



# Application of logistic regression (LR) and frequency ratio (FR) models for landslide susceptibility mapping in Relli Khola river basin of Darjeeling Himalaya, India

Goutam Das<sup>1</sup> · Kabita Lepcha<sup>1</sup>

© Springer Nature Switzerland AG 2019

## Abstract

Landslide is one of the important disasters taking place on earth, which may be either a natural or man-made process. Landslides are more active and disastrous in hilly and mountainous regions. The present study aims to identify the landslide susceptibility areas in the Relli river basin in Darjeeling Himalaya using logistic regression (LR) and frequency ratio (FR) models. The GIS techniques have been used for landslide susceptibility mapping. A total number of 67 landslide locations have been identified from Google Earth images and multiple field surveys. 70% of landslide locations have been randomly selected and used as training data set for preparing landslide susceptibility map, and the remaining 30% have been used as validation data set. For the present study, 20 different factors like drainage density, drainage texture, infiltration number, stream frequency, stream junction frequency, stream power index, lithology, soil, relative relief, slope, maximum relief, drainage intensity, ruggedness number, rainfall, dissection index, aspect, relief class, and distance from stream, topographic wetness index and land use land cover have been used. The application of the logistic regression and frequency ratio model has demonstrated that the lower catchment of the basin has been widely dominated by the most landslide susceptibility areas than other parts of the catchment. Almost 6.92 sq km (4.05%) and 7.44 sq km (4.36%) areas out of 170.61 sq km area of the basin have been observed as very high and high landslide susceptibility categories, respectively, for FR model and 5.75 sq km (3.37%) and 1.86 sq km (1.09%) areas have been under very high and high landslide susceptibility zones for LR model. Finally, the ROC curve has been used to validate the models. The prediction capabilities of the models seem significant as the area under the curve value ranges from 75 to 81%.

**Keywords** Landslide susceptibility mapping · Logistic regression (LR) · Frequency ratio (FR) · ROC curve · Darjeeling Himalaya

## 1 Introduction

Unexpected environmental conditions like tsunami, flood, earthquake, landslides, etc., that cause significant environmental and human loss on the surface of the earth are called natural disaster [48]. Landslide is one of the most dangerous and costly natural disasters of the earth, and it is a form of mass wasting also known as landslip or mudslide which may cause a wide range of ground movements [99]. Gravity of the earth is the main driving force

for landslides occurrences [56]. It is estimated that almost 9 percentages of worldwide natural disasters constitute landslides during the 1990s [35]. Almost 600 people are killed every year throughout the world as a result of slope instability [100]; Xie et al. [102]. Both people and the properties such as building, roads, communication networks, houses, agricultural land, forest are destroyed, every year due to landslides; thus, a huge amount of money is spent globally to mitigate the destruction of landslides [14].

✉ Goutam Das, itisgoutamdas@gmail.com; Kabita Lepcha, kabitaugb@gmail.com | <sup>1</sup>Department of Geography and Applied Geography, University of North Bengal, Shibmandir, Siliguri, West Bengal 734013, India.

SN Applied Sciences (2019) 1:1453 | <https://doi.org/10.1007/s42452-019-1499-8>

Received: 28 May 2019 / Accepted: 14 October 2019 / Published online: 21 October 2019

Himalayan region frequently faced different types of natural hazards, and landslide is one of the prominent which damages property, agriculture and human lives [9]. The area selected for the present study has also suffered a lot of damages due to landslides, triggered by heavy rainfall; thus, the area seems suitable site to evaluate the frequency and distribution of landslides [28]. Rai et al. [78] stated that more than 20,000 landslides were recorded in one day. According to Basu and Pal [9], several landslides have been induced in different parts of Darjeeling Himalaya due to intensive rainfall, earthquake and expansion of anthropogenic activities such as road construction, building construction, resort formation. In 2015 more than 38 people were killed and 500 people were displaced due to landslides during monsoon season in this area [92]. Thus, the landslide mitigation and susceptibility studies become one of the most required fields of studies in this region.

There are several quantitative and qualitative methods for preparing landslide susceptibility maps [3]. Quantitative approaches have been used by number of researchers such as bivariate regression analysis [11, 51], multivariate regression analysis [62, 75, 74], logistic regression analysis [73, 2, 32], fuzzy logic [71], artificial neural network [57, 72]. Many researchers have worked with more than one model and compared to find out which one is most accurate [9]. Recently, machine learning (ML) techniques have become popular in spatial prediction of natural hazards studies such as wildfire [45], sinkhole [91], groundwater and flood [1, 12, 16, 17, 40, 49, 50, 59, 76, 77, 82, 93], droughtiness [80], gully erosion [7, 98], earthquake [4], land/ground subsidence [96] and landslide studies [68, 70, 79, 87, 67, 97]. ML is a subdivision of artificial intelligence (AI) that uses computer techniques to analyze and forecast information by learning from training data. ML algorithms that have been used for landslide prediction include support vector machine [20, 30, 47, 69, 95], artificial neural network [31, 86], decision trees such as naïve Bayes tree (NBT) [22, 85], radial basis function (RBF) [38], kernel logistic regression (KLR) [13, 21], Bayes' net (BN) [19], bivariate statistical index (SI) [23], stochastic gradient descent (SGD) [94], particle swarm optimization (PSO) [18], best-first decision tree (BFDT) [24], random subspace-based support vector machines (RSSVM) [39] and logistic model tree (LMT) [20]. Ensemble models have been used in landslide susceptibility mapping due to their novelty and their ability to comprehensively assess landslide-related parameters for discrete classes of independent factor [15, 16, 25, 44, 49, 63, 67, 87, 93]. To achieve this, a frequency ratio (FR) and logistic regression (LR) models have been applied to obtain maps of landslide susceptibility (spatial prediction) using the ArcGIS software (version 10.2) for the area. The frequency ratio model is useful to analyze the slope instability and treated as one of the best

quantitative approaches [43]. Literature reviews show that the logistic regression model is so accurate to support vector machine, classification tree and likelihood ratios [29, 83]. The logistic regression model is one of the most effective mathematical methods which are useful to find out the relationship between landslide causative factors and landslide locations [8, 28].

Landslide susceptibility mapping is important to ensure the safety of human life and mitigate the negative impact on regional as well as the national economy of a country [48]. This map helps government agencies, policymakers and planners to reduce the damages that a landslide incident can cause. The main objective of this study is to locate probable landslide susceptible areas within the basin using FR and LR models and compare the results using the ROC curve for the most suitable or acceptable method between these two (FR & LR) models.

## 2 Study area

Relli Khola river basin or Relli River is a small Himalayan river of the Indian States of West Bengal flowing through Kalimpong district. The river originates between the Alagara and Lava forest range at an elevation of 2400 meters known as t'Tiffin Dara and joins Teesta River as one of its tributaries. The length of the river is almost 27.38 km. A small village known as Relli is situated on its bank. During Makar Sankranti (January 14), a fair is held annually at the Relli. The basin extends from 26°58'2" to 27°5'31"N and from 88°26'32" to 88°39'14"E which is about 170.61 sq km. Relli Khola river basin is the left bank tributary of the Teesta River. Being a part of the Himalayan region, the area is characterized by intensive rainfall. The total basin is situated over two geological units. From top to bottom, the two geological units are Darjeeling Gneiss, and slate, schists, quartzite. The basin has natural beauty for its surrounding environment. Its culture is full of diversity, people from different parts live in the area. But the area is not safe from natural difficulties. So many natural disturbances such as earthquakes and landslides are common phenomena of the basin. Intensive rainfall (200–250 cm/annually) and a moderate-type temperature are the common characters of the basin. Being a forested area the basin is suffered few for landslide incidents. But where lands are open and deforestation takes place, landslides damage the area. In the lower course of the river, Komesi forest, Suruk Khasmahal and Mezok forest landslides have mainly been observed. But due to low habited place, damages are not found big. The portion that is close to Kalimpong town is a highly settled area and this portion is characterized as a gentle uniform slope, and thus, no such noticeable landslides are found in this portion (Fig. 1).

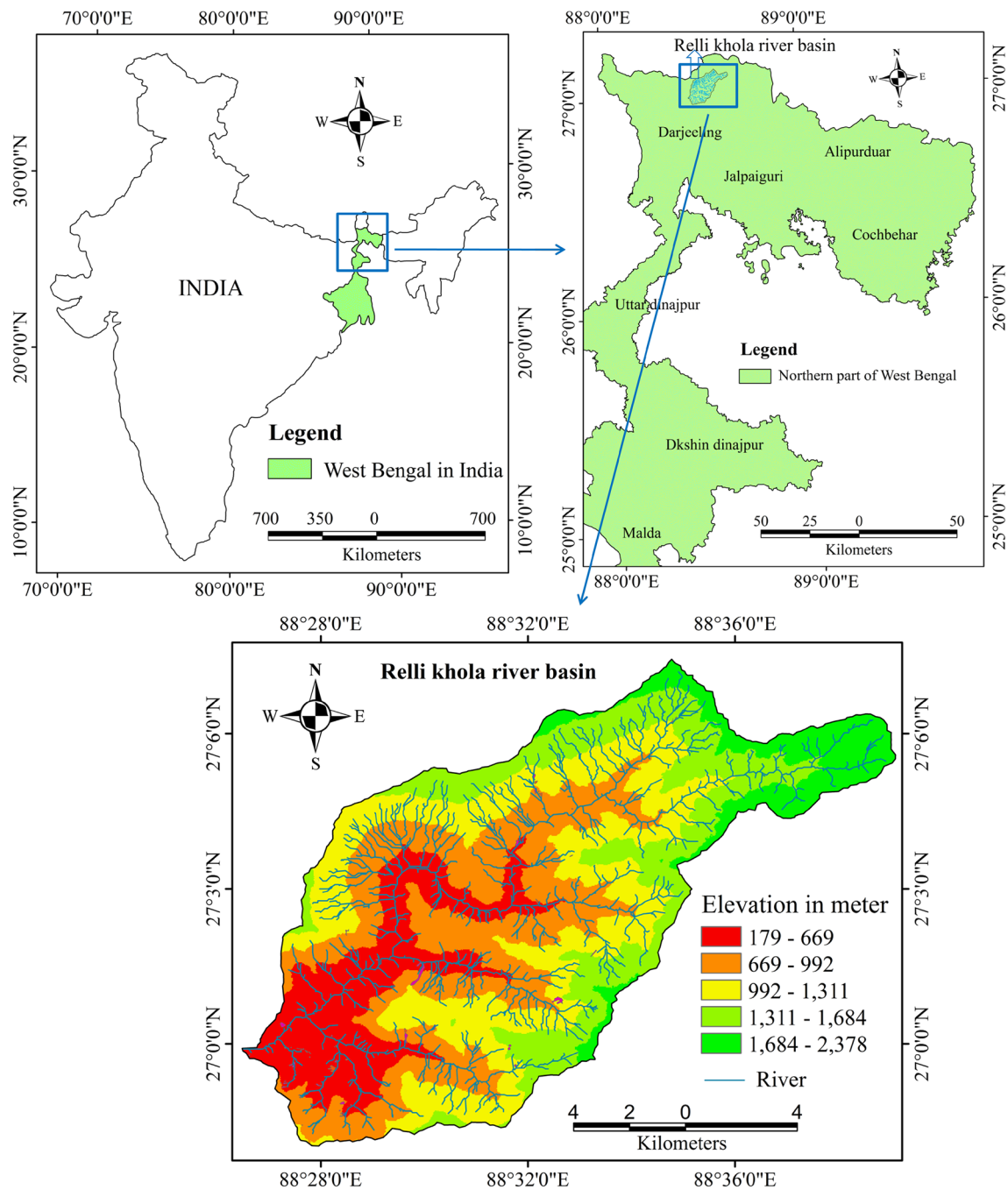


Fig. 1 Location map of the study area

### 3 Data and methods

A digital elevation model (DEM) with the resolution of  $30\text{ m} \times 30\text{ m}$  has been extracted from the ASTER GDEM data of October 2011 and downloaded from USGS Earth Explorer on January 18, 2018. The DEM data have been applied to extract slope, aspect, relief parameters (relative relief, maximum relief, dissection index, TWI, stream power index, etc.) and drainage parameters such as drainage

density, drainage frequency, stream junction frequency, stream junction angle and infiltration number of the study area using ArcGIS 10.2 software. To apply the models, 20 parameters have been selected. These are (a) drainage density, (b) drainage texture, (c) infiltration number, (d) stream frequency, (e) stream junction frequency, (f) stream power, (g) lithology, (h) soil, (i) relative relief, (j) slope, (k) maximum relief, (l) drainage intensity, (m) ruggedness number, (n) rainfall, (o) dissection index, (p) aspect, (q)

relief class and (r) distance from stream, (s) TWI and (t) land use land cover (lulc). These factors have been further classified into five subclasses such as (1) drainage factors, (2) relief factors, (3) hydrological factors, (4) lithological factors and (5) triggering factors. The methodologies for each parameter as stated above have been provided in detail in the following sections (Table 1).

