App. Envi. Res. 43(1) (2021): 102-115

Applied Environmental Research

Journal homepage : http://www.tci-thaijo.org/index.php/aer

Landslide Identification and Zonation Using the Index of Entropy Technique at Ossey Watershed Area in Bhutan

Thongley, Chaiwiwat Vansarochana*

Faculty of Agriculture, Natural Resources and Environment, Naresuan University, Phitsanulok, Thailand

* Corresponding author: chaiwiwatv@nu.ac.th

Article History

Submitted: 25 July 2020/ Revision received: 1 October 2020/ Accepted: 1 October 2020/ Published online: 4 January 2021

Abstract

The landslide is one of the natural disasters which claim human lives and incur huge economic losses, especially in the mountainous area. The main aim of this study is to develop different zones of landslide-prone area using the index of entropy (IOE) at the Ossey watershed area in Bhutan. During the landslide inventory, 164 landslides were identified of which 115 locations were used for the training dataset while the remaining 49 locations were used for the validation dataset. A total of ten causal factors were used for this study including elevation, slope, aspect, slope curvature, stream power index, normalized difference vegetation index (NDVI), distance from the road, distance from the river, lithology, and rainfall. The IOE was used to obtain the relationship between the landslide events and the causal factors. The most influential causal factors were NDVI, slope, and rainfall with the weightage of 0.377, 0.347, and 0.175 respectively as per the IOE. The final landslide susceptibility map was classified into five classes using the geometrical interval classification. The validation was done using the receiver operating characteristic (ROC) curves and the kappa index. The area under the curve (AUC) for the success rate and prediction rate was 0.7821 and 0.8377, respectively. The kappa index using the training dataset and validation dataset were 0.4111 and 0.4898, respectively. The final landslide susceptibility map is accurate enough for the future references by the decision-makers and the engineers.

Keywords: Landslide susceptibility mapping; Index of entropy; Area under the curve; Kappa index

https://doi.org/10.35762/AER.2021.43.1.8

Introduction

Landslide is considered as one of the prominent destructive natural disaster in the mountainous countries [1]. The landslide caused damage to the properties, natural resources, and threaten human lives [2]. The total death caused by the landslides between 1910–2020 was 67052, the total injured was 12060 and the total affected was 14,679,178 globally [3]. The detail of the total death, injured and affected by the landslide in various continents between 1910–2020 is shown in Table 1 which is maintained by international disaster database EM-DAT [3] and can be downloaded from the website https://public.emdat.be. The trend is expected to increase in the future due to urbanization, deforestation, and climate change [4]. Although it is important to understand the landslide mechanism and landslide mapping, the challenges continue to exist due to the complex nature of the landslides [5]. The landslide susceptibility mapping (LSM) is the depiction of the spatial distribution of the probable landslides occurrence in the given area [6]. The LSM act as a standard tool for the decision-maker and engineers for developmental activities [5].

Table 1 The total number of deaths, injured and affected by the landslide from 1910-2020

	Total death	No injured	Total affected
Asia	25,604	5,564	8,821,136
Africa	3,171	442	221,907
America	20,869	5,484	5,574,270
Europe	16,830	518	40,450
Oceania	578	52	21,415
Global data	67,052	12,060	14,679,178

In the last two decades, many scholars have conducted extensive studies on landslides using different approaches worldwide [4]. The reason for international interest on landslide is due to the significant impact of the landslide on the socio-economic and increased pressure on the developmental activities and urbanization [7]. Although there is great progress in the landslide assessment approaches, the limitation still exists [8]. The development of the dynamic models associated with GIS can forecast the propagation of slope failure and to delineate the zones based on the severity [9]. These landslide models linked to GIS are broadly classified into heuristic, deterministic, and statistical approaches [2]. Heuristic approaches are based on the expert's opinion while the deterministic and statistical approaches are based on the numerical expression between the landslide event and landslide causative factors [10]. Deterministic approaches are feasible for the only smaller area due to its requirement of exhaustive geotechnical field investigation [11]. In recent years, the statistical approaches in conjunction with GIS have become more popular [9]. Some of the popular statistical methods are frequency ratio, information value, statistical index, evidence of belief function, the weight of evidence, index of entropy, and the logistic regression [2]. The statistical models require past landslide data and the influencing factors which work on the assumption that future landslides follow the condition of the past [9]. The powerful machine learning approaches such as artificial neural networks, support vector machine, random forest, alternating decision tree, kernel logistic regression, naïve Bayes, naïve Bayes tree, and multivariate adaptive regression were also introduced for the landslide assessment [2].

The focus on the landslide problems was heightened with increasing developmental activities in the geologically fragile Himalayan region [12]. Bhutan is a part of the Himalayan region and it is considered as one of the most landslide susceptible area [13]. During the monsoon season (May-September), the landslide block numerous roads in Bhutan [12]. The landslides in Bhutan are caused by steep terrain, natural geological setting, anthropogenic activities, and unpredictable precipitation during the mon-

soon season [12]. The southern part of Bhutan is cut off from the rest of the places during the monsoon season due to the number of landslides [13]. Although Landslide is one of the major problems in Bhutan, there are very limited studies carried out covering the landslides. The dearth of publications related to the scale of landslides, spatial distribution, and the frequency of occurrence becomes essential in Bhutan [14]. The area chosen for this study is the Ossey watershed area. Ossey watershed area is one of the most landslide-prone areas in Bhutan. The national highway between two districts (Zhemgang and Gelephu) runs through this area. However, during the monsoon season, the road gets closed for months due to the road blockage by the landslides. Recently, the government built an alternative route to avoid landslides. However, the landslide problem continues to persist in many places of Ossey.

This study aimed at developing landslide susceptibility mapping using the geospatial index of entropy model. This study attempts to find the weightage of influence from the individual landslide causal factors as well as to derive the relationship between the individual classes of each causal factor.

