Selection of weightages for causative factors used in preparation of landslide susceptibility zonation (LSZ)

Sharad Kumar Gupta a, Dericks P. Shukla a and Manoj Thakurb

aSchool of Engineering, Indian Institute of Technology, Mandi, India; bSchool of Basic Sciences, Indian Institute of Technology, Mandi, India

ABSTRACT
Most of the models for landslide susceptibility zonation (LSZ) except machine learning needs manual selection of weights or ratings, which are given by expertise knowledge, leading to subjectivity and specificity. Hence, selection of ratings is very important in the preparation of LSZ maps. Here, seven layers/factors viz. aspect, elevation, geology, slope, soil, distance from stream, distance from thrusts are considered for LSZ mapping in Mandakini River basin, Uttarakhand containing 1,805,636 pixels. The weights were calculated using information value (InfoVal) method. The occurrences of landslides (2009 pixels) until 2008 were considered for training of model. Thus for giving rating to each thematic layer, 7! = 5040 permutations (ratings) are possible. Hence, 5040 LSZ maps were prepared and out of them, the best rating was identified using fisher discriminant analysis (FDA) and binary logistic regression (LR). FDA and LR gave similar ratings and the correlation ($r^2$) between their weights was 0.9226. Thematic layers were then multiplied by the corresponding ratings and added to prepare the final LSZ map. The results were validated on the landslide data until 2011, which were not used for training. Out of 223 occurrences of landslides, 39% (87) falls in high susceptible zone followed by 35% (79) in very high susceptible zone according to FDA and similar for LR. The accuracy of the prediction was assessed by Heidke-Skill-Score, which gave score of 0.89 to FDA and 0.90 to LR for 0.5 as threshold of landslide susceptibility index.

KEYWORDS
Fisher discriminant analysis (FDA); Heidke-Skill-Score (HSS); information value (InfoVal); landslide susceptibility zonation (LSZ); logistic regression (LR)

1. Introduction
Landslides are among the most devastating natural disasters. Worldwide, they cause billions of dollars in property damage and thousands of deaths every year (Hong et al. 2007). Nearly 15% of the Indian land area is prone to landslide hazard (GSI, http://www.portal.gsi.gov.in). Not only landslides, but also other natural hazards such as cloud burst, flash floods, earthquakes, etc. are very frequent in this tectonically active Himalayan belt. Thus, the preparation of zonation maps in this region plays a critical role on which government can take necessary actions for proper and phased development in this region. Government of India prepared landslide zonation for major pilgrimage routes in the Northwestern Himalaya as mentioned on BHUVAN portal (http://bhuvan-noeda.nrsc.gov.in/disaster/disaster/disaster.php?id=landslide), an Indian geo-portal of ISRO (Indian Space Research Organisation). Majority of these landslides are mapped using AHP (analytic hierarchy process) method along roads and near settlements. Most of these roads are in the valley region, thus the proximity of rivers (such as Alakmnanda, Bhagirathi, Mandakini, etc.) is quite prominent.
In general, there are four approaches for preparation of such zonation maps. They are deterministic approach, heuristic approach, probabilistic and machine learning approach. Deterministic or physically based models are based on physical laws of conservation of mass, energy or momentum (Terlien et al. 1995). The parameters used in these models can be determined in the field or in the laboratory. Most deterministic models are site-specific and do not take into account the spatial distribution of the input parameters. In heuristic approach, the expert opinion of the geomorphologists is used for assignment of weightages to various factors in zonation, e.g. analytic hierarchy process, etc. (Soeters and Westen 1996). The probabilistic approach is based on relationship observed between causative factors and landslide distribution in an area within a probabilistic framework, e.g. weight of evidence method under Bayesian probability model, certainty factor method under favourability mapping model, etc. (Kanungo et al. 2009). Since landslides occur due to the complex topographic and geological conditions, where identification of such relationships is quite challenging; flexible nonlinear methods such as machine learning techniques are being used these days. Machine learning methods learn the spatial distribution of landslides from the data and thus models the unknown dependency between potential landslide causative factors and the presence or absence of landslide (Michelutti et al. 2014), e.g. artificial neural network (ANN), support vector machine, etc. (Kumar et al. 2017).