### 3.1 Drainage factors

#### 3.1.1 Drainage density (Dd)

Drainage density is the length of stream per unit area of a river basin. Landslides are prominent in such areas where drainage density is high and the soil layer is too thin [65]. Figure 2a shows the drainage density of the Relli Khola river basin. The drainage density of the basin ranges from 0.13 to 5.84 km/sq km (Fig. 2a). The formula is given below [41].

$$Dd = \frac{L\mu}{A} \tag{1}$$

where Dd is drainage density,  $L\mu$  is the length of the stream and A is the total area. The grid method has been used to calculate the specific drainage density of the basin.

#### 3.1.2 Stream frequency (Fs)

Stream frequency is one of the important factors for landslide susceptibility measuring. Stream frequency (Fs) is the number of streams per unit area of the basin [41]. The stream frequency of the basin ranges from 0 to 20 stream/sq km (Fig. 2b). The value close to 0 means less diversity of slope and less landslide susceptible areas, and the higher value means high diversity of slope and high probability of landslides. Stream frequency is calculated in the following way [41]

$$Fs = \frac{N\mu}{A} \tag{2}$$

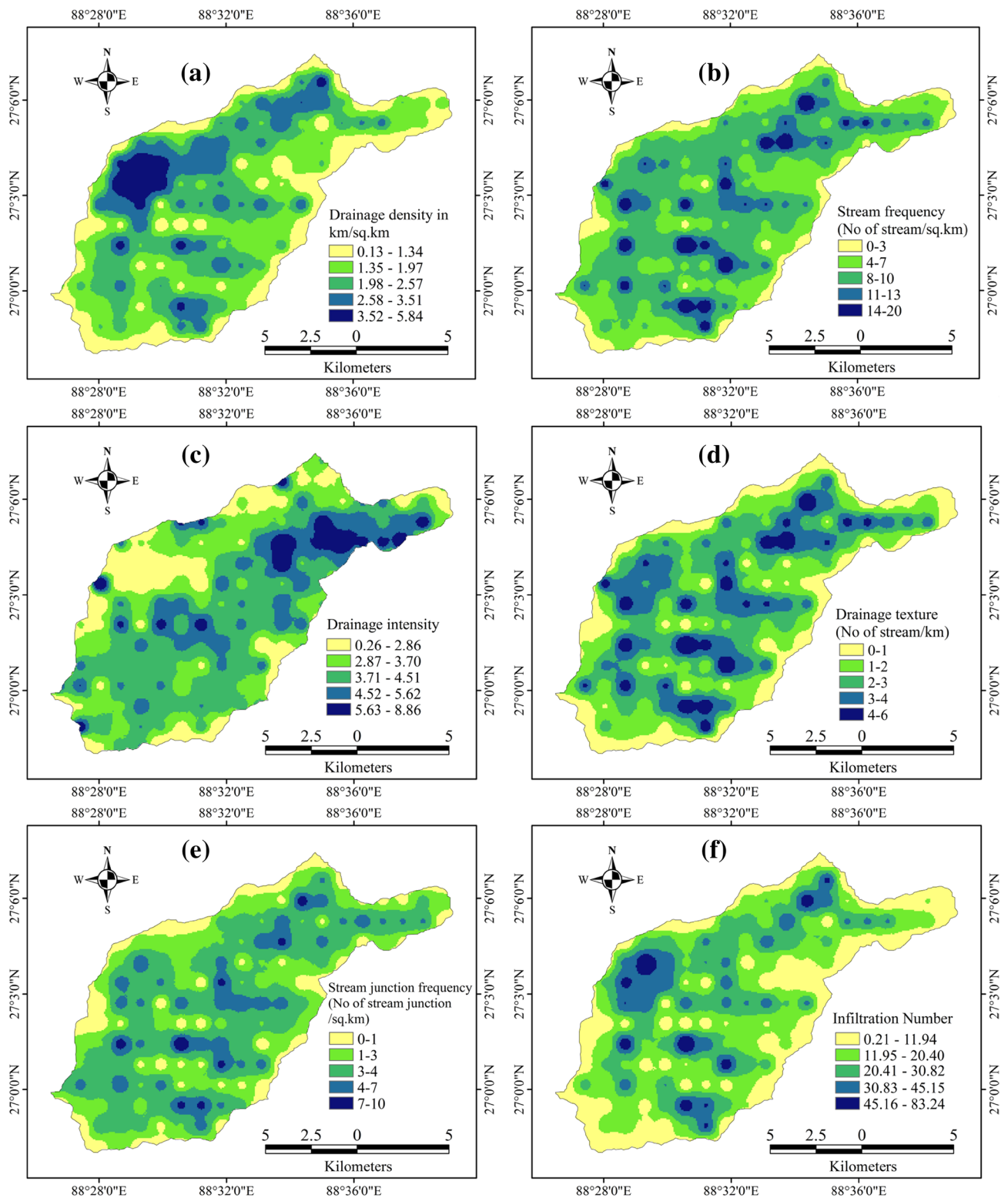
where Fs is stream frequency,  $N\mu$  is the total number of stream and A is the total area of the basin or region adopted for the present study.

#### 3.1.3 Drainage intensity (Id)

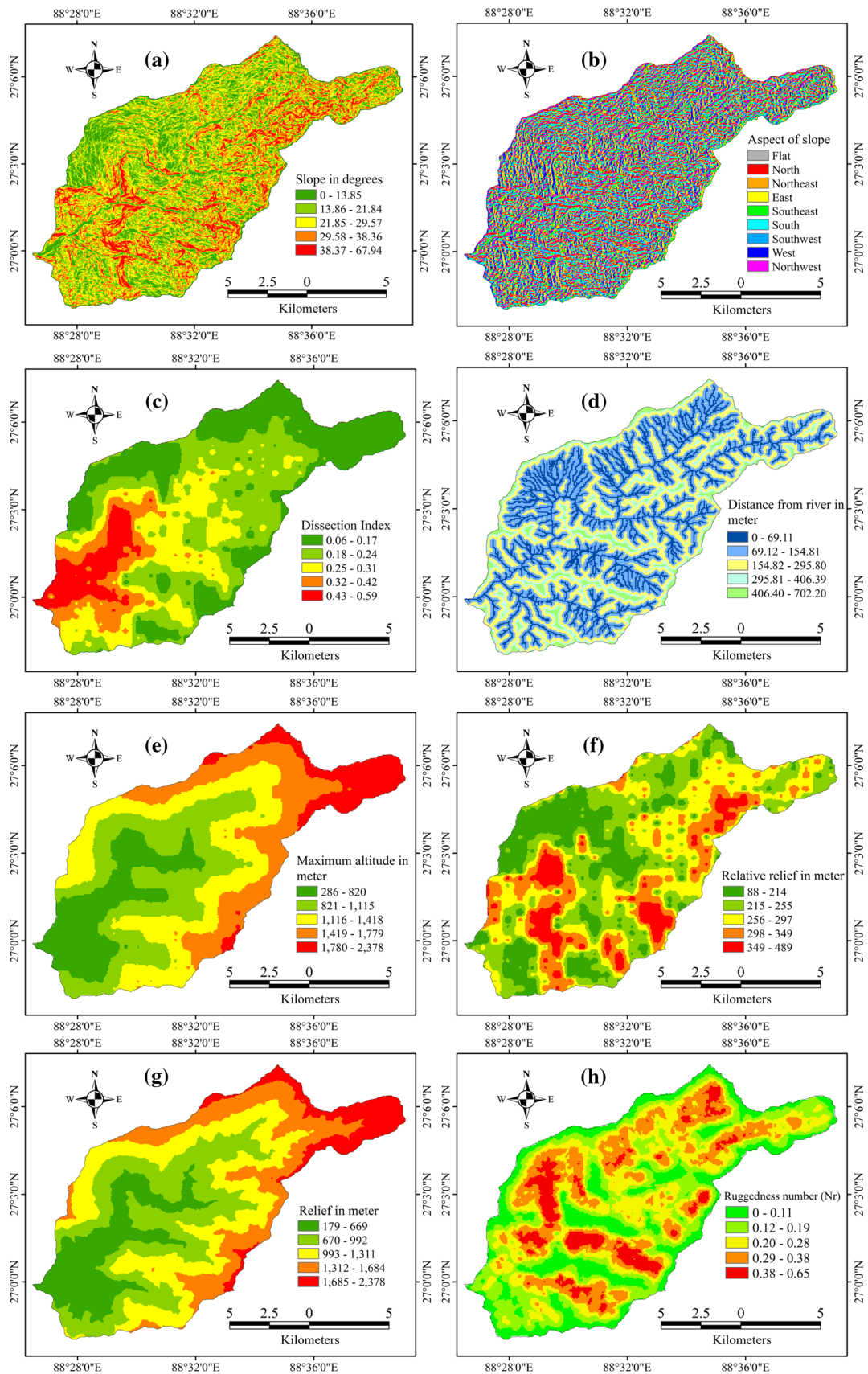
Drainage intensity (Id) denotes the ratio between stream frequency (Fs) and drainage density (Dd) [33]. The value ranges from 0.26 to 8.86 (Fig. 2c). High drainage intensity indicates a high probability of landslide, and low drainage intensity indicates less probability of landslide. Faniran [33] calculated drainage intensity in the following way.

**Table 1** Database of the current study

Datasets	Parameters	Source	Scale or resolution	Classification method
ASTER GDEM	Slope	USGS Earth Explorer	30 m × 30 m	Natural break
	Aspect		30 m × 30 m	Equal interval
	Dissection index		30 m × 30 m	Natural break
	Distance from river		30 m × 30 m	Natural break
	Maximum relief		30 m × 30 m	Natural break
	Relative relief		30 m × 30 m	Natural break
	Relief		30 m × 30 m	Natural break
	Ruggedness number		30 m × 30 m	Natural break
	Drainage density		30 m × 30 m	Natural break
	Stream frequency		30 m × 30 m	Natural break
	Drainage intensity		30 m × 30 m	Natural break
	Drainage texture		30 m × 30 m	Natural break
	Stream junction frequency		30 m × 30 m	Natural break
	Infiltration number		30 m × 30 m	Natural break
Stream power index (SPI)	30 m × 30 m	Natural break		
Topographic wetness index(TWI)	30 m × 30 m	Natural break		
Geological map	Lithology	Geological Survey of India, Kolkata	1:250,000	Lithological units
Soil map	Soil	NBSS & LUP Regional Centre, Kolkata	1:250,000	Textural units
Rainfall map	Rainfall	<a href="https://pmm.nasa.gov/data-access/downloads/trmm">https://pmm.nasa.gov/data-access/downloads/trmm</a>	0.25° × 0.25°	Natural break
LANDSAT 8 OLI/TIRS	Land use land cover	USGS Earth Explorer	30 m × 30 m	Supervised classification



**Fig. 2** Raster layer of drainage parameter **a** drainage density, **b** stream frequency, **c** drainage intensity, **d** drainage texture, **e** stream junction frequency, **f** infiltration number



◀**Fig. 3** Raster layers of relief factors **a** slope, **b** aspect, **c** dissection index, **d** distance from river, **e** maximum relief, **f** relative relief, **g** relief and **h** ruggedness number

$$I_d = \frac{F_s}{D_d} \quad (3)$$

where  $I_d$  is drainage intensity,  $F_s$  is the stream frequency and  $D_d$  is the drainage density.

### 3.1.4 Drainage texture (T)

Drainage texture is also an important morphometric factor that indicates a relative spacing of streams per unit length. Drainage texture is the ratio between the number of streams and the length of the perimeter of the basin [42]. The value of drainage texture ranges from 0 to 6 stream per km (Fig. 2d). Horton [42] gave the formula for calculating drainage texture as stated below.

$$T = \frac{N\mu}{P} \quad (4)$$

where  $T$  is the drainage texture of the basin,  $N\mu$  is the number of streams and  $P$  is the perimeter of the basin.