Materials and methods

1) Description of the study area

The Ossey watershed area (Figure 1) is located in the northern part of Sarpang district of Bhutan. It is situated in between 26°50'00"N – 27°15'00"N latitude and 90°10'00"E – 90°50'00"E longitude and it covers an area of 820.72 km^2. The geomorphology of the area is characterized by steep and rugged terrain varying from 190 m – 4194 m above mean sea level. The Ossey watershed area is located on the south-facing slope of the Himalaya. Most of the people in this region depend on agriculture and livestock. However, most of the lands are unfeasible for agricultural purposes due to the rugged terrain. The study area is composed of sandstone, schist, phyllite, quartzite, limestone, marble, and granite. The intense precipitation during the monsoon season triggers multiple landslides every year. The national highway (Gelephu-Zhemgang highway) passes through this basin. However, the highway gets block continuously for months during the monsoon season causing numerous inconveniences to the local people due to the landslides.

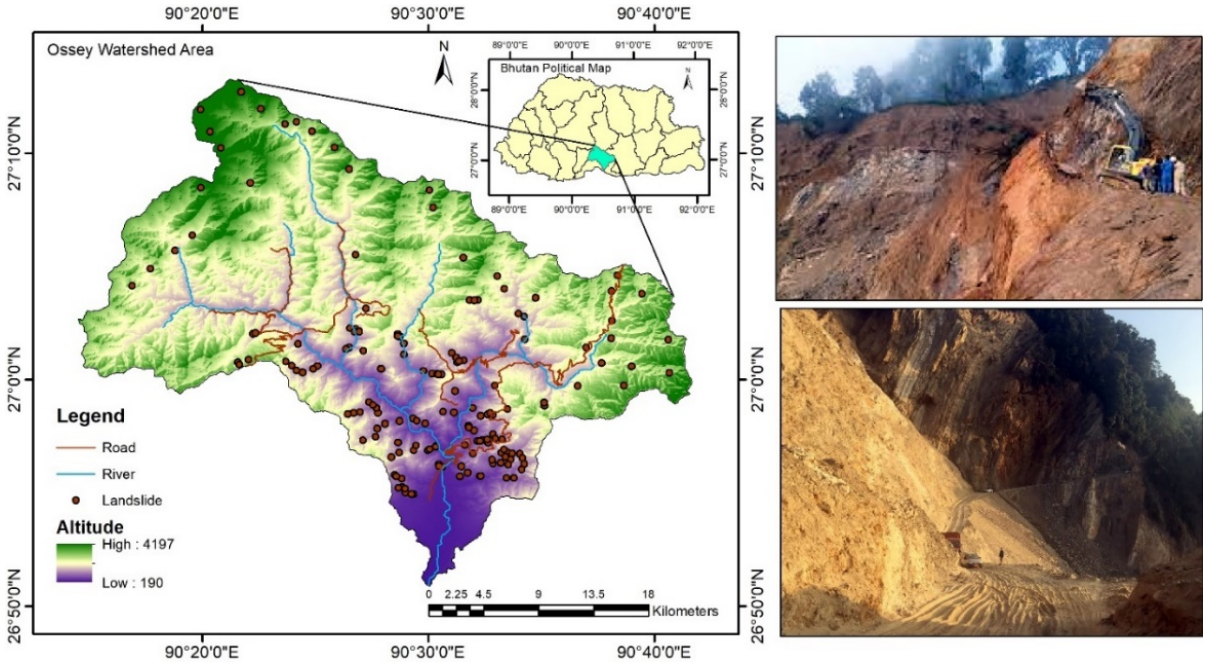

Figure 1 The study area with the landslide location.

2) Preparation of landslide inventory

The landslide inventory is the collection of basic landslide information of the existing landslide through the catalogue of existing landslide and interpretation of aerial photographs [15]. The landslide inventory is fundamental for the derivation of the relation between the landslides and the causal factors [16]. The accuracy of the landslide location during the landslide inventory is crucial for the landslide susceptibility analysis [8]. For this study, the interpretation of satellite image, google earth, and field investigation were employed for the landslide inventory mapping. During the landslide inventory, a total of 164 landslide locations were identified as shown the Figure 1. The landslide inventory was divided into a training dataset (70%, 115 locations) and validation dataset (30%, 49 locations).

3) Landslide causal factors

It is important to select only the responsible causal factors for the development of the landslide predictive models [17]. The usage of many causal factors for the LSM is not necessary because the inclusion of the noise factor may reduce the accuracy of the results [8]. This study uses ten effective causal factors. They are elevation, slope, aspect, slope curvature, stream power index (SPI), normalized difference vegetation index (NDVI), distance from the road, distance from the river, lithology, and rainfall.

Based on the SRTM DEM (30 m spatial resolution), five causal factors were constructed namely elevation, slope, aspect, slope curvature, and SPI. The elevation Figure 2(a) is used as one of the causal factors because the tendency of landslide occurrence is different in specific elevation range [8]. The stability of the slope depends on the interaction of slope angle with the slope forming materials [18] and it is shown in Figure 2(b). The different slope aspects influence the temperature, precipitation, evaporation, etc. [2]. The aspect map is shown in Figure 2(c). The slope curvature (Figure 2(d)) affects the landslide due to its differences in the rate of change of the slope gradient or aspect in a specific direction [2]. The SPI (Figure 2(e)) measures the erosive strength of the flowing water [19] and it is calculated based on an equation given by Moore et al. [20].

$$SPI = A_s tan\,\beta \quad \text{(Eq. 1)}$$

Where *As* is the specific catchment area and β is the slope gradient in degrees.

