with TLIV_e zonation

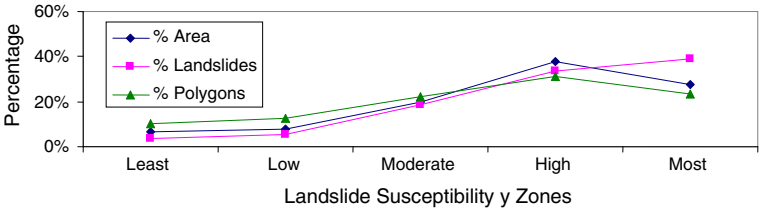

Fig. 11 *Line graph* for percentage of area, polygons and landslides in TLIV_e zonation

Table 7 Percentage and density of landslides in TLIV_e type of zonation

Type of zones	No of polygons with landslides	Landslide density	Prediction accuracy
Least vulnerable	21 (3.42 %)	3.31	91 %
Low vulnerable	34 (5.54 %)	4.74	
Moderate vulnerable	115 (18.735)	6.18	
High vulnerable	206 (33.55 %)	5.77	
Most vulnerable	238 (38.76 %)	9.13	
Total	614 (100 %)	6.54	

associated with the higher three landslide susceptibility zones out of the total number of polygons with the landslides lying in all the susceptibility zones in the study area. The Shannon's entropy-based TLIV_e values provided significantly high prediction accuracy of 91 %. Based on the TLIV_e values, a susceptibility zonation map was generated superimposed with the actual landslides events of the study area (Fig. 13).

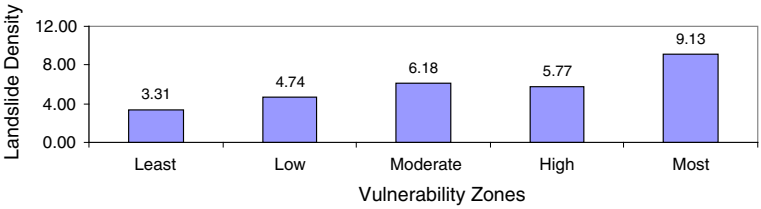

Fig. 12 Landslide densities in TLIV_e-based zonation

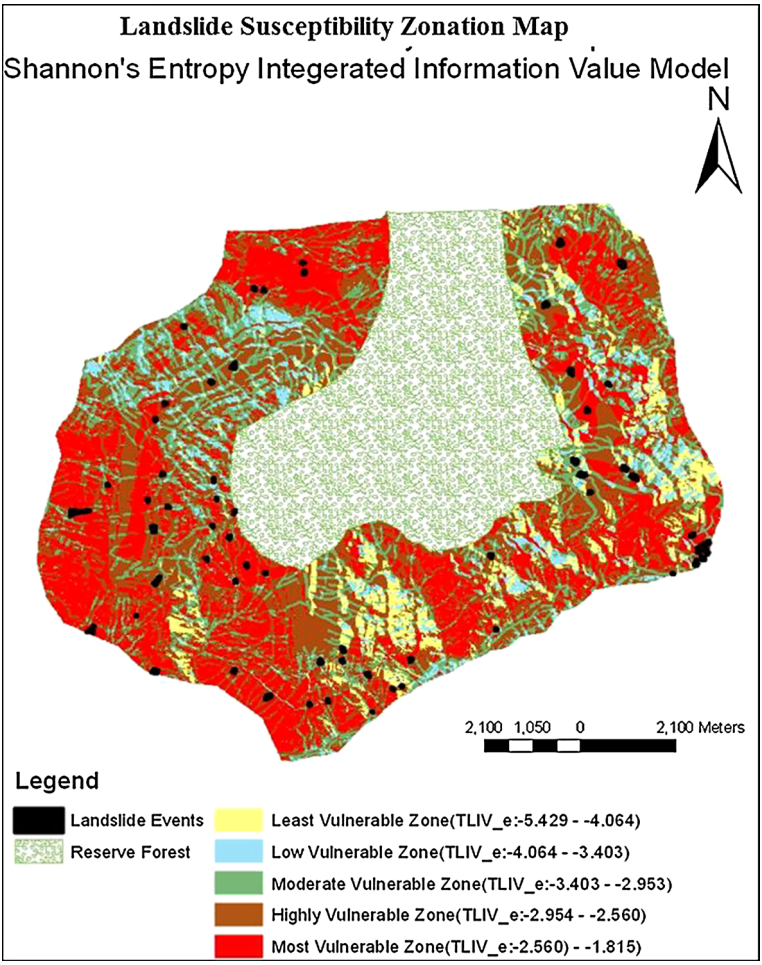

Fig. 13 Susceptibility zonation based on TLIV_e values

3 Discussion and conclusions

In the present study, comparative assessment was performed between the proposed Shannon's entropy integrated information value model and standard information value