### 3.1.5 Stream junction frequency

Stream junction frequency is the number of stream junction points within a unit area of a drainage basin. Being a part of the source region, i.e., mountainous region, the river has many stream junction points throughout the basin. Stream junction frequency indirectly indicates the slope's instability, because the break of slope occurs where two or more streams join in a single point. The value ranges from 0 to 10 stream junctions/sq km (Fig. 2e). Stream junction frequency is calculated in the following formula.

$$F_{sj} = \frac{f_j}{A} \quad (5)$$

where  $F_{sj}$  is the frequency of stream junctions,  $f_j$  is the number of stream junction points and  $A$  is the area of the basin

### 3.1.6 Infiltration number (If)

Faniran [33] also defines infiltration number ( $I_f$ ) as the multiplication of both stream frequency ( $F_s$ ) and drainage density ( $D_d$ ). The infiltration number value of the basin ranges from 0 to 176.19 (Fig. 2f). The value close to 0 indicates the high infiltration and low surface runoff and

higher value indicate the opposite, i.e., low infiltration and high surface runoff. The infiltration number is calculated in the following ways [33].

$$I_f = F_s \times D_d \quad (6)$$

where  $I_f$  is the infiltration number,  $F_s$  is the stream frequency and  $D_d$  is drainage density.

## 3.2 Relief parameter

### 3.2.1 Slope

Slope is one of the most dominant factors of landslide occurrences [27, 6, 36]. Occurrences of landslides are directly affected by the slope angle of an area [53]. ASTER GDEM data have been used for preparing the slope map of the basin with 30 m resolution in ArcGIS 10.2 in degree form. The slope of the basin ranges from 0 to 67.48 (Fig. 3a)

### 3.2.2 Aspect

Aspect of slope describes the slope direction of an area. This is also an important factor of landslide and exposure to sunlight, drying winds, rainfall (degree of saturation), discontinuities are the aspect-associated parameters are also important factors of landslides [52]. Aspect map of the basin has been prepared from ASTER GDEM data with 30 m resolution in ArcGIS 10.2. Figure 3b shows the aspect map of the basin.

### 3.2.3 Dissection index

Dissection index is one of the most important factors to understand the relief as it is defined as the ratio between relative relief and absolute relief [64]. The value of DI of the area ranges from 0.06 to 0.58 (Fig. 3c). The lower portion of the basin is highly dissected, whereas the upper portion is less dissected which indicates the relative relief is low in case of the upper portion and high in case of lower portion of the basin. It is calculated in the following formula [64].

$$DI = \frac{Rr}{Ar} \quad (7)$$

where  $DI$  is the dissection index of the basin,  $Rr$  is relative relief and  $Ar$  is the absolute relief of the basin. The value of  $DI$  ranges from 0 to 1. 0 means complete absence of dissection, and 1 means vertical cliff.

### 3.2.4 Distance from stream

Distances from stream have been measured by using ArcGIS 10.2 software. It is also an important factor for landslides. The area close to streams can get water from the

streams that help the rock or soil to be fragile for eroding or sliding. Therefore, there is a possibility of landslides close to the streams. The map shows that the value of the distance from the stream of the basin ranges from 0 to 704.96 m (Fig. 3d)

### 3.2.5 Maximum relief

Maximum relief is simply defined as the highest altitude of an area. It is also known as absolute relief. The map is prepared by using the grid method in ArcGIS 10.2. The maximum relief of each grid has then been used to prepare the maximum relief map using IDW (inverse distance weighted) method. The value of maximum relief ranges from 286 to 2378 ms (Fig. 3e). The source region shows the maximum relief of the basin.

### 3.2.6 Relative relief

Relative relief is another form of representing the slope of a terrain. It is the difference between the highest altitude and lowest altitude. Therefore, a high relative relief zone has a chance of landslide. This map is also prepared using the grid method and IDW technique in ArcGIS. The formula is given in the equation. The maximum and minimum values of relative relief of the basin are 489 and 88 meters, respectively (Fig. 3f). Smith [88] gave the formula for calculating relative relief in this way.

$$RR = H - h \quad (8)$$

where RR is relative relief,  $H$  is the highest altitude and the lowest altitude of the basin.

### 3.2.7 Relief class

Relief class has been done based on the DEM classification of the basin. It is done basically to understand what range of relief class has been dominating the high landslide areas. The minimum and maximum reliefs of the basin are 179 and 2378 meters, respectively (Fig. 3g).

### 3.2.8 Ruggedness number (Nr)

Ruggedness number is a unitless value as both relative relief and drainage density are expressed in the same units and help to combine the slope steepness and [90]. The value of the ruggedness number of the basin ranges from 0 to 1.88 (Fig. 3h). The ruggedness number is calculated by using the following formula.

$$Nr = \frac{Dd \times Rr}{K(1000)} \quad (9)$$

where  $Dd$  is drainage density and  $K$  is a conversion constant (5280 in case of mile grid and when relative relief is expressed in feet and drainage density in miles/sq. mile and is 1000 when relative relief is expressed in meter and drainage density in meter/sq meter

## 3.3 Hydrological parameters

### 3.3.1 Stream power index (SPI)

Stream power index is the power of stream to move sediment, and thus, it is the potential of flowing water to complete geomorphic works such as incise, widen or aggrades of channel. It is estimated that if discharge or slope is increased, stream power is also increased proportionally. Stream power is low in the case of flat areas and high in the case of rugged topography. The stream power index is calculated with the following equation (Fig. 4a) [61]

$$SPI = A_s \times \tan \beta \quad (10)$$

where  $A_s$  is the specific catchment's area (sq m/m) and  $\beta$  is the slope gradient

### 3.3.2 Topographic wetness index (TWI)

Beven and Kirkby [10] developed the topographic wetness index (TWI) which is commonly used to measure and quantify the topographic control on hydrological processes [89]. If the moisture in the soil is high, the strength of soil will decrease and this enhances landslides. Wilson and Gallant [101] defined TWI in the following way (Fig. 4b).

$$TWI = \frac{A_s}{\tan \beta} \quad (11)$$

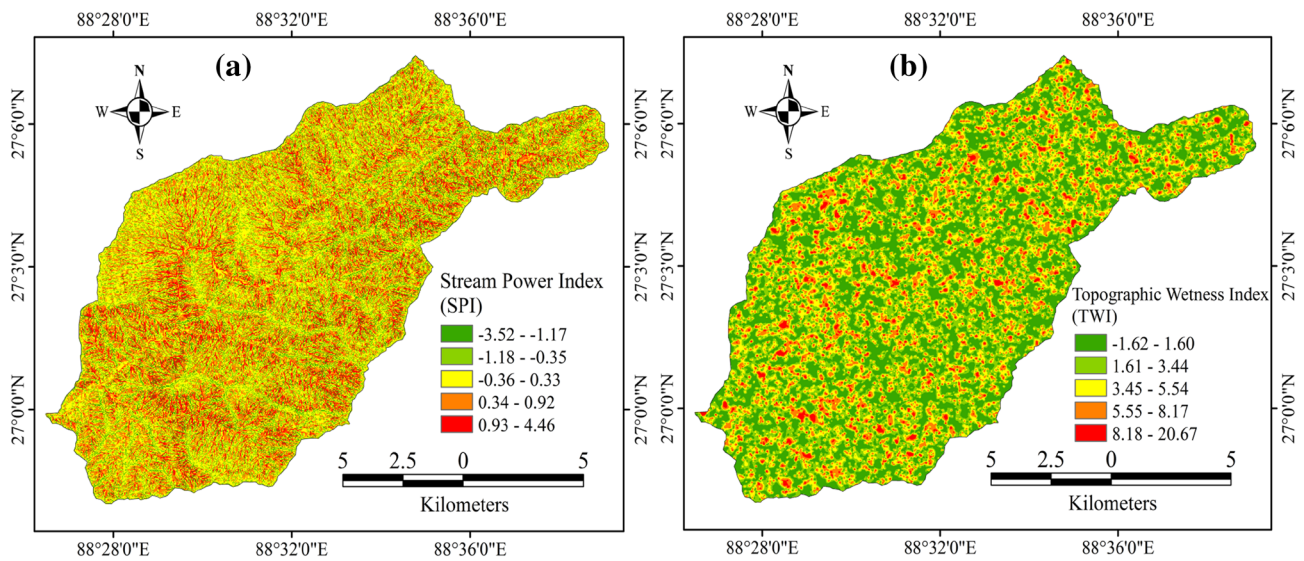
where  $A_s$  is the specific catchment's area (sq m/m) and  $\beta$  is the slope gradient.

## 3.4 Lithological factors

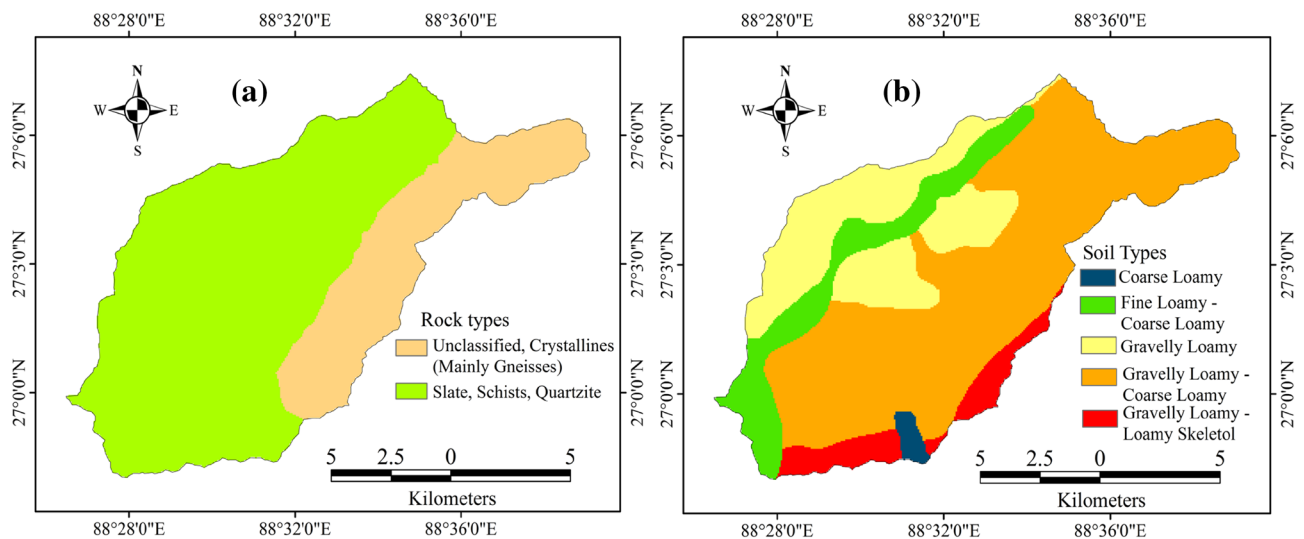
### 3.4.1 Lithology

Himalayan mountain region belongs to a very special convergence zone of two plates, e.g., (a) Indian Plate and (b) Eurasian Plate. Therefore, small and medium earthquakes have been occurring throughout the years. Landslide is also related to the earthquake. An earthquake can increase the rate and dimension of landslide. The lithological unit map has been collected from the Geological Survey of





**Fig. 4** Raster layers of hydrologic parameters **a** stream power index (SPI), **b** topographic wetness index (TWI)



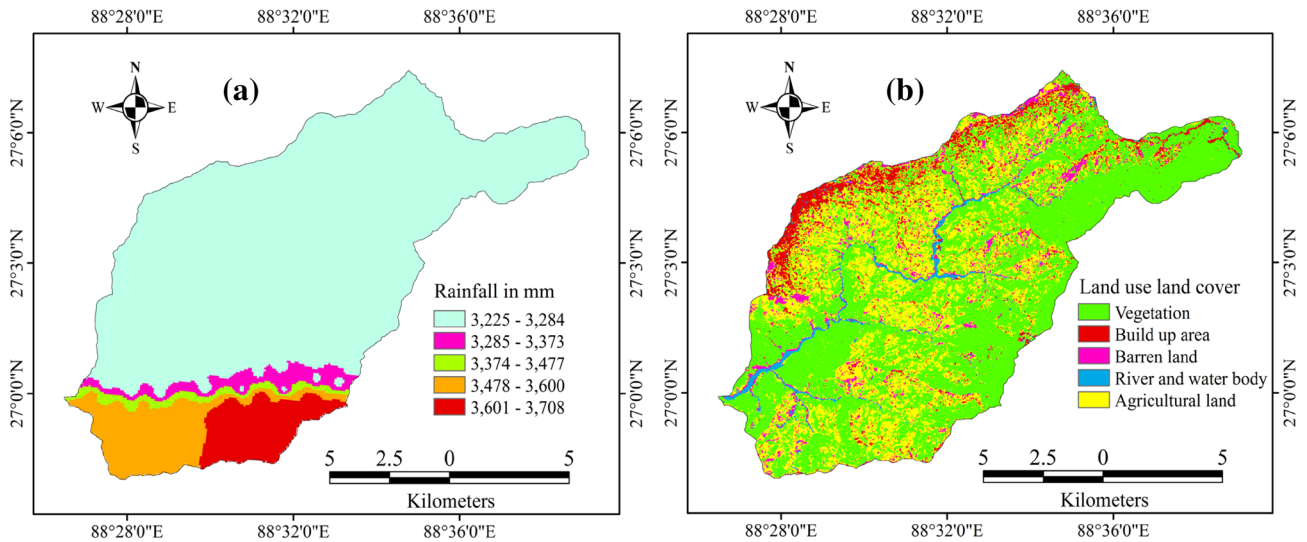
**Fig. 5** Lithological parameters **a** lithology and **b** soil

India and was grouped into two classes according to their character and lithological ages (Fig. 5a).