The NDVI shows the health of the vegetation and the root of vegetation act as reinforcement to the landslide. The sentinel-2 was used to derive the NDVI map (Figure 2(f)) and it is calculated using the Eq. 2.

$$NDVI = \frac{NIR-R}{NIR+R} \quad \text{(Eq. 2)}$$

Where NIR is Near Infrared and R is the red band of sentinel-2

The river affects the stability of the slope forming materials and erodes a lower portion of the slope [18]. Similarly, road construction also destabilises an area as the anthropogenic activity changes the equilibrium of the slope [8]. The digital topographic map of Bhutan (1:25,000) is used to extract the road and river locations and five classes each for distance from river and road were constructed (Figure 2 (g) and Figure 2(h)).

The different lithologic formation, its composition, structure and permeability influence the stability of the slope [4]. The Geological Map of Bhutan (1:500,000) which was prepared by Long et al. [21] is used for the extraction of lithological data (Figure 2 (i)) and its lithological description of the study area is elaborated in Table 2.

The rainfall is considered as one of the triggering factors for the landslides. The places which receive heavy rainfall experiences relatively

more landslides. The rainfall data of 21 years (1996–2017) from twenty rain gauge stations across Bhutan were used for this study. The inverse distance weighting (IDW) interpolation is used to derive the rainfall map (Figure 2 (j)).

All the factors are converted into 30 m x 30 m grid and the study area constitutes 911,947 cells with 1376 rows and 1497 columns grid cells.

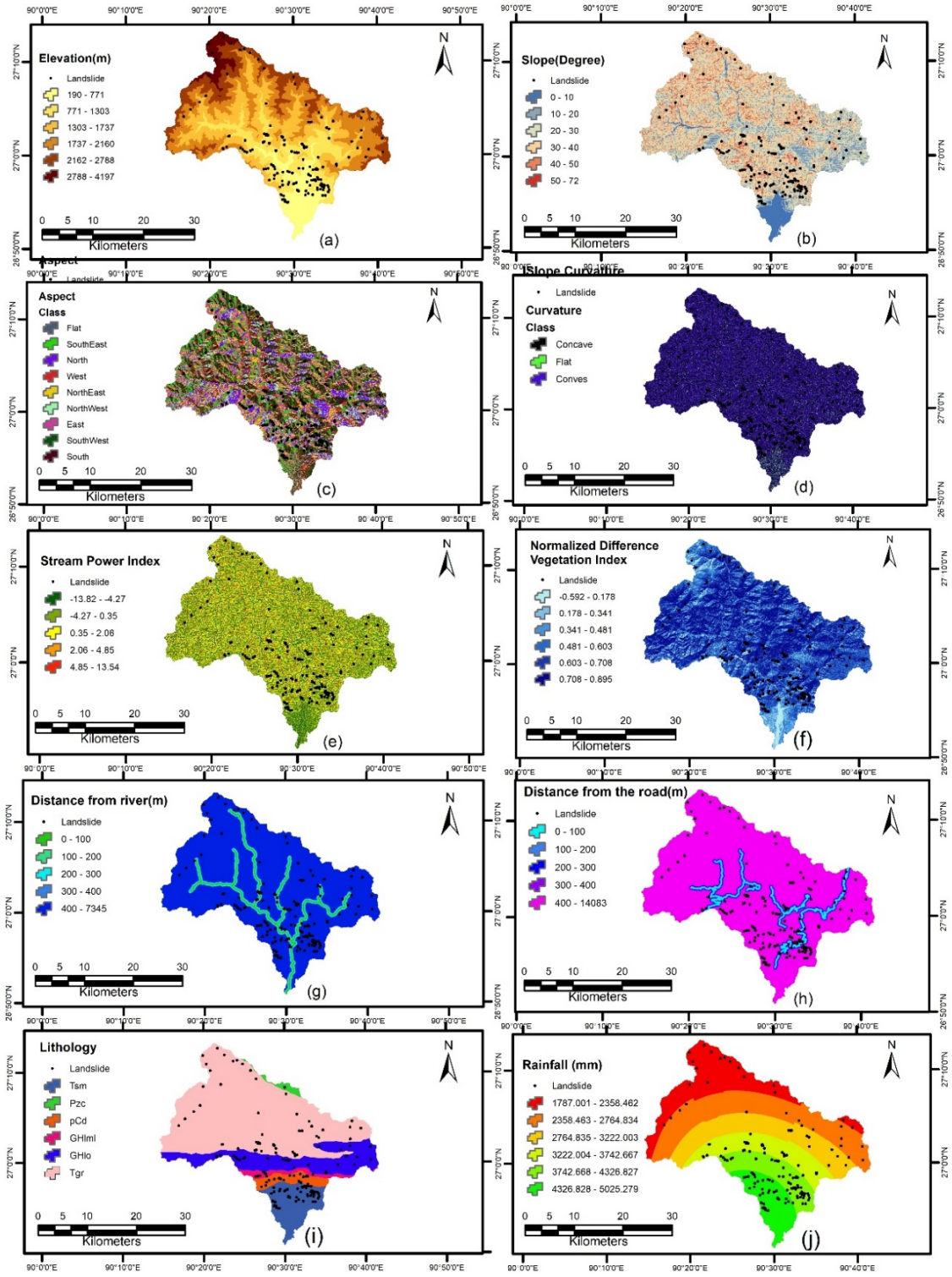

Figure 2 Causal factors (a) elevation, (b) slope (c) aspect, (d) slope curvature, (e) stream power index, (f) normalized difference vegetation index, (g) distance from river, (h) distance from road, (i) lithology, and (j) rainfall.

Table 2 Lithological description at Ossey watershed area

Geologic age	Code	Lithological description
Miocene-Pliocene	Tsm	Sandstone and cobble-conglomeratic sandstone
Paleoproterozoic	pCd	Schist, phyllite, quartzite, and limestone.
Ordovician or younger	Pzc	Micaceous quartzite, schist, marble, phyllite, and phyllitic quartzite.
Neoproterozoic-Cambrian	GHlml	Quartzite, schist, kyanite, sillimanite, and staurolite
Cambrian-Ordovician	GHlo	Granite, schist, quartzite and granite
Miocene	Tgr	Massive to foliated, syn-Himalayan leucogranite plutons.