Table 8 Comparison of area and landslides in TLIV- and TLIV_e-based zonation

Susceptibility zones	Area (Sq Km)		Polygons with landslides	
	TLIV zonation	TLIV_e zonation	TLIV zonation	TLIV_e zonation
Least	4.49 (4.78 %)	8.08 (8.60 %)	17 (2.77 %)	14 (2.28 %)
Low	16.59 (17.66 %)	10.12 (10.77 %)	77 (12.54 %)	18 (2.93 %)
Moderate	25.97 (27.65 %)	21.77 (23.18 %)	122 (19.87 %)	71 (11.56 %)
High	29.12 (31 %)	25.86 (27.53 %)	212 (34.53 %)	155 (25.24 %)
Most	17.75 (18.90 %)	28.09 (29.91 %)	186 (30.29 %)	356 (57.98 %)
Total	93.92 (100 %)	93.92 (100 %)	614 (100 %)	614 (100 %)

Table 9 Comparison of landslide densities under each susceptibility zone in TLIV- and TLIV_e-based zonation

Susceptibility zones	Landslide densities	
	TLIV zonation	TLIV_e zonation
Least vulnerable zone	3.786191537	1.732673267
Low vulnerable zone	4.641350211	1.778656126
Moderately vulnerable zone	4.697728148	3.261368856
Highly vulnerable zone	7.28021978	5.993812838
Most vulnerable zone	10.47887324	12.67354931

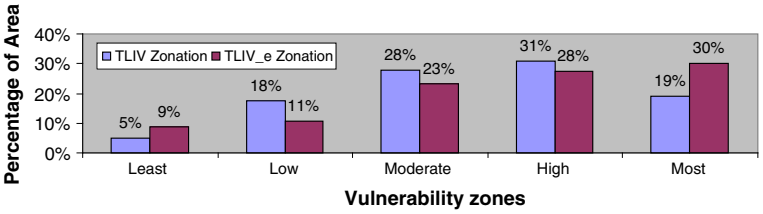

Fig. 14 *Bar chart* showing percentage of area under each susceptibility zone in TLIV- and TLIV_e-based zonation

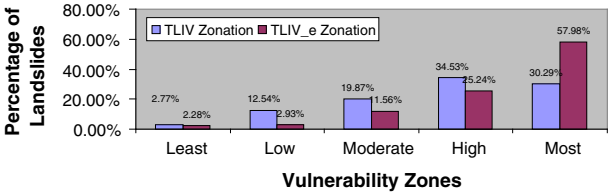

Fig. 15 *Bar chart* showing percentage of past landslides under each susceptibility zone in TLIV- and TLIV_e-based zonation

model in terms of their potential of categorizing the study area into accurate landslide susceptibility zones. This task comprised the following steps. First, the TLIV and TLIV_e was computed for each representative polygon of the study area employing the standard

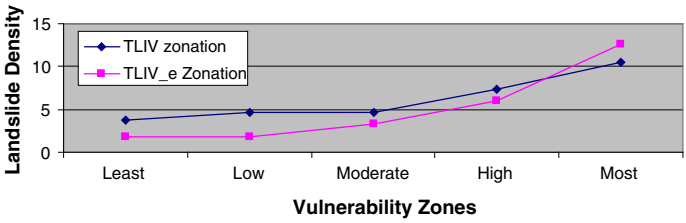

Fig. 16 *Line graph* showing landslide densities under each susceptibility zone in TLIV- and TLIV_e-based zonation

information value model and Shannon's entropy integrated information value model, respectively. Using the natural breaks (Jenks) method of ArcGIS, the TLIVs and TLIV_e values of the polygons were categorized into five classes of landslide susceptibility viz. least, low, moderate, high and most vulnerable zones in order of their increasing susceptibility to landslides (Figs. 9, 13). For each landslide susceptibility zone, the percentage of the number of polygons associated with the landslides and the landslide density were determined. Then the prediction accuracy of each model was estimated as the percentage of the polygons with landslides associated with only the higher three susceptibility zones out of the five zones delineated.