### 3.4.2 Soil

Five soil classes are observed in the basin (Fig. 6b). Soil demarcates the land use pattern. Soils having shallow depth

on steep slopes are affected most by landslides [84]. Soil map has been collected from the National Bureau of Soil Survey (NBSS) and Land Use Planning (LUP), Kolkata.



**Fig. 6** Raster layer of triggering factors **a** rainfall and **b** land use land cover

### 3.5 Triggering factors

#### 3.5.1 Rainfall

The most important triggering factor of landslide is rainfall. In Darjeeling Himalaya, intensive rainfall (almost 300 to 350 cm annually) is recorded every year. The duration of rainfall is also important for landslide. Landslides mainly occur during the monsoon season (July–Aug). The mean annual rainfall map has been prepared using TRMM (Tropical Rainfall Measuring Mission) data of the last 19 years (1998–2017) for the study area and downloaded from <https://pmm.nasa.gov/data-access/downloads/trmm> website. The thematic layer of rainfall has been prepared using the interpolation method of IDW (Inverse Distance Weighted) in a GIS platform. Figure 6a shows the annual rainfall map of the basin.

#### 3.5.2 Land use land cover (lulc)

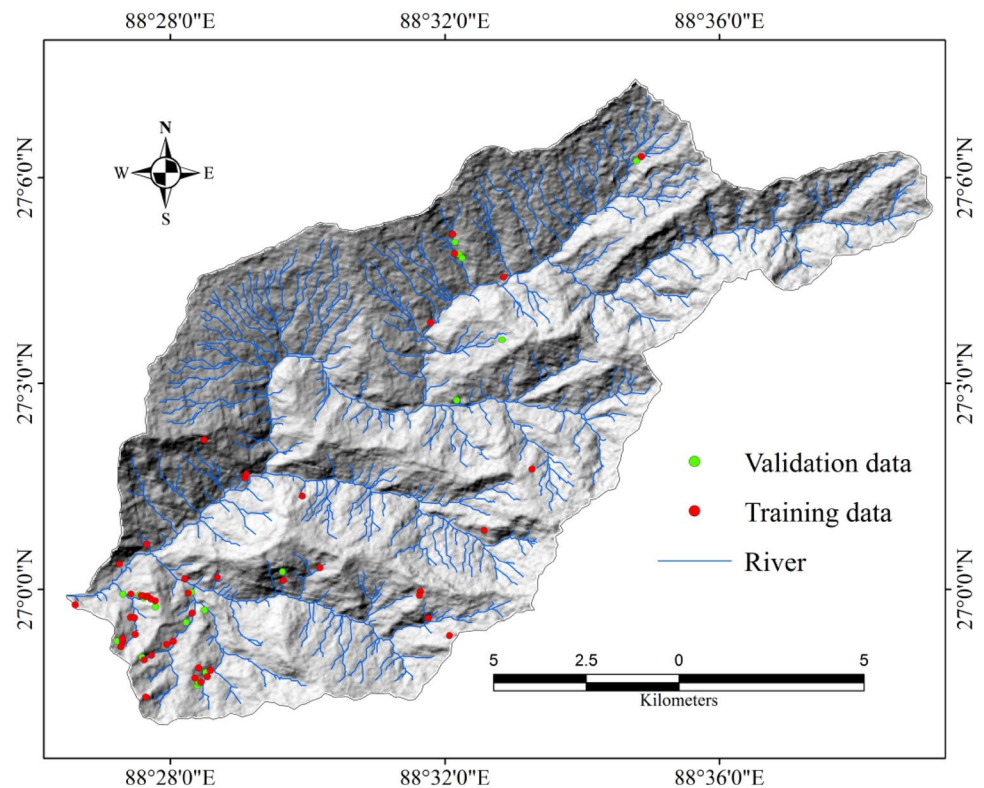
The upper layer of the earth's crust is used for different purposes by man. Each land use land cover has different intensities of landslide, e.g., forest can reduce landslide rate and open bare surface; build-up areas can increase the rate of the landslide. Five land use classes have been identified based on supervised image classification (Fig. 6b). The Landsat 8 OLI images with the resolution of 30 m × 30 m have been used to extract land use map using ArcGIS 10.2 software. The Landsat 8 OLI images have been downloaded from USGS Earth Explorer on January 18, 2018.

### 3.6 Landslide inventory map

Guzzetti et al. [37] stated that the landslide inventory map is an important part of analyzing landslide susceptibility, hazard as well as risk assessment. A landslide distribution map or landslide inventory map (Fig. 2) has been prepared to determine the landslide affected areas (%) and frequency of landslides of each class of possible landslide causing factors [60]. The landslide locations have been identified using Google Earth imagery and multiple field survey to cross-check the prepared landslide map. Ten-day (December 30, 2017, to January 8, 2018) extensive field survey and observing Google Earth imagery have been done for identifying landslide locations. In this current study, a total of 67 landslides have been identified. Out of the total, 47 (70%) landslides have been used as a training data set and 20 (30%) landslides have been used as validation data set. After that landslide inventory map has been prepared in the GIS environment to run the models and identify the probable landslide susceptible areas (Fig. 7). All the possible landslide causing factors have been incorporated with this landslide inventory map to understand the degree of importance of each possible landslide causing factors [60].

### 3.7 Frequency ratio model

The frequency ratio model is a well-accepted and popular quantitative approach for preparing landslide susceptibility mapping [60]. Lee and Talib [55], Lee and Pradhan [54], Jadda et al. [46], Avinash and Ashamanjari [5], Mondal and Maiti [60] successfully applied frequency ratio model for preparing susceptible map. To obtain the frequency ratio of each

**Fig. 7** Landslide inventory map of the study area

class of all the data layers, the following equation has been applied.

$$FR = \frac{N_{\text{pix}(S_i)} / N_{\text{pix}(N_i)}}{\sum_j N_{\text{pix}(S_j)} / \sum_j N_{\text{pix}(N_j)}} \quad (12)$$

where  $N_{\text{pix}(S_i)}$  is the number of landslide pixels containing in class  $i$ ,  $N_{\text{pix}(N_i)}$  is the total number of landslide pixels having class  $i$  in the watershed,  $\sum_j N_{\text{pix}(S_j)}$  is the total number of pixels containing landslide and  $\sum_j N_{\text{pix}(N_j)}$  is the total number of pixels in the watershed. The total number of pixels containing landslide is 400 out of 189,500 pixels (almost 170.61 sq km) in the watershed. Most of the landslides in the study area are rainfall-induced shallow types. The derived frequency ratio value of more than 1 means strong and positive relationship between landslide occurrences and concerned class of the selected data layers and high landslide probability; on the other hand, frequency ratio value of less than 1 means poor and negative correlation between landslide occurrences and concerned class of the data layer and low landslide probability, whereas 1 means moderate relationship. After calculating the frequency ratio, all the raster map parameters of frequency ratio have

been summed up to make landslide susceptibility index value (LSIV) using the following equation.

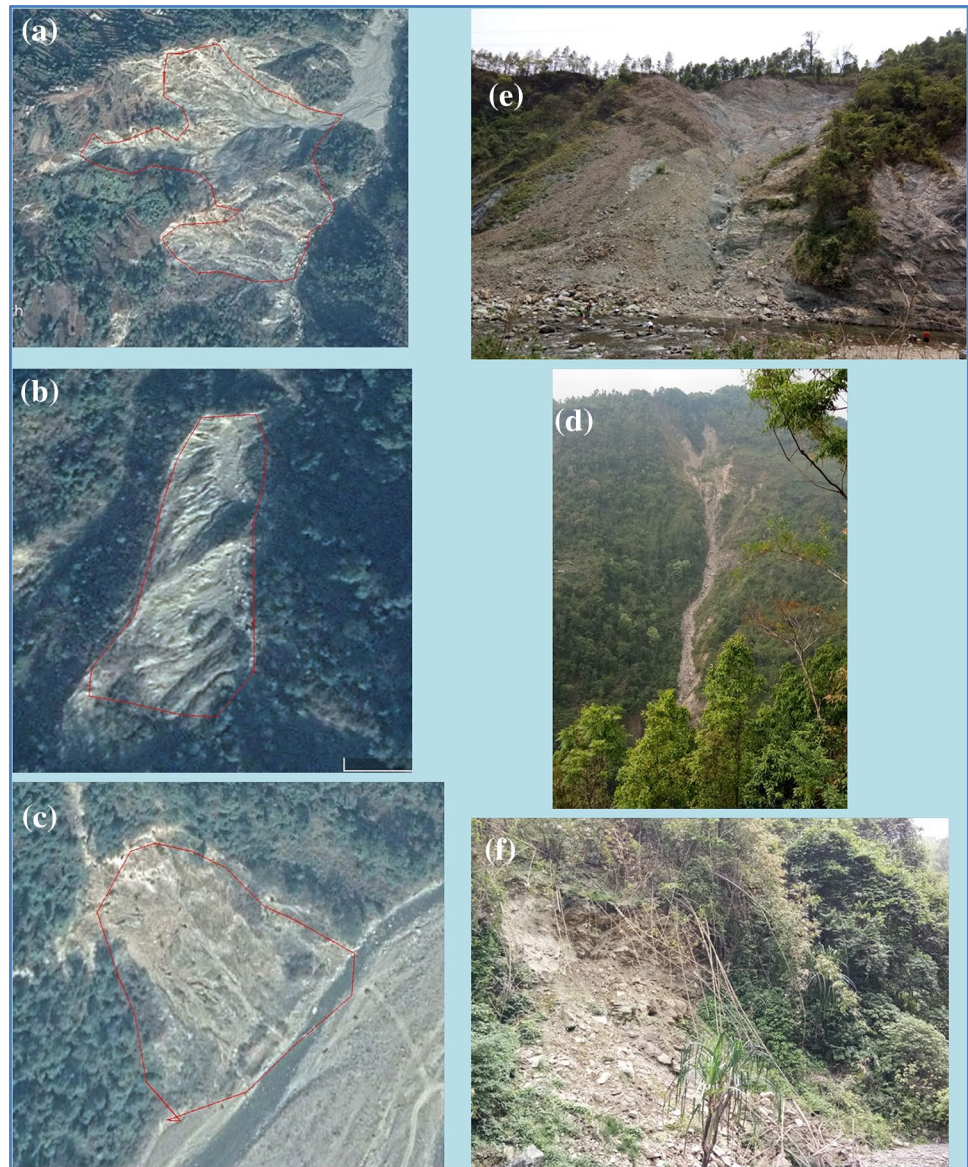
$$LSIV = Fr_1 + Fr_2 + \dots + Fr_n \quad (13)$$

where LSIV is landslide susceptibility index value and  $Fr_1, Fr_2, Fr_n$  is the frequency ratio of the raster data layers and  $n$  is the total number of factors for the study [20]. Higher value indicates high landslide susceptibility, and lower value indicates low susceptibility and vice versa.

### 3.7.1 Logistic regression model

The logistic regression model is also known as multivariate analysis is measured with dichotomous variables such as 1 or 0 (presence or absence), and it is determined by one or more independent variables [58]. The general purpose of the study is to determine the best-fitting model to describe the relationship between dependent variables (Landslide occurrence) and many independent variables, e.g., slope, lulc, lithology, rainfall, etc. (Kavzoglu et al. 2013). Dai and Lee [27] argued that the advantage of the model is that the dependent variable can have only two values, presence (Value 1) or not presence (Value 0). The logistic

**Fig. 8** Picture showing the landslide sites—from the Google Earth image **a** near Suruk Khasmahal (26°59'31.03"N, 88°28'14.43"E), **b** Mezoek forest (26°59'36.23"N, 88°27'28.89"E), **c** near confluence with Teesta River (27°00'22.89"N, 88°27'15.37"E) and from the field **d** Dalapchan Slip Reserve Forest (27°04'53.47"N, 88°32'08.78"E), **e** near Komesi Forest (27°01'37.43"N, 88°29'05.67"E) and **f** Suruk Khasmahal (26°58'51.62"N, 88°28'25.30"E)



regression model is based on the generalized linear model and can be calculated by the following equation.

$$P = 1 / (1 + e^{-z}) \quad (14)$$

where  $p$  is the probability of landslide occurrence and  $z$  is the linear regression model.

$$z = b_0 + b_1x_1 + b_2x_2 + \dots + b_nx_n \quad (15)$$

$b_0$  is the intercept of the model,  $n$  is the number of independent variables,  $b_1, b_2, \dots, b_n$  are the coefficients and  $x_1, x_2, \dots, x_n$  are the landslide causing factors.