4) Index of entropy

The entropy shows the degree of imbalance, disorder, unstable behaviour, and uncertainty in the system [8]. With regards to the landslides, entropy refers to the degree of influence or instability caused by the causal factors on the landslides occurrence [22]. The entropy value is used to calculate the weight of the individual causal factors [23]. The Eq. 3 to Eq. 8 is used for the calculation of the weight of the landslides causal factors as shown in Table 3.

For the LSM, the causal factors are secondarily reclassified using the value of probability density (P_{ij}). The algebraic sum (Eq. 9) of the weighted secondarily reclassified causal factors produce LSM [24].

$$P_{ij} = \frac{b}{a} \quad \text{(Eq. 3)}$$

$$(P_{ij}) = \frac{P_{ij}}{\sum_{i=1}^{sj} P_{ij}} \quad \text{(Eq. 4)}$$

$$H_j = -\sum_{i=1}^{sj}(Pij)\log_2(P_{ij}),\ j=1,\ldots,\ n \quad \text{(Eq. 5)}$$

$$H_{jmax} = log_2 Sj \quad \text{(Eq. 6)}$$

$$I_j = \frac{H_{jmax} - H_j}{H_{jmax}} \qquad I=(0,\ 1),\quad j=1,\ldots,\ n \quad \text{(Eq. 7)}$$

$$W_j = I_j P_{ij} \quad \text{(Eq. 8)}$$

Where b is the percentage of landslide pixels in a class; a is the percentage of pixels in a class; P_{ij} is the frequency ratio, (P_{ij}) is the probability density, H_j and H_{jmax} is the entropy values, s_j is the number of classes, I_j is the information coefficient value, and W_j is the weight of the causal factor.

$$LSM = \sum C\ x\ W_j \quad \text{(Eq. 9)}$$

Where LSM is the landslide susceptibility mapping, C is the value of the class after secondary reclassification based on the probability density (P_{ij}), and W_j is the resultant weight of a causal factor [24].

Table 3 Spatial relationship between the landslide causal factors and the landslide events using the index of entropy

Causal factor	Class	No of Pixel in domain	% of Domain	No of landslide	% of landslide	P_{ij}[1]	$(P_{ij})^2$[2]	H_j[3]	H_{jmax}[4]	I_j[5]	W_j[6]
Elevation (m)	190 - 771	96582	10.59	32	27.83	2.63	0.376	2.239	2.585	0.134	0.156
	771 - 1303	152224	16.69	38	33.04	1.98	0.283				
	1303 - 1737	236318	25.91	14	12.17	0.47	0.067				
	1737 - 2160	247540	27.14	19	16.52	0.61	0.087				
	2160 - 2788	142060	15.58	8	6.96	0.45	0.064				
	2788 - 4197	37223	4.08	4	3.48	0.85	0.122				
Slope (Degree)	0-10	64424	7.06	5	4.35	0.62	0.061	2.052	2.585	0.206	0.347
	10-20	149925	16.44	15	13.04	0.79	0.078				
	20-30	285396	31.3	26	22.61	0.72	0.071				
	30-40	284202	31.16	31	26.96	0.86	0.086				
	40-50	109832	12.04	26	22.61	1.88	0.186				
	50<	18168	1.99	12	10.43	5.24	0.518				
Aspect	Flat	285	0.03	0	0.00	0.00	0.000	2.839	3.17	0.104	0.090
	North	118174	12.96	6	5.22	0.4	0.052				
	North-East	97997	10.75	9	7.83	0.73	0.094				
	East	110602	12.13	14	12.17	1.00	0.129				
	South-East	129200	14.17	33	28.7	2.03	0.26				
	South	120454	13.21	18	15.65	1.19	0.152				
	South-West	126820	13.91	18	15.65	1.13	0.145				
	West	111498	12.23	7	6.09	0.50	0.064				
	North-West	96917	10.63	10	8.70	0.82	0.105				
Curvature	Concave	434525	47.65	70	60.87	1.28	0.476	1.521	1.585	0.04	0.036
	Flat	36424	3.99	3	2.61	0.65	0.243				
	Convex	440998	48.36	42	36.52	0.76	0.281				
SPI	-13.82 - -4.27	233383	25.59	18	15.65	0.61	0.092	2.135	2.322	0.081	0.107
	-4.27 - 0.35	240632	26.39	28	24.35	0.92	0.139				
	0.35 - 2.06	281843	30.91	37	32.17	1.04	0.157				
	2.06 - 4.85	129181	14.17	23	20.00	1.41	0.213				
	4.85 - 13.54	26908	2.95	9	7.83	2.65	0.399				
NDVI	-0.592 - 0.178	25989	2.85	17	14.78	5.19	0.461	2.066	2.585	0.201	0.377
	0.1778 - 0.341	69043	7.57	26	22.61	2.99	0.265				
	0.341 - 0.481	103152	11.31	14	12.17	1.08	0.096				
	0.481 - 0.603	157319	17.25	15	13.04	0.76	0.067				
	0.603 - 0.708	257631	28.25	26	22.61	0.80	0.071				
	0.708 - 0.895	298813	32.77	17	14.78	0.45	0.04				
Distance from river (m)	0-100	31639	3.47	7	6.09	1.75	0.328	2.159	2.322	0.07	0.075
	100-200	27341	3	1	0.87	0.29	0.054				
	200-300	29785	3.27	4	3.48	1.06	0.199				
	300-400	25379	2.78	4	3.48	1.25	0.234				
	400<	797803	87.48	99	86.09	0.98	0.184				
Distance from road (m)	0-100	30136	3.305	5	4.35	1.32	0.228	2.165	2.322	0.068	0.078
	100-200	22593	2.477	3	2.61	1.05	0.182				
	200-300	23374	2.563	6	5.22	2.04	0.352				
	300-400	19606	2.15	1	0.87	0.40	0.07				
	400<	816238	89.505	100	86.96	0.97	0.168				
Lithology	Tsm	96424	10.57	32	27.83	2.63	0.273	2.379	2.585	0.08	0.128
	Pzc	9617	1.05	1	0.87	0.82	0.086				
	pCd	34574	3.79	12	10.43	2.75	0.286				
	GHlml	22953	2.52	5	4.35	1.73	0.179				
	GHlo	162532	17.82	23	20.00	1.12	0.117				
	Tgr	585847	64.24	42	36.52	0.57	0.059				
Rainfall (mm)	1787.00 - 2358.46	186862	20.49	15	13.04	0.64	0.084	2.225	2.585	0.139	0.175
	2358.46 - 2764.83	237368	26.03	13	11.30	0.43	0.058				
	2764.83 - 3222.00	173635	19.04	9	7.83	0.41	0.054				
	3222.00 - 3742.67	120060	13.17	18	15.65	1.19	0.157				
	3742.67 - 4326.83	99686	10.93	36	31.30	2.86	0.379				
	4326.83 - 5025.28	94336	10.34	24	20.87	2.02	0.267				