The following inferences are drawn based on the comparative assessment of the results obtained from both the models. Some of the landslide affected polygons that were underestimated and categorized under less vulnerable zones by the TLIV model were accurately delineated as being highly vulnerable through the TLIV_e model (Tables 8, 9, Figs. 14, 15 and 16). As a result of the upward shift in the vulnerability status of the polygons determined from the TLIV-based model to TLIV_e-based model, there occurred corresponding increase in the landslide density in the higher vulnerable zones of the TLIV_e model. This further resulted in significant increase in the prediction accuracy from 85 % in TLIV model to 95 % in TLIV_e model.

The investigation performed in the present research effectively demonstrated that integration of the concept of Shannon's entropy into the standard information value model can result in improved categorization of the polygons into the accurate vulnerability zones and thereby, increasing the prediction accuracy of the occurrence of landslides in the study area.

References

Atkinson PM, Massari R (1998) Generalized linear modeling of susceptibility to land sliding in the central Apennines. Italy Comput Geo-Sci 24:373–385

Avinash KG, Ashamanjari KG (2010) A GIS and frequency ratio based landslide susceptibility mapping: Aghnashini river catchment, Uttarkhand, India. Int J Geomat Geosci 1(3):343–354

Carro M, Amicis MD, Luzi L, Marzorati S (2003) The application of predictive modeling techniques to landslides induced by earthquakes: the case study of the 26 September 1997 Umbria–Marche earthquake (Italy). Eng Geol 69(1–2):139–159

Dhakal AS, Sidle RC (2002) Physically based landslide hazard model 'U' method and issues. EGS XXVII General Assembly, Nice, pp 21–26

Ercanoglu M, Gokceoglu C (2004) Use of fuzzy relations to produce landslide susceptibility slope failure risk maps of the Altindag (settlement) region in Turkey. Eng Geol 55:277–296

George YL, Long SC, David WW (2007) Vulnerability assessment of rainfall-induced debris flows in Department of Earth Systems and GeoInformation Sciences, College of Science, George Mason University, Fairfax, VA 22030,ETATS-UNIS. http://www.springerlink.com/content/uj26871v2831nx44. Accessed 10 Jan 2011

Gupta M, Ghose MK, Sharma LP (2009) Application of remote sensing and GIS for landslides hazard and assessment of their probabilistic occurrence-A case study of NH31A between Rangpo and Singtam. J Geomat 3(1):13–17

Jadda M, Shafri Helmi ZM, Manssor Shattri B, Mohammad S, Saeid P (2009) Landslide susceptibility evaluation and factor effect analysis using probabilistic- frequency ratio model. Eur J Sci Res 33(4):654–668

Lee S (2004) Application of likelihood ratio and logistic regression models to landslide susceptibility mapping using GIS. Environ Manage 34(2):223–232

Lee S (2005) Application of logistic regression model and its validation for landslide susceptibility mapping using GIS and remote sensing data. Int J Remote Sensing 26:1477–1491

Lee S (2007) Application and verification of fuzzy algebraic operators to landslide susceptibility mapping. Environ Geol 52:615–623

Lee S, Pradhan Biswajeet (2006) Landslide hazard mapping at Selangor, Malaysia using frequency ratio and logistic regression models. Landslides 4(1):33–41

Lee S, Sambath T (2006) Landslide susceptibility mapping in the Damrei Romel area, Cambodia using frequency ratio and logistic regression. Environ Geol 50:847–855

Lee S, Ryu J-H, Min K, Won J-S (2003a) Landslide susceptibility analysis using GIS and artificial neural network. Earth Surf Process Landforms 28:361–1376

Lee S, Ryu JH, Lee MJ, Won JS (2003b) Landslide susceptibility analysis using artificial neural network at Bonn, Korea. Environ Geol 44:820–833

Lee S, Ryu JH, Won JS, Park HJ (2004) Determination and application of the weights for landslide susceptibility mapping using an artificial neural network. Eng Geol 71:289–302

Pachauri AK, Pant M (1992) Landslide hazard mapping based on geological attributes. Eng Geol 32:81–100

Pachauri AK, Gupta PV, Chander R (1998) Landslide zoning in a part of the Garhwal Himalayas. Environ Geol 36(3–4):325–334

Patanakanog B (2001) Landslide hazard potential area in 3 dimension by remote sensing and gis technique. Land Development Department, Thailand. www.ecy.wa.gov/programs/sea/landslides/help/drainage.html. Accessed 13 Jan 2010