$P$ , probability of vulnerability, varies from 0 to 1. Whenever it is nearer to 1, it indicates high vulnerable, and whenever it is nearer to 0, it indicates very low vulnerable. The entire calculation of logistic regression has been done in SPSS software (Fig. 8).

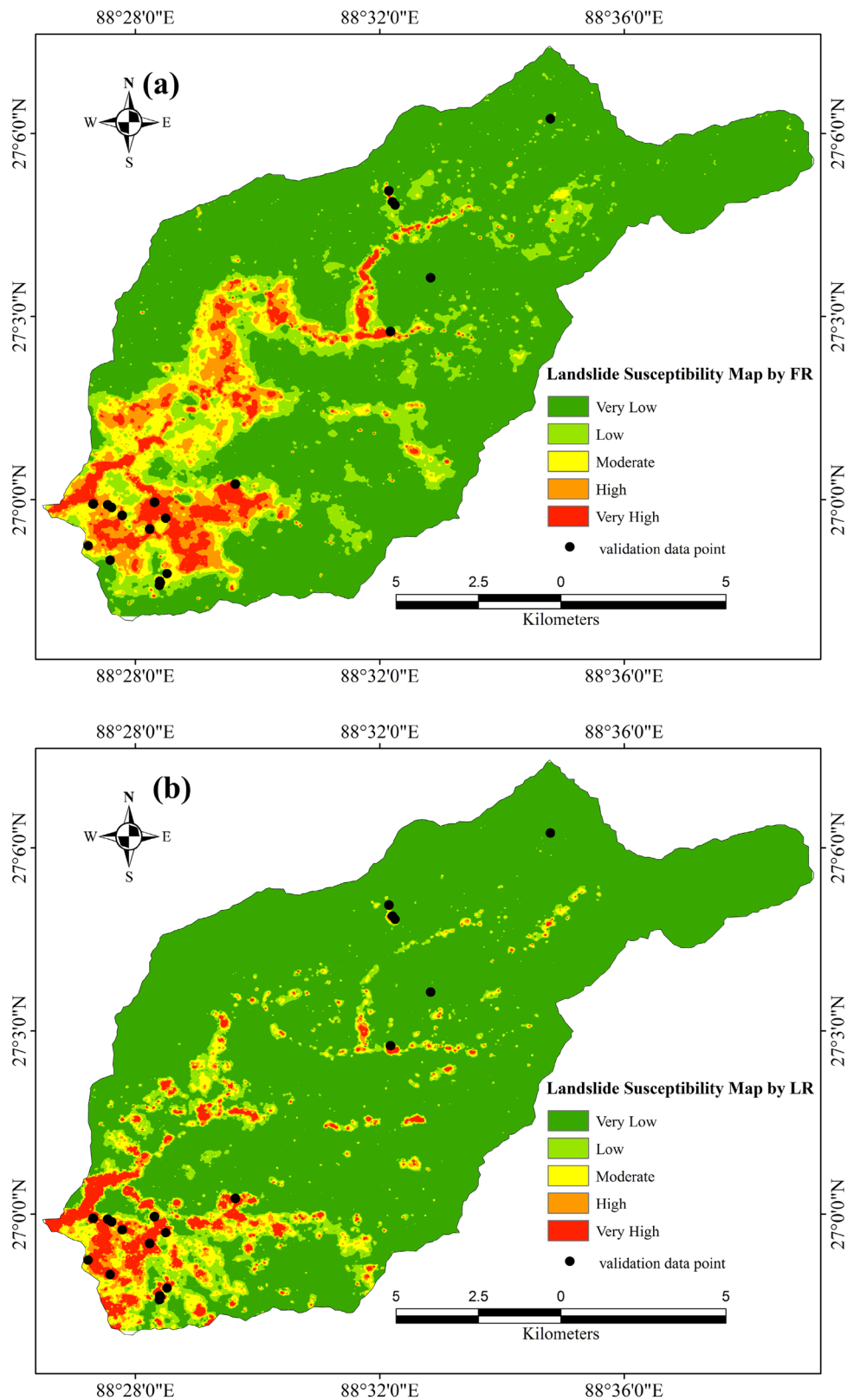
## 4 Results and analysis

FR and LR models have been developed to prepare two landslide vulnerability maps (Fig. 9) based on raster inputs, and each vulnerable map has been classified into five subtypes indicating varying intensity of vulnerability [66]

### 4.1 Frequency ratio model

The susceptibility map based on the frequency ratio model has been classified into 5 groups such as very high, high, moderate, low and very low which represent 4.05%, 4.36%, 9.12%, 14.34 and 68.11% area of the total basin, respectively (see Table 2). The lower portion of the basin has

**Fig. 9** Landslide susceptibility maps (a, b) of Relli Khola river basin (a) frequency ratio model and (b) logistic regression model



mainly been characterized as high landslide-prone area (Fig. 9). The landslide areas have been mostly dominated along the river channel of the basin.

The frequency ratio for different classes of each parameter helps us to understand the importance or probability of subclass under landslide occurrences. For example,

**Table 2** Area under different landslide vulnerable classes based on frequency ratio model and logistic regression model

Landslide vulnerable class	FR model			Logistic regression		
	Pixel	Area(sq km)	Area in %	Pixel	Area(sq km)	Area in %
Very low	129,068	116.2	68.11	152,343	137.16	80.39
Low	27,190	24.48	14.34	17,761	15.99	9.37
Moderate	17,292	15.57	9.12	10,943	9.85	5.77
High	8268	7.44	4.36	2062	1.86	1.09
Very high	7682	6.92	4.05	6391	5.75	3.37

frequency ratio of river and barren land (subcategories of lulc) is 12.84 and 5.74 indicating landslide vulnerable of these zones (Table 3). That is the significance of this model. The frequency ratio of the fourth class (5.56–8.17) of topographic wetness index (TWI) is high (2.04) compared to other subclasses of this parameter (Table 3). Therefore, it can be said that this zone is highly landslide prone and this zone is more responsible for landslide occurrences. Thus, this model helps us to understand the importance of each subclass.

Slope is a major parameter of landslide. The higher slope indicates the risk of landslide. In Table 3 it is observed that the frequency ratio of fifth (38.37 to 67.48 degrees) subclass of the slope is high (2.97). In the case of maximum relief, the result is different, i.e., higher value represents a low-frequency ratio and the lower value represents a high-frequency ratio (Table 3). It seems that not only maximum altitude is responsible for landslide but also other parameters are playing the dominant role for the basin's landslide. In the case of relative relief, it gives a good result with our view that higher relative relief (rugged topography) represents a high-frequency ratio, i.e., high probability of landslides (Table 3). The two classes of land use land cover (lulc), i.e., (a) river and water body and (b) barren land, have been dominated with high landslides as a frequency ratio of these two subclasses are 12.84 and 5.74 respectively (Table 3). Vegetation cover is the least subclass of land use land cover which interrupts landslide incidents. Slate, Schists, Quartzite rock is dominated by a high-frequency ratio (Table 3). The higher value of stream power index (SPI) indicates a high risk of landslide occurrences. It has been noticed that the frequency ratio of the fifth class of ruggedness number value is maximum (1.72) and lower classes have been gradually decreased. It seems that high ruggedness number values play a significant role in landslide occurrences.

## 4.2 Logistic regression model

Logistic regression is a useful method to determine the magnitude of the correlation between landslide locations and affective factors [34]. Table 2 shows that 5.75

sq km (3.37%) and 1.86 sq km (1.09%) areas of the total area of the Relli Khola basin are under very high and high landslide susceptibility zones, respectively. This map also shows almost the same result that the lower portion of the basin has been dominated by a highly landslide-prone area. Landslide is dependent on different factors (it can be physiographic or it can be anthropogenic), and logistic regression is useful to predict the future landslide trend based on the factors, and it also helps us to predict the most dominating factors of landslide occurrences [34]. Among the 20 parameters drainage density, slope, aspect, lulc, soil and rainfall have been identified as dominating factors of landslide occurrences (Table 4). The significance levels of these 3 parameters are 95 to 100. The steep slope is responsible for slope failure. The areas which have steep slope represent rugged topography and high gravity power and hence are at high risk of landslide occurrences. Lulc is also a major factor of landslide occurrences. It is the upper layer of the soil.

A particular land use has a high capability to slide such as bare land or open land tends high landslide. On the other hand, the forest cover area has a low tendency of landslide. Rainfall can increase the rate of landslide as it can make the soil fragile and detach soil from other soil molecules.

Table 4 shows the logistic regression result calculated in SPSS software. This result shows that the rate of landslide occurrences is noticeably and positively determined by stream frequency (Wald = 0.028, Exp(B) = 7.067,  $df = 1$ ), stream power (Wald = 0.001, Exp(B) = 1.001,  $df = 1$ ), lithology (Wald = 0.604, Exp(B) = 1.830,  $df = 1$ ), relative relief (Wald = 0.005, Exp(B) = 1.005,  $df = 1$ ), Slope (Wald = 0.048, Exp(B) = 1.049,  $df = 1$ ), ruggedness number (Wald = 2.152, Exp(B) = 8.604,  $df = 1$  and rainfall (Wald = 0.004, Exp(B) = 1.004,  $df = 1$ ) (Table 4). For the categorical factors of soil, aspect, some subclasses have been categorized as positively determined and some subclasses have been categorized as negatively determined. But for lulc it has been positively determined (Table 4). Other parameters and their logistic results are shown in Table 3. This result denotes that landslide occurrence is not controlled by a single parameter; rather, it is the result of multiple factors.

**Table 3** Class frequency ratios of selected parameters

Parameters	Classes	No of pixels [ $N_{pix(N_i)}$ ]	% of $N_{pix(N_i)}$	Landslide pixels [ $N_{pix(S_i)}$ ]	% of $N_{pix(S_i)}$	FR
Topographic wetness index (1)	−0.72–1.63	68,619	36.21	98	24.5	0.67
	1.64–3.46	52,812	27.86	107	26.75	0.96
	3.47–5.55	38,544	20.33	75	18.75	0.92
	5.56–8.17	22,958	12.11	99	24.75	2.04
	8.18–21.51	6567	3.46	21	5.25	1.51
	Total	189,500	100	400	100	–
Drainage density (meter per sq km) (2)	0–1.18	42,591	22.48	22	5.5	0.24
	1.19–2.13	45,827	24.18	141	35.25	1.46
	2.14–3.08	46,038	24.29	149	37.25	1.53
	3.09–4.14	38,079	20.09	84	21	1.05
	4.15–6.44	16,965	8.95	4	1	0.11
	Total	189,500	100	400	100	–
Soil types (3)	Gravelly Loamy	44,108	23.28	26	6.5	0.28
	Fine Loamy—Coarse Loamy	24,149	12.74	103	25.75	2.02
	Gravelly Loamy—Coarse Loamy	107,257	56.6	271	67.75	1.2
	Gravelly Loamy—Loamy Skeletal	11,810	6.23	0	0	0
	Coarse Loamy	2176	1.15	0	0	0
	Total	189,500	100	400	100	–
Lithology (4)	Unclassified, Crystallines (Mainly Gneisses)	49,182	25.95	62	15.5	0.6
	Slate, Schists, Quartzite	140,318	74.05	338	84.5	1.14
	Total	189,500	100	400	100	–
Distance from river (meter) (5)	0–99.20	94,672	49.95	223	55.75	1.11
	99.21–221.82	52,024	27.45	86	21.5	0.78
	221.83–313.70	22,744	12	79	19.75	1.64
	313.71–409.02	13,041	6.88	12	3	0.43
	409.03–704.96	7019	3.7	0	0	0
	Total	189,500	100	400	100	–
Slope in degrees (6)	0–13.76	30,057	15.86	15	3.75	0.23
	13.77–21.70	51,565	27.21	72	18	0.66
	21.71–29.37	52,423	27.66	104	26	0.93
	29.38–38.37	38,752	20.44	104	26	1.27
	38.38–67.48	16,703	8.81	105	26.25	2.97
	Total	189,500	100	400	100	–
Lulc (7)	Vegetation	104,234	55	93	23.25	0.42
	Settlement	12,521	6.61	15	3.75	0.57
	Barren land	9244	4.88	112	28	5.74
	River and water body	4354	2.3	118	29.5	12.84
	Agricultural land	59,147	31.21	62	15.5	0.5
	Total	189,500	100	400	100	–
Stream power index (8)	−1.81	18,669	9.85	33	8.25	0.83
	−1.37	34,109	17.99	36	9	0.5
	−0.69	67,570	35.65	97	24.25	0.68
	0.34–0.92	52,643	27.77	165	41.25	1.48
	0.93–4.46	16,509	8.71	69	17.25	1.98
	Total	189,500	100	400	100	–