Note: [1]frequency ratio, [2]probability density, [3] and [4]entropy values, [5]information coefficient value, [6]weight of the causal factor.

5) Accuracy assessment

This study uses area under the curve (AUC) of the receiver operating characteristics (ROC) curve and the kappa index for the validation of the model and LSM. These validation parameters were derived based on the true positive (TP), false positive (FP), true negative (TN), and false negative (FN). These parameters are used to calculate sensitivity and specificity. The sensitivity is the proportion of landslide pixels that are correctly classified as landslides while specificity is the proportion of the non-landslide pixel that are correctly classified as non-landslides [5]. The sensitivity and specificity are calculated using Eq. 10 and Eq. 11.

The performance of the landslide susceptibility model is assessed based on the value of AUC of the ROC curve. The ROC curve is plotted using 1-specificity on the x-axis and sensitivity on the y-axis [23]. The training dataset is used for the success rate while the prediction rate is evaluated using the validation dataset. The ROC curve is developed for both training datasets and validation datasets. The success rate will evaluate the goodness of fit of the model while the prediction rate is used for checking the predictive ability of the LSM [6]. The AUC is calculated using the Eq. 12 and its value ranges from 0.5–1. The interpretation of accuracy are: excellent (0.9–1), very good (0.8–0.9), good (0.7–0.8), moderate (0.6–0.7), and poor (0.5–0.6) [8].

The kappa index is calculated using the Eq. 13 to Eq. 15 and it is used for checking the reliability of the model [25]. Kappa index describes the ability of the models to classify the landslide pixels [5]. It is calculated using the agreement between the observed value and the expected value [23]. The interpretation of the kappa index are: 0.8–1.0 almost perfect, 0.6–0.8 substantial, 0.4–0.6 moderate, 0.2–0.4 fair, 0–0.2 slight, and ≤0 poor [26].

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP+FN}} \quad \text{(Eq. 10)}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN+FP}} \quad \text{(Eq. 11)}$$

$$\text{AUC} = \frac{\sum \text{TP} + \sum \text{TN}}{\text{P+N}} \quad \text{(Eq. 12)}$$

Where TP is true positive, TN is true negative, FP is false positive, FN is false negative, P is the total number of landslides, and N is the total number of non-landslides.

$$\text{Kappa} = \frac{\text{P}_{\text{obs}} - \text{P}_{\text{exp}}}{1 - \text{P}_{\text{exp}}} \quad \text{(Eq. 13)}$$

$$P_{obs} = \frac{\text{TP+TN}}{\text{n}} \quad \text{(Eq. 14)}$$

$$P_{exp} = \frac{(\text{TP+FN})(\text{TP+FP}) + (\text{FP+TN})(\text{FN+TN})}{\text{n}^2} \quad \text{(Eq. 15)}$$

Where P_{obs} is the observed agreement, P_{exp} is the expected agreement, TP is true positive, TN is true negative, FP is a false positive, and FN is false negative, n is the total number of landslide and non-landslide pixels used for the validation purpose.

Results and discussion

1) The spatial relationship between the landslide occurrence and causal factors.

The W_j of Table 3 indicates the magnitude of influence of the causal factors on the landslide occurrences [27]. Among the ten causal factors, the most dominant causal factor was NDVI (0.377), followed by the slope gradient (0.347), rainfall (0.175), and lithology (0.128). On the other hand, the least influential factor on the landslide was slope curvature with a weight of 0.036. The weight of the elevation, aspect, SPI, distance from the river, and distance from the road were 0.156, 0.090, 0.107, 0.075, and 0.078, respectively.

Every class of the causal factors contributes differently to the landslides event. The probability density (P_{ij}) is calculated using the Eq. 2 and it is used to compare how the individual classes of each landslide causal factors contribute to the landslides events. From the result of (P_{ij}) of Table 3, it is noticed that the lowest elevation class (190 m – 171 m) is highly susceptible to the landslides with a probability density (P_{ij}) 0.376. As the elevation increases, the (P_{ij}) decreases. This indicates that the landslide probability decreases with an increase in altitude. This may be due to the coincidence of the lowest elevation with the heavy rainfall and steep slope. A similar result was also obtained by Mondal and Mandal [27].

Regarding the slope gradient, it is noticed that the (P_{ij}) value increases with increases in the slope gradient. The field verification also confirmed that the steeper slopes exhibit more landslides. The result is agreed with the results of Singh et al. [28].