Pistocchi A, Luzi L, Napolitano P (2002) The use of predictive modeling techniques for optimal exploitation of spatial databases: a case study in landslide hazard mapping with expert system-like methods. Environ Geol 58:251–270

Pradhan B, Ahmed MY (2010) Manifestation of remote sensing data and GIS on landslide hazard analysis using spatial-based statistical models. Arab J Geosci 3:319–326

Pradhan B, Lee S (2010) Delineation of landslide hazard areas on Penang Island, Malaysia, by using frequency ratio, logistic regression, and artificial neural network models. Environ Earth Sci 60(5):1037–1054

Ramakrishna D, Ghose MK, Vinu Chandra R, Jeyaram A (2005) Probabilistic techniques, GIS and remote sensing in landslide hazard mitigation: a case study from Sikkim Himalayas, India. Geocarto Int 20(4):53–58

Sakellariou MG, Ferentinou MD (2001) GIS-based estimation of slope stability. Nat Hazards Rev 2(1):12–21

Sarkar S, Kanungo DP, Patra AK, Kumar Pushpendra (2008) GIS based spatial data analysis for landslide susceptibility mapping. J Mt Sci 5:52–62

Sharma LP, Patel N, Ghose MK, Debnath P (2009a) Geographical information system based landslide probabilistic model with tri-variate approach—a case study in Sikkim Himalayas. In: 18th United Nation's regional cartographic conference-Asia and the Pacific, Bangkok 26–29 October 2009

Sharma LP, Patel N, Ghose MK, Debnath P (2009b) Landslide susceptibility. Coordinates V(11):31–34

Sharma LP, Patel N, Ghose MK, Debnath P (2011) Landslide vulnerability assessment and zonation through ranking of causative parameters based on landslide density-derived statistical indicators. Geocarto Int 26(6):491–504

Sharma LP, Patel N, Ghose MK, Debnath P (2012a) Influence of Shannon's entropy on landslide-causing parameters for vulnerability study and zonation—a case study in Sikkim, India. Arab J Geosci 5(3):421–431

Sharma LP, Patel N, Ghose MK, Debnath P (2012b) Assessing landslide vulnerability from soil characteristics—a GIS based analysis. Arab J Geosci 5(4):789–796

Sharma LP, Patel N, Ghose MK, Debnath P (2012c) Geo-spatial technology based landslide vulnerability assessment and zonation in Sikkim Himalayas in India. J Geomat 6(2):51–57

Sharma LP, Patel N, Ghose MK, Debnath P (2012d) Application of frequency ratio and likelihood ratio model for geo-spatial modeling of landslide hazard vulnerability assessment and zonation: a case study from the Sikkim Himalayas in India. Jeocarto Int 10(1080/10106049):748830

Sharma LP, Patel N, Ghose MK, Debnath P (2013) Synergistic application of fuzzy logic and geo-informatics for landslide vulnerability zonation—a case study in Sikkim Himalayas, India. Appl Geomat 5(4):271–284. doi:10.1007/s12518-013-0115-7

Shu-Quin W, Unwin DJ (1992) Modeling landslide distribution on loess soils in China: an investigation. Int J Geogra Inf syst 6(5):391–405

Sinha BN, Varma RS, Paul DK (1975) Landslides in Darjeeling District (West Bengal) and adjacent areas. Bull Geol Surv India B(36):45

Sivakumar GL, Mukesh MD (2002) Landslide analysis in geographic information systems. Department of Civil Engineering, Indian Institute of science, Bangalore, India. www.gisdevelopment.net/application/natural_hazards/landslides/nhls0011pf.htm. Accessed 26 Feb 2011

Uromeihy A, Mahdavifar MR (2000) Landslide hazard zonation of Khorshrostam area, Iran. Bull Eng Geol Environ 58:207–213

Van Westen CJ, Rengers N, Terlien MTJ, Soeters R (1997) Prediction of the occurrence of slope instability phenomena through GIS-based hazard zonation. Geol Rundsch 86:404–414

Yilmaz I (2009a) Landslide susceptibility mapping using frequency ratio, logistic regression, artificial neural networks and their comparison: a case study from Kat landslides (Tokat-Turkey). Comput Geosci 35(6):1125–1138

Yilmaz I (2009b) A case study from Koyulhisar (Sivas-Turkey) for landslide susceptibility mapping by artificial neural networks. Bull Eng Geol Environ 68(3):297–306

Yin KL, Yan TZ (1988) Statistical prediction models for slope instability of metamorphosed rocks. In: International symposium on landslides, Lausanne, pp 1269–1272