**Table 3** (continued)

Parameters	Classes	No of pixels [ $N_{pix(N_i)}$ ]	% of $N_{pix(N_i)}$	Landslide pixels [ $N_{pix(S_i)}$ ]	% of $N_{pix(S_i)}$	FR
Relief (meter) (9)	179–669	32,737	17.28	171	42.75	2.47
	670–992	52,455	27.68	129	32.25	1.16
	993–1311	51,673	27.27	90	22.5	0.82
	1312–1684	35,740	18.86	10	2.5	0.13
	1685–2378	16,895	8.92	0	0	0
	Total	189,500	100	400	100	
Ruggedness number (10)	0–0.31	41,901	22.11	46	11.5	0.52
	0.32–0.56	48,182	25.42	103	25.75	1.01
	0.57–0.79	48,869	25.78	113	28.25	1.09
	0.80–1.06	36,506	19.26	87	21.75	1.12
	1.07–1.88	14,042	7.41	51	12.75	1.72
	Total	189,500	100	400	100	–
Maximum relief (meter) (11)	286.88–820.10	33,261	17.55	177	44.25	2.52
	820.11–1115.43	50,887	26.85	114	28.5	1.06
	1115.44–1418.95	50,550	26.67	99	24.75	0.92
	1485.96–1779.90	36,939	19.49	10	2.5	0.12
	1779.91–2378.75	17,863	9.42	0	0	0
	Total	189,500	100	400	100	
Aspect (direction of slope) (12)	Flat	3	0.0016	0	0	0.00
	North	28,988	15.29	74	18.5	1.21
	Northeast	16,193	8.54	28	7	0.82
	East	13,699	7.22	71	17.75	2.46
	Southeast	25,393	13.4	54	13.5	1.01
	South	34,587	18.25	62	15.5	0.85
	Southwest	23,806	12.56	85	21.25	1.69
	West	19,427	10.25	17	4.25	0.41
	Northwest	27,404	14.49	9	2.25	0.16
	Total	189,500	100	400	100	
Stream junction frequency (no. of stream junction/sq km) (13)	0–1.84	75,204	39.69	175	43.75	1.1
	1.85–4.09	61,354	32.38	104	26	0.8
	4.10–6.75	36,051	19.02	77	19.25	1.01
	6.76–10.54	14,101	7.44	38	9.5	1.27
	10.55–26.11	2790	1.47	6	1.5	1.01
	Total	189,500	100	400	100	
Infiltration number (14)	0–13.12	71,553	37.75	108	27	0.71
	13.13–31.09	55,675	29.37	164	41	1.39
	31.10–53.20	36,858	19.45	118	29.5	1.51
	53.21–82.22	19,215	10.13	10	2.5	0.24
	82.23–176.19	6199	3.27	0	0	0
	Total	189,500	100	400	100	
Relative relief in meter (15)	88.09–214.13	40,088	21.15	88	22	1.03,996
	214.14–255.09	54,803	28.92	73	18.25	0.63,106
	255.10–297.63	49,113	25.92	42	10.5	0.40514
	297.64–349.62	30,339	16.01	109	27.25	1.70206
	349.63–489.83	15,157	8.00	88	22	2.75054
	Total	189,500	100	400	100	



**Table 3** (continued)

Parameters	Classes	No of pixels [ $N_{\text{pix}(N_i)}$ ]	% of $N_{\text{pix}(N_i)}$	Landslide pixels [ $N_{\text{pix}(S_i)}$ ]	% of $N_{\text{pix}(S_i)}$	FR
Dissection index (16)	0.08–0.17	56,698	29.92	0	0	0
	0.18–0.24	60,244	31.79	88	22	0.69
	0.25–0.31	39,554	20.87	99	24.75	1.18
	0.32–0.42	17,229	9.09	171	42.75	4.7
	0.43–0.58	15,775	8.32	42	10.5	1.26
	Total	189,500	100	400	100	
Rainfall in mm/year (17)	3225.19–3284	152,058	80.24	248	62	0.77
	3284.01–3373.15	6114	3.22	45	11.25	3.48
	3373.16–3477.49	3546	1.87	19	4.75	2.53
	3477.50–3600.79	16,078	8.48	83	20.75	2.44
	3600.80–3708.92	11,704	6.17	5	1.25	0.2
	Total	189,500	100	400	100	–
Drainage intensity (per sq km) (18)	0.005–4.67	132,246	69.79	318	79.5	1.13
	4.68–18.08	55,969	29.54	82	20.5	0.69
	18.09–86.40	1229	0.65	0	0	0
	86.41–296.55	33	0.02	0	0	0
	296.56–595.44	23	0.01	0	0	0
	Total	189,500	100	400	100	
Stream frequency (no. of stream/sq km) (19)	0.001–5.00	37,028	19.54	44	11	0.56
	5.01–8.46	55,257.6	29.16	137	34.25	1.17
	8.47–12.04	50,706.5	26.76	148	37	1.38
	12.05–16.66	34,752.5	18.34	66	16.5	0.89
	16.67–32.68	11,755.3	6.20	5	1.25	0.2
	Total	189,500	100	400	100	
Drainage texture (No of stream per 1 km perimeter) (20)	0–0.66	37,028	19.53	44	11	0.56
	0.67–1.13	55,258	29.15	137	34.25	1.17
	1.14–1.62	50,707	26.75	148	37	1.38
	1.63–2.22	34,752	18.33	66	16.5	0.89
	2.23–4.37	11,755	6.2	5	1.25	0.2
	Total	189,500	100	400	100	

## 5 Validation of FR and LR models

After preparing landslide susceptibility, map validation is necessary. Otherwise, it has no use and has no scientific importance [26]. A receiver operating characteristic curve (ROC) has been used to validate these models. The ROC curve measures the goodness of fit from the area it falls under the curve [9]. There are five categories of AUC (area under the curve) value under the ROC curve such as excellent (0.90–1.00), good (0.80–0.90), fair (0.70–0.80), poor (0.60–0.70) and fail (0.50–0.60) to understand the accuracy level [81].

Figure 10 shows the ROC curves of landslide maps using FR and LR models. The ROC curves of both models have been prepared using SPSS software. The values of AUC (area under the curve) of the FR model and LR model are

0.814 and 0.751, respectively. The value of the FR model has indicated a good accuracy level with 81% area under the curve, and the value of the LR model has indicated a fair accuracy level having 75% area under the curve. Thus, it can be said that these two models have a considerable amount of accuracy and that can be used for further study.

## 6 Conclusion

Logistic regression and frequency ratio models have been used for the present study of landslide susceptibility of the Relli Khola river basin, a small tributary of the Teesta River in Darjeeling Himalaya. A total number of 20 possible parameters have been identified to prepare landslide susceptibility maps of the basin. Out of

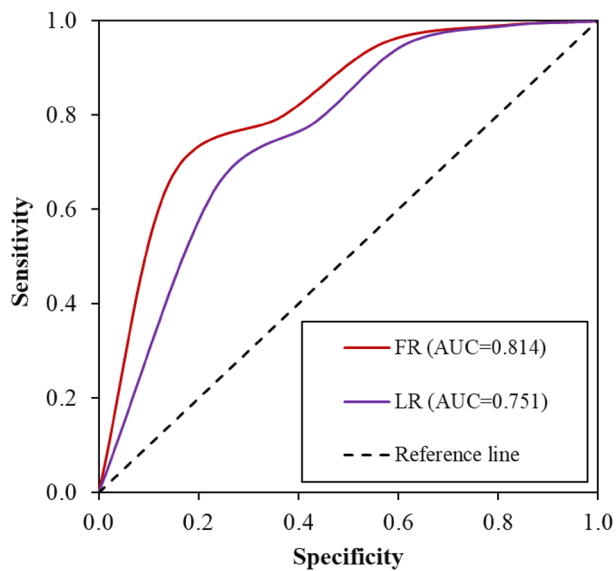
**Table 4** Coefficient of logistic regression for different factors

Parameters	B	S.E.	Wald	df	Sig.	Exp(B)
DD	-1.026	0.510	4.050	1	0.044	0.358
Drainage intensity	-0.093	0.102	0.816	1	0.366	0.912
Drainage texture	-13.652	87.031	0.025	1	0.875	0.000
Infiltration number	-0.021	0.024	0.818	1	0.366	0.979
Stream frequency	1.955	11.639	0.028	1	0.867	7.067
Stream junction frequency	-0.025	0.085	0.086	1	0.770	0.975
Stream power	0.001	0.131	0.000	1	0.994	1.001
Lithology			1.535	1	0.215	1.830
Lithology (1)	0.604	0.488	3.215	1	0.072	1.812
Lithology (2)	0.345	0.654	0.231	1	0.628	2.354
Soil			9.482	4	0.050	
Soil (1)	-0.953	0.470	4.112	1	0.043	0.386
Soil (2)	0.159	0.463	0.118	1	0.731	1.172
Soil (3)	-15.832	2518.445	0.000	1	0.995	0.000
Soil (4)	0.486	0.264	3.406	1	0.065	1.626
Soil (5)	0.398	0.345	2.564	1	0.111	0.234
RR	0.005	0.006	0.688	1	0.407	1.005
Maximum relief	-0.003	0.003	1.084	1	0.298	0.997
Slope	0.048	0.011	20.153	1	0.000	1.049
Aspect			22.099	9	0.009	
Aspect(Flat)	-13.561	40,192.970	0.000	1	1.000	0.000
Aspect(N)	0.529	0.645	0.673	1	0.412	1.697
Aspect(NE)	0.550	0.617	0.796	1	0.372	1.734
Aspect(E)	1.401	0.573	5.966	1	0.015	4.058
Aspect(SE)	1.037	0.590	3.089	1	0.079	2.821
Aspect(S)	0.461	0.602	0.587	1	0.444	1.586
Aspect(SW)	0.884	0.589	2.254	1	0.133	2.420
Aspect(W)	0.190	0.631	0.091	1	0.763	1.210
Aspect(NW)	-0.732	0.719	1.035	1	0.309	0.481
Dissection index	-6.006	3.132	3.676	1	0.055	0.002
Ruggedness number	2.152	1.611	1.784	1	0.182	8.604
Distance from river	-0.002	0.002	2.512	1	0.113	0.998
Relief	-0.001	0.003	0.323	1	0.570	0.999
Rainfall	0.004	0.001	25.270	1	0.000	1.004
Lulc			129.367	4	0.000	
Lulc (vegetation cover)	0.081	0.332	0.060	1	0.806	1.085
Lulc (build-up area)	1.824	0.596	9.363	1	0.002	6.196
Lulc (barren land)	2.387	0.338	49.937	1	0.000	10.883
Lulc (river)	3.071	0.390	62.030	1	0.000	21.553
Lulc (agricultural land)	2.312	0.448	30.451	1	0.004	13.214
TWI	-0.009	0.024	0.152	1	0.697	0.991
Constant	-14.769	3.502	17.790	1	0.000	0.000

Lithology (1) Unclassified, Crystallines (Mainly Gneisses), Lithology (2) Slate, Schists, Quartzite; Soil (1) Gravelly Loamy, Soil (2) Fine Loamy—Coarse Loamy, Soil (3) Gravelly Loamy—Coarse Loamy, Soil (4) Gravelly Loamy—Loamy Skeletal, Soil (5) Coarse Loamy

the total basin area, almost 7–14 sq km area has come under high landslide susceptible zones. Both the maps have shown that high landslide susceptibility zones have been located at the lower portion of the basin and along the river channel where the soil is saturated, open to

sky and surface is unprotected, whereas the upper portion of the basin has very less susceptible as this portion has been covered with healthy forest cover that protects the soil from sliding. The susceptibility maps have been proved with satisfying accuracy of the ROC



**Fig. 10** Validation of landslide susceptible maps using frequency ratio (FR) and logistic regression (LR) models by ROC curve

curve (Fig. 10). The resulting maps have provided the spatial distribution of landslide occurrences, but it cannot forecast the time, degree of landslide occurrences and how often it can occur. Therefore, the government and responsible authorities should take responsibility and steps to mitigate the problem. The government and higher authorities should keep notice so that any kind of development project or market, towns, roads, tourist places, etc., will not grow in the future in the susceptible areas within the basin. These maps will be also useful for the planners and decision-makers to build up policies to restrict and save the destruction from landslides.