In the case of aspect, the slope facing southeast experiences more landslide with (P_{ij}) value 0.260, followed by south (0.152), southwest (0.145), and east (0.129) direction. The rest of the slope directions are less susceptible to the landslides. Kaur et al. [29] studied landslide at Sikkim which is one of the states of India and neighbour to Bhutan and it is also a part of the Himalayan region. Kaur et al. [29] observed that the south, southeast, and southwest with a higher value of (P_{ij}). This may be due blockage of wind that comes from the south (Indian Ocean) by the giant Himalayan mountain. The condense wind finally drops as precipitation which triggers relatively more landslides on the south-facing slope.

The concave slope curvature ((P_{ij})=0.476) is more susceptible to landslide, followed by convex slope curvature ((P_{ij})=0.281) and flat area ((P_{ij})=0.243). This result is evident that concave terrain holds more water which decreases the soil strength and augments soil erosion. Youssef et al. [18] also found that the concave has a higher value of (P_{ij}).

The landslide is also influenced significantly by the stream power index (W_j=0.107). It was noticed that the (P_{ij}) increases with the increases in SPI value. This indicates that the erosive power of the stream increases the landslide event. A similar result was observed by Wang et al. [30].

The NDVI value -1 to 0 represents water, 0 to 0.2 represent barren land, rock, and built-up area, and 0.2 to 1 represents the vegetated area [31]. The shallow landslides take place within the forest rooting zone while the deep-seated landslides take place deep below the forest root zones. The NDVI plays an important role in shallow landslides due to its activities within the foresting root zone. For this study, NDVI ranges from -0.592 to 0.895 and it is the top influential factor (W_j=0.377). The (P_{ij}) decreases with an increase in NDVI value. This indicates that the landslide occurrence decreases with the increase in vegetation strength and vice versa. The vegetation roots reinforce the slope surface and minimize the landslides. The result is agreed with the theoretical concept of reinforcement of soil by the vegetation roots and a similar result was obtained by Mondal and Mandal [27].

Regarding the distance from the river, it was noticed that the landslides are more prominent within a distance of 100 m with a (P_{ij}) value of 0.328. Beyond 100 m, the (P_{ij}) value is lesser than 0.328 indicating minimal landslides events. The field verification also noticed that the landslide is relatively more near the river.

In the case of a road, the distance from the road doesn't follow the systematic trend of the (P_{ij}) value. However, it was noticed that the landslides are more common within 300 m from the road.

Pertaining to the lithology, the landslide is more common for the class pCd ((P_{ij})=0.286), followed by Tsm ((P_{ij})=0.273), GHlml ((P_{ij})= 0.179), and GHlo ((P_{ij})=0.117). The least landslide-prone lithology are Pzc ((P_{ij})=0.086) and Tgr ((P_{ij})=0.059). Regarding the rainfall, the (P_{ij}) also increases with the increase in rainfall amount. This shows that the more landslides events take place in the area having a high amount of rainfall. A similar result was obtained by Tien Bui et al. [23].

2) Landslide susceptibility mapping using the index of entropy

For the LSM, all the factors were secondarily reclassified using the value of landslide probability density (P_{ij}). The LSM is developed by adding all the weighted secondarily reclassified causal factors using the Eq. 14.

The classification of LSM necessary for better visualization and interpretation (Figure 3). The geometrical interval classification was introduced for this study due to its better accuracy compared to other classification methods. The class break in geometrical interval is based on the class intervals that have a geometrical series. Table 4 shows the reclassified LSM with its size and percentage of the area of an individual landslide susceptibility zone. As per the geometrical interval classification, 17.37% (142.53 km^2), 28.45% (233.50 km^2), and 26.11% (214.26 km^2) of the total area fall under very low, low, and moderate susceptibility zones, respectively.

$$\begin{aligned} LSM = {} & (\text{Elevation x } 0.156) + (\text{Slope x } 0.347) + (\text{Aspect x } 0.090 \\ & +(\text{Curvature x } 0.036) + (\text{SPI x } 0.107) + (\text{NDVI x } 0.377) \\ & +(\text{Dist. from river x } 0.075) + (\text{Dist. from road x } 0.078) \\ & +(\text{Lithology x } 0.128) + (\text{Rainfall x } 0.175) \end{aligned} \quad \text{(Eq. 14)}$$

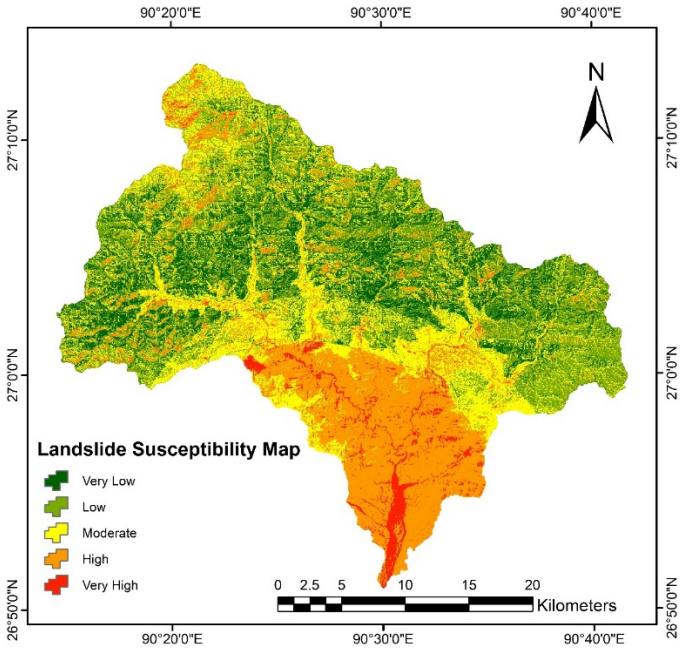

Figure 3 Landslide susceptibility mapping using the index of entropy.

A total of 24.81% (203.62 km^2) and 3.27% (26.84 km^2) area falls under high and very high susceptible zone, respectively and these areas are precisely falling in the risky area during the field verification. It was verified that among 164 landslides from the landslide inventory, 84 landslides were located in high susceptibility zone and 40 landslides were in very high landslide susceptibility zone. The field verification of the very high susceptibility zone concludes that it is not advisable to construct any infra-structure and agricultural activities in the very high risky area. However, in the high susceptible zone, locals can use the area for agricultural purposes with only inevitable engineering structures.

Table 4 The area and percentage of landslide susceptibility zones

Zone	Area (km^2)	Percentage
Very low	142.53	17.37
Low	233.50	28.45
Moderate	214.26	26.11
High	203.62	24.81
Very high	26.84	3.27

3) Landslide susceptibility map verification

The result of True Positive (TP), True Negative (TN), False Positive (FP), False Negative (FN), sensitivity, specificity, the area under the curve (AUC), and kappa index for the training and validation datasets are shown in Table 5.

The TP shows the correctly classified landslide pixel while the TN shows the correctly classified non-landslide pixel. The misclassified pixels are FP and FN Pixel. The FP is the wrongly classified non-landslide pixel as the landslide pixel while FN is incorrectly classified landslide pixel as the non-landslide pixel. The number of correctly classified (TP and TN) and misclassified (FP and FN) pixels for training and validation datasets were given in Table 5.

Table 5 Model performance on the training and validation datasets

	Training dataset	Validation dataset
True positive (TP)	86	45
True negative (TN)	77	28
False positive (FP)	38	21
False negative (FN)	30	4
Sensitivity	0.7414	0.9184
Specificity	0.6696	0.5714
AUC	0.7821	0.8377
Kappa	0.4111	0.4898

The AUC of the success rate and prediction rate were 0.7821 and 0.8377 (Figure 4 (a) and Figure 4(b)). As per Shirani et al. [8], the success rate (0.7821) and falls under the good category (0.7-0.8) while the prediction rate (0.8377) falls under a very good category (0.8-0.9). This concludes that the LSM is accurate enough for the future references by the researchers and engineers.

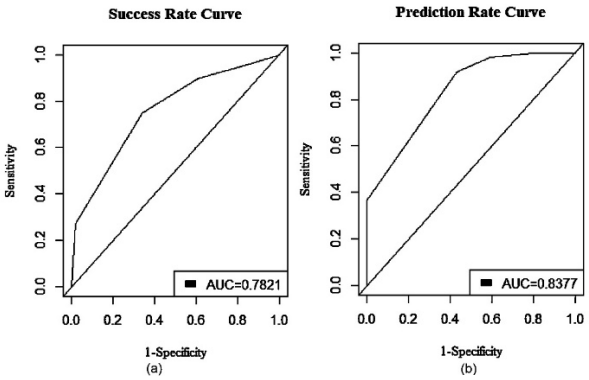

Figure 4 Model validation: (a) Success rate curve, (b) Prediction rate curve.

As per the Landis and Koch [26], both the kappa index for the training dataset (0.4111) and the validation dataset (0.4898) fall under the moderate (0.4-0.6) category. Therefore, the index of entropy model is a reliable model for developing the LSM.

Conclusions

This paper presents LSM using the IOE at the Ossey watershed area in Bhutan and the LSM is classified into five classes based on the severity of landslide using the geometrical interval. It is not advisable to plan developmental activities in the high and very high landslide susceptibility zone which covers 24.81% for high and 3.81% very high landslide susceptibility zone. As per the IOE, the most influential factors were NDVI, slope, and rainfall with a weightage of 0.377, 0.347, and 0.175, respectively. The landslide probability density (P_{ij}) of Table 3 shows the relationship between the individual classes of the causal factors and the landslide events. The LSM has a predictive capability of 83.77% which is accurate enough for future use. Similarly, as per the kappa index for the training dataset and validation dataset are 0.4111 and 0.4898, which shows the IOE is a reliable model for the landslide assessment.

The geotechnical field investigation is recommended for future study at the few selected sites of the study area to further check the reliability of the developed landslides susceptibility map.

Acknowledgement

The authors would like to express sincere gratitude to His Majesty the King of Bhutan and Naresuan University, Thailand for funding this scholarship.

References

[1] Zhang, T., Han, L., Han, J., Li, X., Zhang, H., Wang, H. Assessment of landslide susceptibility using integrated ensemble fractal dimension with Kernel logistic regression model. Entropy, 2019, 21(2), 218.

[2] Liu, J., Duan, Z. Quantitative assessment of landslide susceptibility comparing statistical index, index of entropy, and weights of evidence in the Shangnan area, China. Entropy, 2018, 20(11), 868.

[3] EM-DAT, C. The International disaster database. Brussels, Belgium: UCLouvain, 2020.

[4] Nohani, E., Moharrami, M., Sharafi, S., Khosravi, K., Pradhan, B., Pham, B.T., ..., Melesse, AM. Landslide susceptibility mapping using different GIS-based bivariate models. Water, 2019, 11(7), 1402.

[5] Bui, D.T., Tuan, T.A., Klempe, H., Pradhan, B., Revhaug, I. Spatial prediction models for shallow landslide hazards: a comparative assessment of the efficacy of support vector machines, artificial neural networks, kernel logistic regression, and logistic model tree. Landslides, 2016, 13(2), 361–378.

[6] Bui, D.T., Pradhan, B., Revhaug, I., Nguyen, D.B., Pham, H.V., Bui, Q.N. A novel hybrid evidential belief function-based fuzzy logic model in spatial prediction of rainfall-induced shallow landslides in the Lang Son city area (Vietnam). Geomatics, Natural Hazards and Risk, 2015, 6(3), 243–271.