**Acknowledgements** The authors would like to express cordial thanks to our respected teachers of the Department of Geography, University of Gour Banga, who have always been mentally, economically and infrastructurally supported ourselves. At last, authors would like to acknowledge all of the agencies and individuals, especially the Irrigation and Waterways Department (Govt. of West Bengal), Survey of India (SOI), Geological Survey of India (GSI) and USGS for obtaining the maps and data required for the study.

## Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflict of interest.

## References

1. Ahmadlou M, Karimi M, Alizadeh S, Shirzadi A, Parvinnejhad D, Shahabi H, Panahi M (2019) Flood susceptibility

assessment using integration of adaptive network-based fuzzy inference system (ANFIS) and biogeography-based optimization (BBO) and BAT algorithms (BA). *Geocarto International* 34(11):1252–1272

2. Akgun A (2012) A comparison of landslide susceptibility maps produced by logistic regression, multi-criteria decision, and likelihood ratio methods: a case study at İzmir, Turkey. *Landslides* 9(1):93–106. <https://doi.org/10.1007/s10346-011-0283-7>
3. Aleotti P, Chowdhury R (1999) Landslide hazard assessment: summary review and new perspectives. *Bull Eng Geol Environ* 58(1):21–44. <https://doi.org/10.1007/s100640050066>
4. Alizadeh M, Alizadeh E, Asadollahpour Kotenaee S, Shahabi H, Beiranvand Pour A, Panahi M, Saro L (2018) Social vulnerability assessment using artificial neural network (ANN) model for earthquake hazard in Tabriz city Iran. *Sustainability* 10(10):3376
5. Avinash KG, Ashamanjari KG (2010) A GIS and frequency ratio based landslide susceptibility mapping: aghnashini river catchment, Uttara Kannada, India. *Int J Geomatics Geosci* 1(3):343
6. Ayalew L, Yamagishi H, Ugawa N (2004) Landslide susceptibility mapping using GIS-based weighted linear combination, the case in Tsugawa area of Agano River, Niigata Prefecture Japan. *Landslides* 1(1):73–81. <https://doi.org/10.1007/s10346-003-0006-9>
7. Azareh A, Rahmati O, Rafiei-Sardooi E, Sankey JB, Lee S, Shahabi H, Ahmad BB (2019) Modelling gully-erosion susceptibility in a semi-arid region, Iran: investigation of applicability of certainty factor and maximum entropy models. *Sci Total Environ* 655:684–696
8. Bai SB, Wang J, Lü GN, Zhou PG, Hou SS, Xu SN (2010) GIS-based logistic regression for landslide susceptibility mapping of the Zhongxian segment in the Three Gorges area China. *Geomorphology* 115(1–2):23–31
9. Basu T, Pal S (2017) Identification of landslide susceptibility zones in Gish River basin, West Bengal, India. *Georisk: Assess Manag Risk Eng Syst Geohazards* 12(1):14–28. <http://dx.doi.org/10.1080/17499518.2017.1343482>
10. Beven KJ, Kirkby MJ (1979) A physically based, variable contributing area model of basin hydrology/Un modèle à base physique de zone d'apport variable de l'hydrologie du bassin versant. *Hydrol Sci Bull* 24(1):43–69. <https://doi.org/10.1080/0262667909491834>
11. Bijukchhen SM, Kayastha P, Dhital MR (2012) A comparative evaluation of heuristic and bivariate statistical modelling for landslide susceptibility mappings in Ghurmi-Dhad Khola, east Nepal. *Arab J Geosci* 6(8):2727–2743. <https://doi.org/10.1007/s12517-012-0569-7>
12. Bui DT, Panahi M, Shahabi H, Singh VP, Shirzadi A, Chapi K, Ahmad BB (2018) Novel hybrid evolutionary algorithms for spatial prediction of floods. *Sci Rep* 8(1):15364
13. Bui DT, Tuan TA, Klempe H, Pradhan B, Revhaug I (2016) Spatial prediction models for shallow landslide hazards: a comparative assessment of the efficacy of support vector machines, artificial neural networks, kernel logistic regression, and logistic model tree. *Landslides* 13(2):361–378
14. Burns WM, Mickelson KA, Saint-Pierre CE (2013) Landslide inventory maps. Portland: Oregon Department of Geology and Mineral Industries. Interpretive map 54–55, 2 pls., scale 1:8,000
15. Chang KT, Hwang JT, Liu JK, Wang EH, Wang CI (2011) Apply two hybrid methods on the rainfall-induced landslides interpretation. In: 2011 19th international conference on geoinformatics, IEEE, pp 1–5
16. Chapi K, Singh VP, Shirzadi A, Shahabi H, Bui DT, Pham BT, Khosravi K (2017) A novel hybrid artificial intelligence approach

- for flood susceptibility assessment. *Environ Model Softw* 95:229–245
17. Chen W, Hong H, Li S, Shahabi H, Wang Y, Wang X, Ahmad BB (2019a) Flood susceptibility modelling using novel hybrid approach of reduced-error pruning trees with bagging and random subspace ensembles. *J Hydrol*
  18. Chen W, Panahi M, Tsangaratos P, Shahabi H, Ilia I, Panahi S, Ahmad BB (2019) Applying population-based evolutionary algorithms and a neuro-fuzzy system for modeling landslide susceptibility. *CATENA* 172:212–231
  19. Chen W, Peng J, Hong H, Shahabi H, Pradhan B, Liu J, Duan Z (2018) Landslide susceptibility modelling using GIS-based machine learning techniques for Chongren County, Jiangxi Province, China. *Sci Total Environ* 626:1121–1135
  20. Chen W, Shahabi H, Shirzadi A, Hong H, Akgun A, Tian Y, Li S (2018b) Novel hybrid artificial intelligence approach of bivariate statistical-methods-based kernel logistic regression classifier for landslide susceptibility modeling. *Bull Eng Geol Environ*:1–23
  21. Chen W, Shahabi H, Shirzadi A, Li T, Guo C, Hong H, Xi M (2018) A novel ensemble approach of bivariate statistical-based logistic model tree classifier for landslide susceptibility assessment. *Geocarto Int* 33(12):1398–1420
  22. Chen W, Shirzadi A, Shahabi H, Ahmad BB, Zhang S, Hong H, Zhang N (2017) A novel hybrid artificial intelligence approach based on the rotation forest ensemble and naïve Bayes tree classifiers for a landslide susceptibility assessment in Langao County, China. *Geomatics, Nat Hazards Risk* 8(2):1955–1977
  23. Chen W, Xie X, Peng J, Shahabi H, Hong H, Bui DT, Zhu AX (2018) GIS-based landslide susceptibility evaluation using a novel hybrid integration approach of bivariate statistical based random forest method. *CATENA* 164:135–149
  24. Chen W, Zhang S, Li R, Shahabi H (2018) Performance evaluation of the GIS-based data mining techniques of best-first decision tree, random forest, and naïve Bayes tree for landslide susceptibility modeling. *Sci Total Environ* 644:1006–1018
  25. Chen W, Zhao X, Shahabi H, Shirzadi A, Khosravi K, Chai H, Wang, X (2019c) Spatial prediction of landslide susceptibility by combining evidential belief function, logistic regression and logistic model tree. *Geocarto Int* (just-accepted) 1–25
  26. Chung CF, Fabbri AG (2003) Validation of spatial prediction models for landslide hazard mapping. *Nat Hazards* 30(3):451–472. <https://doi.org/10.1023/b:nhaz.0000007172.62651.2b>
  27. Dai F, Lee C (2002) Landslide characteristics and slope instability modeling using GIS, Lantau Island Hong Kong. *Geomorphology* 42(3–4):213–228. [https://doi.org/10.1016/s0169-555x\(01\)00087-3](https://doi.org/10.1016/s0169-555x(01)00087-3)
  28. Das I, Sahoo S, Westen CV, Stein A, Hack R (2010) Landslide susceptibility assessment using logistic regression and its comparison with a rock mass classification system, along a road section in the northern Himalayas (India). *Geomorphology* 114(4):627–637. <https://doi.org/10.1016/j.geomorph.2009.09.023>
  29. Demir G, Aytikin M, Akgun A (2014) Landslide susceptibility mapping by frequency ratio and logistic regression methods: an example from Niksar-Resadiye (Tokat, Turkey). *Arab J Geosci* 8(3):1801–1812. <https://doi.org/10.1007/s12517-014-1332-z>
  30. Dou J, Paudel U, Oguchi T, Uchiyama S, Hayakawa YS (2015) Shallow and deep-seated landslide differentiation using support vector machines: a case study of the chuetsu area, Japan. *Terr Atmos Ocean Sci* 26(2)
  31. Dou J, Yamagishi H, Zhu Z, Yunus AP, Chen CW (2018) TXT-tool 1.081-6.1 A comparative study of the binary logistic regression (BLR) and artificial neural network (ANN) models for GIS-based spatial predicting landslides at a regional scale. In: *Landslide dynamics: ISDR-ICL landslide interactive teaching tools*. Springer, Cham pp 139–151
  32. Eker R, Aydin A (2014) Assessment of forest road conditions in terms of landslide susceptibility: a case study in Yığılca Forest Directorate (Turkey). *Turk J Agric For* 38:281–290. <https://doi.org/10.3906/tar-1303-12>
  33. Faniran A (1968) The index of drainage intensity-A provisional new drainage factor. *Aust J Sci* 31:328–330
  34. Gayen A, Saha S (2018) Deforestation probable area predicted by logistic regression in Pathro river basin: a tributary of Ajay River. *Spat Inf Res* 26(1):1–9. <https://doi.org/10.1007/s41324-017-0151-1>
  35. Gokceoglu C, Sonmez H, Nefeslioglu HA, Duman TY, Can T (2005) The 17 March 2005 Kuzulu landslide (Sivas, Turkey) and landslide-susceptibility map of its near vicinity. *Eng Geol* 81(1):65–83
  36. Gómez 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. <https://doi.org/10.1016/j.enggeo.2004.10.004>
  37. Guzzetti F, Reichenbach P, Ardizzone F, Cardinali M, Galli M (2006) Estimating the quality of landslide susceptibility models. *Geomorphology* 81(1–2):166–184
  38. He Q, Shahabi H, Shirzadi A, Li S, Chen W, Wang N, Wang X (2019) Landslide spatial modelling using novel bivariate statistical based Naïve Bayes, RBF Classifier, and RBF Network machine learning algorithms. *Sci Total Environ* 663:1–15
  39. Hong H, Liu J, Zhu AX, Shahabi H, Pham BT, Chen W, Bui DT (2017) A novel hybrid integration model using support vector machines and random subspace for weather-triggered landslide susceptibility assessment in the Wuning area (China). *Environ Earth Sci* 76(19):652
  40. Hong H, Panahi M, Shirzadi A, Ma T, Liu J, Zhu AX, Kazakis N (2018) Flood susceptibility assessment in Hengfeng area coupling adaptive neuro-fuzzy inference system with genetic algorithm and differential evolution. *Sci Total Environ* 621:1124–1141
  41. Horton RE (1932) Drainage-basin characteristics. *Trans Am Geophys Union* 13(1):350. <https://doi.org/10.1029/tr013i001p00350>
  42. Horton RE (1945) Erosional development of streams and their drainage basins; hydrophysical approach to quantitative morphology. *Geol Soc Am Bull* 56(3):275. [http://dx.doi.org/10.1130/0016-7606\(1945\)56\[275:edosat\]2.0.co;2](http://dx.doi.org/10.1130/0016-7606(1945)56[275:edosat]2.0.co;2)
  43. Intrawichian N, Dasananda S (2011) Frequency ratio model based landslide susceptibility mapping in lower Mae Chaem watershed Northern Thailand. *Environ Earth Sci* 64(8):2271–2285. <https://doi.org/10.1007/s12665-011-1055-3>
  44. Jaafari A, Panahi M, Pham BT, Shahabi H, Bui DT, Rezaie F, Lee S (2019) Meta optimization of an adaptive neuro-fuzzy inference system with grey wolf optimizer and biogeography-based optimization algorithms for spatial prediction of landslide susceptibility. *CATENA* 175:430–445
  45. Jaafari A, Zenner EK, Panah M, Shahabi H (2019) Hybrid artificial intelligence models based on a neuro-fuzzy system and metaheuristic optimization algorithms for spatial prediction of wildfire probability. *Agric For Meteorol* 266:198–207
  46. Jadda M, Shafri HZ, Mansor SB, Sharifikia M, Pirasteh S (2009) Landslide susceptibility evaluation and factor effect analysis using probabilistic-frequency ratio model. *Eur J Sci Res* 33(4):654–668
  47. Kavzoglu T, Colkesen I, Sahin, EK (2019) Machine learning techniques in landslide susceptibility mapping: a survey and a case study. In: *Landslides: theory, practice and modelling*. Springer, Cham pp 283–301
  48. Kavzoglu T, Sahin EK, Colkesen I (2014) Landslide susceptibility mapping using GIS-based multi-criteria decision analysis, support vector machines, and logistic regression. *Landslides* 11(3):425–439. <https://doi.org/10.1007/s10346-013-0391-7>