[7] Aleotti, P., Chowdhury, R. Landslide hazard assessment: summary review and new perspectives. Bulletin of Engineering Geology and the Environment, 1999, 58(1), 21–44.

[8] Shirani, K., Pasandi, M., Arabameri, A. Landslide susceptibility assessment by dempster–shafer and index of entropy models, Sarkhoun basin, southwestern Iran. Natural Hazards, 2018, 93(3), 1379-1418.

[9] Van Westen, C., Van Asch, T.W., Soeters, R. Landslide hazard and risk zonation—why is it still so difficult? Bulletin of Engineering Geology and the environment, 2006, 65(2), 167–184.

[10] Mandal, S., Mandal, K. Bivariate statistical index for landslide susceptibility mapping in the Rorachu river basin of eastern Sikkim Himalaya, India. Spatial Information Research, 2018, 26(1), 59–75.

[11] Zhu, L., Huang, J.-f. GIS-based logistic regression method for landslide susceptibility mapping in the regional scale. Journal of Zhejiang University-Science A, 2006, 7(12), 2007–2017.

[12] Keunza, K., Dorji, Y., Wangda, D. Landslides in Bhutan. Country report, Department of Geology and Mines, Royal Government of Bhutan, Thimpu, 2004, 8.

[13] Dikshit, A., Sarkar, R., Pradhan, B., Acharya, S., Dorji, K. Estimating rainfall thresholds for landslide occurrence in Bhutan the Himalayas. Water, 2019, 11(8), 1616.

[14] Dunning, S., Rosser, N., Petley, D., Massey, C. Formation and failure of the Tsatichhu landslide dam, Bhutan. Landslides, 2006, 3(2), 107–113.

[15] Guzzetti, F., Carrara, A., Cardinali, M., Reichenbach, P. Landslide hazard evaluation: a review of current techniques and their application in a multi-scale study, Central Italy. Geomorphology, 1999, 31 (1–4), 181–216.

[16] Regmi, A.D., Devkota, K.C., Yoshida, K., Pradhan, B., Pourghasemi, H.R., Kumamoto, T., Akgun, A. Application of frequency ratio, statistical index, and weights -of-evidence models and their comparison in landslide susceptibility mapping in Central Nepal Himalaya. Arabian Journal of Geosciences, 2014, 7(2), 725–742.

[17] Jaafari, A., Najafi, A., Pourghasemi, H., Rezaeian, J., Sattarian, A. GIS-based frequency ratio and index of entropy models for landslide susceptibility assessment in the Caspian forest, northern Iran. International Journal of Environmental Science and Technology, 2014, 11(4), 909–926.

[18] Youssef, A.M., Al-Kathery, M., Pradhan, B. Landslide susceptibility mapping at Al-Hasher area, Jizan (Saudi Arabia) using GIS-based frequency ratio and index of entropy models. Geosciences Journal, 2015, 19(1), 113–134.

[19] Sun, X., Chen, J., Bao, Y., Han, X., Zhan, J., Peng, W. Landslide susceptibility mapping using logistic regression analysis along the Jinsha river and its tributaries close to Derong and Deqin County, Southwestern China. ISPRS International Journal of Geo-Information, 2018, 7(11), 438.

[20] Moore, I.D., Grayson, R., Ladson, A. Digital terrain modelling: A review of hydrological, geomorphological, and biological applications. Hydrological Processes, 1991, 5(1), 3–30.

[21] Long, S., McQuarrie, N., Tobgay, T., Grujic, D., Hollister, L. Geologic map of Bhutan. Journal of Maps, 2011, 7(1), 184–192.

[22] Pourghasemi, H.R., Mohammady, M., Pradhan, B. Landslide susceptibility mapping using the index of entropy and conditional probability models in GIS: Safarood Basin, Iran. Catena, 2012, 97, 71–84.

[23] Tien Bui, D., Shahabi, H., Shirzadi, A., Chapi, K., Alizadeh, M., Chen, W., ..., Hong, H. Landslide detection and susceptibility mapping by airsar data using support vector machine and index of entropy models in Cameron highlands, Malaysia. Remote Sensing, 2018, 10(10), 1527.

[24] Constantin, M., Bednarik, M., Jurchescu, M. C., Vlaicu, M. Landslide susceptibility assessment using the bivariate statistical analysis and the index of entropy in the Sibiciu Basin (Romania). Environmental Earth Sciences, 2011, 63(2), 397–406.

[25] Cohen, J. A coefficient of agreement for nominal scales. Educational and Psychological Measurement, 1960, 20(1), 37–46.

[26] Landis, J.R., Koch, G.G. The measurement of observer agreement for categorical data. Biometrics, 1977, 159–174.

[27] Mondal, S., Mandal, S. Landslide susceptibility mapping of Darjeeling Himalaya, India using the index of entropy (IOE) model. Applied Geomatics, 2019, 11(2), 129–146.

[28] Singh, P., Sharma, A., Sur, U., Rai, P.K. Comparative landslide susceptibility assessment using statistical information value and index of entropy model in Bhanupali-Beri region, Himachal Pradesh, India. Environment, Development and Sustainability, 2020, 1–18.

[29] Kaur, H., Gupta, S., Parkash, S., Thapa, R., Gupta, A., Khanal, G.C. Evaluation of landslide susceptibility in a hill city of Sikkim Himalaya with the perspective of hybrid modelling techniques. Annals of GIS, 2019, 25(2), 113–132.

[30] Wang, Q., Li, W., Yan, S., Wu, Y., Pei, Y. GIS based frequency ratio and index of entropy models to landslide susceptibility mapping (Daguan, China). Environmental Earth Sciences, 2016, 75(9), 780.

[31] Alex, E.C., Ramesh, K., Sridevi, H. Quantification and understanding The observed changes in land cover patterns in Bangalore. International Journal of Civil Engineering and Technology, 2017, 8(4), 597–603.