49. Khosravi K, Pham BT, Chapi K, Shirzadi A, Shahabi H, Revhaug I, Bui DT (2018) A comparative assessment of decision trees algorithms for flash flood susceptibility modeling at Haraz watershed, northern Iran. *Sci Total Environ* 627:744–755
50. Khosravi K, Shahabi H, Pham BT, Adamowski J, Shirzadi A, Pradhan B, Hong H (2019) A comparative assessment of flood susceptibility modeling using multi-criteria decision-making analysis and machine learning methods. *J Hydrol* 573:311–323
51. Kayastha P, Dhital MR, Smedt FD (2013) Evaluation and comparison of GIS based landslide susceptibility mapping procedures in Kulekhani watershed Nepal. *J Geol Soc India* 81(2):219–231. <https://doi.org/10.1007/s12594-013-0025-7>
52. Lee CF, Li J, Xu ZW, Dai FC (2001) Assessment of landslide susceptibility on the natural terrain of Lantau Island Hong Kong. *Environ Geol* 40(3):381–391. <https://doi.org/10.1007/s002540000163>
53. Lee S, Min K (2001) Statistical analysis of landslide susceptibility at Yongin Korea. *Environ Geol* 40(9):1095–1113. <https://doi.org/10.1007/s002540100310>
54. Lee S, Pradhan B (2007) Landslide hazard mapping at Selangor, Malaysia using frequency ratio and logistic regression models. *Landslides* 4(1):33–41. <https://doi.org/10.1007/s10346-006-0047-y>
55. Lee S, Talib JA (2005) Probabilistic landslide susceptibility and factor effect analysis. *Environ Geol* 47(7):982–990. <https://doi.org/10.1007/s00254-005-1228-z>
56. Mangeney A (2011) Geomorphology: landslide boost from entrainment. *Nat Geosci* 4(2):77
57. Melchiorre C, Matteucci M, Azzoni A, Zanchi A (2008) Artificial neural networks and cluster analysis in landslide susceptibility zonation. *Geomorphology* 94(3–4):379–400. <https://doi.org/10.1016/j.geomorph.2006.10.035>
58. Menard S (2001) Applied logistic regression analysis, vol 106, 2nd edn. Sage Publication Thousand Oaks, California
59. Miraki S, Zanganeh SH, Chapi K, Singh VP, Shirzadi A, Shahabi H, Pham BT (2019) Mapping groundwater potential using a novel hybrid intelligence approach. *Water Resour Manage* 33(1):281–302
60. Mondal S, Maiti R (2013) Integrating the analytical hierarchy process (AHP) and the frequency ratio (FR) model in landslide susceptibility mapping of Shiv-khola watershed, Darjeeling Himalaya. *Int J Disaster Risk Sci* 4(4):200–212. <https://doi.org/10.1007/s13753-013-0021-y>
61. Moore ID, Grayson RB, Ladson AR (1991) Digital terrain modelling: a review of hydrological, geomorphological, and biological applications. *Hydrol Process* 5(1):3–30. <https://doi.org/10.1002/hyp.3360050103>
62. Nandi A, Shakoor A (2010) A GIS-based landslide susceptibility evaluation using bivariate and multivariate statistical analyses. *Eng Geol* 110(1–2):11–20. <https://doi.org/10.1016/j.enggeo.2009.10.001>
63. Nguyen VV, Pham BT, Vu BT, Prakash I, Jha S, Shahabi H, Tien Bui D (2019) Hybrid machine learning approaches for landslide susceptibility modeling. *Forests* 10(2):157
64. Nir D (1957) The ratio of relative and absolute altitudes of Mt. Carmel: a contribution to the problem of relief analysis and relief classification. *Geogr Rev* 47(4):564. <http://dx.doi.org/10.2307/211866>
65. Onda Y (1993) Underlying rock type controls of hydrological processes and shallow landslide occurrence. *Sedim Probl Strateg Monit Predict Control* 217:47–55
66. Pal S, Talukdar S (2018) Application of frequency ratio and logistic regression models for assessing physical wetland vulnerability in Punarbhaba river basin of Indo-Bangladesh. *Human and Ecol Risk Assess Int J* 24(5):1291–1311. <https://doi.org/10.1080/10807039.2017.1411781>
67. Pham BT, Bui DT, Prakash I (2018) Bagging based support vector machines for spatial prediction of landslides. *Environ Earth Sci* 77(4):146
68. Pham BT, Prakash I, Dou J, Singh SK, Trinh PT, Tran HT, Bui DT (2019) A novel hybrid approach of landslide susceptibility modelling using rotation forest ensemble and different base classifiers. *Geocarto Int*:1–25
69. Pham BT, Prakash I, Khosravi K, Chapi K, Trinh PT, Ngo TQ, Bui DT (2018b) A comparison of Support Vector Machines and Bayesian algorithms for landslide susceptibility modelling. *Geocarto Int*:1–23
70. Pham BT, Prakash I, Singh SK, Shirzadi A, Shahabi H, Bui DT (2019) Landslide susceptibility modeling using Reduced Error Pruning Trees and different ensemble techniques: hybrid machine learning approaches. *CATENA* 175:203–218
71. Pourghasemi HR, Pradhan B, Gokceoglu C (2012) Application of fuzzy logic and analytical hierarchy process (AHP) to landslide susceptibility mapping at Haraz watershed Iran. *Natural Hazards* 63(2):965–996. <https://doi.org/10.1007/s11069-012-0217-2>
72. Pradhan B, Buchroithner MF (2010) Comparison and validation of landslide susceptibility maps using an artificial neural network model for three test areas in Malaysia. *Environ Eng Geosci* 16(2):107–126. <https://doi.org/10.2113/gsegeosci.16.2.107>
73. Pradhan B, Lee S (2010) Landslide susceptibility assessment and factor effect analysis: backpropagation artificial neural networks and their comparison with frequency ratio and bivariate logistic regression modelling. *Environ Model Softw* 25(6):747–759. <https://doi.org/10.1016/j.envsoft.2009.10.016>
74. Pradhan B, Mansor S, Pirasteh S, Buchroithner MF (2011) Landslide hazard and risk analyses at a landslide prone catchment area using statistical based geospatial model. *Int J Remote Sens* 32(14):4075–4087
75. Pradhan B, Youssef AM (2009) Manifestation of remote sensing data and GIS on landslide hazard analysis using spatial-based statistical models. *Arab J Geosci* 3(3):319–326. <https://doi.org/10.1007/s12517-009-0089-2>
76. Rahmati O, Naghibi SA, Shahabi H, Bui DT, Pradhan B, Azareh A, Melesse AM (2018) Groundwater spring potential modelling: comprising the capability and robustness of three different modeling approaches. *J Hydrol* 565:248–261
77. Rahmati O, Samadi M, Shahabi H, Azareh A, Rafiei-Sardooi E, Ailou H, Shirzadi A (2019) SWPT: An automated GIS-based tool for prioritization of sub-watersheds based on morphometric and topo-hydrological factors. *Geosci Front*
78. Rai PK, Mohan K, Kumra VK (2014) Landslide hazard and its mapping using remote sensing and GIS. *J Sci Res* 58:1–13
79. Rodriguez-Galiano V, Sanchez-Castillo M, Chica-Olmo M, Chica-Rivas M (2015) Machine learning predictive models for mineral prospectivity: an evaluation of neural networks, random forest, regression trees and support vector machines. *Ore Geol Rev* 71:804–818
80. Roodposhti MS, Safarrad T, Shahabi H (2017) Drought sensitivity mapping using two one-class support vector machine algorithms. *Atmos Res* 193:73–82
81. Rasyid AR, Bhandary NP, Yatabe R (2016) Performance of frequency ratio and logistic regression model in creating GIS based landslides susceptibility map at Lompobattang Mountain, Indonesia. *Geoenviro Disasters*, 3(1). <http://dx.doi.org/oi:10.1186/s40677-016-0053-x>
82. Shafizadeh-Moghadam H, Valavi R, Shahabi H, Chapi K, Shirzadi A (2018) Novel forecasting approaches using combination of machine learning and statistical models for flood susceptibility mapping. *J Environ Manage* 217:1–11

83. Shahabi H, Khezri S, Ahmad BB, Hashim M (2014) Landslide susceptibility mapping at central Zab basin, Iran: a comparison between analytical hierarchy process, frequency ratio and logistic regression models. *CATENA* 115:55–70. <https://doi.org/10.1016/j.catena.2013.11.014>
84. Sharma LP, Patel N, Debnath P, Ghose MK (2012) Assessing landslide vulnerability from soil characteristics—a GIS-based analysis. *Arab J Geosci* 5(4):789–796. <https://doi.org/10.1007/s12517-010-0272-5>
85. Shirzadi A, Bui DT, Pham BT, Solaimani K, Chapi K, Kavian A, Revhaug I (2017) Shallow landslide susceptibility assessment using a novel hybrid intelligence approach. *Environ Earth Sci* 76(2):60
86. Shirzadi A, Shahabi H, Chapi K, Bui DT, Pham BT, Shahedi K, Ahmad BB (2017) A comparative study between popular statistical and machine learning methods for simulating volume of landslides. *CATENA* 157:213–226
87. Shirzadi A, Soliamani K, Habibnejhad M, Kavian A, Chapi K, Shahabi H, Ahmad A (2018) Novel GIS based machine learning algorithms for shallow landslide susceptibility mapping. *Sensors* 18(11):3777
88. Smith G (1935) The relative relief of Ohio. *Geogr Rev* 25(2):272. <https://doi.org/10.2307/209602>
89. Sørensen R, Zinko U, Seibert J (2006) on the calculation of the topographic wetness index: evaluation of different methods based on field observations. *Hydrol Earth Syst Sci Dis* 10(1):101–112
90. Strahler AN (1957) Quantitative analysis of watershed geomorphology. *Trans Am Geophys Union* 38(6):913. <https://doi.org/10.1029/tr038i006p00913>
91. Taheri K, Shahabi H, Chapi K, Shirzadi A, Gutiérrez F, Khosravi K (2019) Sinkhole susceptibility mapping: a comparison between Bayes-based machine learning algorithms. *Land Degrad Dev* 30(7):730–745
92. The Indian Express. July 2, 2015. <http://indianexpress.com/article/india/west-bengal/landslides-claim-18-lives-in-darjeeling/>
93. Tien Bui D, Khosravi K, Li S, Shahabi H, Panahi M, Singh V, Bin Ahmad B (2018) New hybrids of anfis with several optimization algorithms for flood susceptibility modeling. *Water* 10(9):1210
94. Tien Bui D, Shahabi H, Omidvar E, Shirzadi A, Geertsema M, Clague JJ, Barati Z (2019) Shallow landslide prediction using a novel hybrid functional machine learning algorithm. *Remote Sens* 11(8):931
95. Tien Bui D, Shahabi H, Shirzadi A, Chapi K, Alizadeh M, Chen W, Tian Y (2018) Landslide detection and susceptibility mapping by airsar data using support vector machine and index of entropy models in cameron highlands, malaysia. *Remote Sens* 10(10):1527
96. Tien Bui D, Shahabi H, Shirzadi A, Chapi K, Pradhan B, Chen W, Saro L (2018) Land subsidence susceptibility mapping in south korea using machine learning algorithms. *Sensors* 18(8):2464
97. Tien Bui D, Shahabi H, Shirzadi A, Kamran Chapi K, Hoang ND, Pham B, Saro L et al (2019) Erratum: dieu, TB A Novel Integrated Approach of Relevance Vector Machine Optimized by Imperialist Competitive Algorithm for Spatial Modeling of Shallow Landslides. *Remote Sens*. 2018, 10, 1538. *Remote Sens* 11(1):57
98. Tien Bui D, Shirzadi A, Shahabi H, Chapi K, Omidav E, Pham BT, Talebpour Asl D, Khaledian H, Pradhan B, Panahi M (2019) A novel ensemble artificial intelligence approach for gully erosion mapping in a semi-arid watershed (Iran). *Sensors* 19:2444
99. Varnes DJ (1978) Slope movement types and processes. *Spe Rep* 176:11–33
100. Varnes DJ (1981) Slope-stability problems of circum-Pacific region as related to mineral and energy resources
101. Wilson JP, Gallant JC (eds) (2000) *Terrain analysis: principles and applications*. Wiley, Hoboken
102. Xie M, Esaki T, Zhou G, Mitani Y (2003) Geographic information systems-based three-dimensional critical slope stability analysis and landslide hazard assessment. *J Geotech Geoenviron Eng* 129(12):1109–1118

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.