Among these approaches, various methods such as frequency ratio, discriminant analysis, direct mapping, regression analysis, fuzzy logic, ANN, neuro-fuzzy method and other techniques are quite popular and can be applied for preparation of zonation maps. Specifically for the preparation of landslide susceptibility zonation maps, various authors have used various approaches using these techniques such as statistical technique (Akgun and Türk 2010; Shahabi and Hashim 2015), logistic regression (LR) (Lee and Pradhan 2007; Mathew et al. 2009; Nourani et al. 2014), probabilistic models (Lee and Pradhan 2006; Pourghasemi et al. 2013) and ANN (Arora et al. 2004; Chauhan et al. 2010; Kanungo et al. 2011; Bui et al. 2016). Landslides usually occur as soil resistance deteriorates in the presence of the acting stresses developed due to a number of reasons such as presence of tectonically active thrusts, high slopes, increased soil moisture content, change in land use, etc.; thus failing the slopes. These conditions are site-specific and vary from one geo-environmental setting to other. These different methods apply different rating system as well as the weights, which are highly dependent on study area, and the controlling factors (Shukla et al. 2016; Kumar et al. 2017). Therefore, these weights and ratings play a vital role in the preparation of susceptibility maps using any method. Many of the works for LSZ mapping in Himalayan regions employ rating system based on the expertise experience (Rautela and Thakur 1999; Saha et al. 2005; Pandey et al. 2008; Devkota et al. 2013; Ahmed 2015), which tend to be dependent on the knowledge of expert or mapping geomorphologists and less on the learning capabilities from the data provided. Even though machine-learning approach has been applied in Himalayan region (Pham et al. 2016a, 2016b; Kumar et al. 2017) as well as elsewhere (Pradhan 2013; Bui et al. 2016; Hong et al. 2016) but still in many practices, the ranking is given based on the earlier experiences. This leads to biasness in the results and the output tends to give the results based on the initial weight given by the author. When such models, where user gives ranking based on earlier experiences, are used for preparation of LSZ map (Rawat et al. 2015; Ghosh et al. 2017; Ramesh et al. 2017), the choice of rating needs to be analysed. Depending on the number of layers being used in the analysis, the combination of these ratings will vary in factorial terms. For example, if there are seven layers being used as input then their rating could be $7! = 5040$ possibilities. Therefore, the choice of best rating needs to be assessed when either expert rating system is not available or not applicable. To cater these uncertainties in this research, frequency ratio has been applied for preparation of LSZ map using seven factors namely elevation, slope, aspect, geology/lithology, buffer of thrusts/ faults, buffer of streams and soil. There are 5040 outputs of LSZ and the choice of best combination for LSZ model is extracted using fisher discriminant analysis (FDA) and LR for assigning ranks to all the causative factors of landslide occurrences.
2. Materials

2.1. Study area

The LSZ has been prepared for Mandakini River basin of Garhwal Himalaya in Uttarakhand. The Mandakini River basin shown in Figure 1 covers an area of about 1625 sq. km and is situated between 30°19’00”N to 30°49’00”N latitude and 78°49’00”E to 79°21’13”E longitude falling in Survey of India Toposheet Nos. 53 J and 53 N. Mandakini River originates from Chorabari Glacier and Companion Glacier near Kedarnath and travels 92 km southward till Rudra Prayag where it meets Alaknanda River. On its way from Chorabari Glacier to Rudraprayag, Mandakini River passes from Gaurikund, Sonprayag, Okhimath, Guptkashi, Kund and Rudra Prayag. This region has highly rugged topography, deep gorges, high peaks where higher areas are mostly snow-covered forming the U-shaped wide valleys of glacial landscape. This part of Garhwal Himalaya is prone to landslides and every year many landslides occur in this area. Hence, the landslide susceptibility mapping in this area is of utmost importance.

2.2. Geological details

The Mandakini river basin presents a typical section of Himalayan orogeny where the rocks of Vaikrita, Munsiari and Ramgarh formation are predominantly visible as shown in Figure 2. In the northern portion of the basin, Vaikrita formation is present which is mainly composed of coarse grained mica-garnet-kyanite and sillimanite bearing psammitic metamorphics, forming most of the greater/higher Himalaya in Garhwal (Valdiya and Goel 1983). These rocks form highly rugged topography, deep gorges, high peaks and at the higher altitudes most of the areas are snow-covered forming the U-shaped wide valleys of glacial landscape. This group is intruded by Tertiary pegmatites and granites (Badrinath Granite) leading to creation of migmatites and biotite-rich schists and calc-schists. The mineral assemblages suggest a medium to high grade metamorphism similar to amphibolite facies observed in the rocks of the Vaikrita Group (Valdiya and Goel 1983). The bedding plane has strike orientation of approximately N300° and 40° dip amount, while the joint orientations are approximately N285° with dip amount 90°. South of this is Munsiari formations which are mainly composed of meso- to epi-grade para- and ortho-gneisses, schist, calc silicate, marble, granites and their mylonitic equivalent (Choubey et al. 1999). The orientation of the bedding planes are strike N320° and dip 38° towards north-east. South of Munsiari formation, the Ramgarh group is present, which has predominantly garnetiferous quartzites with schistose gneisses, impure marble and schists. The age range for inner Ramgarh group is 1856–1868 Ma. It is approximately 1000 Ma older than oldest outer sequence Ramgarh group of rocks and thus is the basement rocks of lesser Himalayan sequence (Célerier et al. 2009). The southernmost area has huge succession of massive, coarse-grain and usually very sericitic quartz arenite of Bering Formation (Valdiya and Goel 1983).

Major thrusts present in this area are Vaikrita Thrust or Main Central Thrust (MCT-I), Munsiari/Jutogh Thrust or Main Central Thrust (MCT-II) and Ramgarh/Chail Thrust or Main Central Thrust (MCT-III) (Ray and Srivastava 2010; Shukla et al. 2014). Due to the presence of MCT Thrust zone, these rocks are highly fractured and sheared. This makes this area highly fragile and prone to landslides. The high susceptibility to landslides in the Mandakini River basin is mainly due to complex geological settings, varying slopes and relief, heavy rainfall, along with ever-increasing human interference in the ecosystem. Furthermore, extreme climatic events of cloud burst and flash flood increase the instability of the terrain with example including the Kedarnath disaster (Dubey et al. 2013). Some of the major landslides that have occurred in the past are near Okhimath in 1997, 1998, 2010, 2012, 2013; in Phata-Byung area in 2001, 2005, 2013; in Madhya-Maheshwar area in 1998, 2005, 2013, etc. They are used for preparation of landslide inventory map along with some other landslides. These are dependent on various factors such as geology, structure, land use, old slides, slope, slope aspect and drainage in the area (Kumar et al. 2017).
Figure 1. Location map of the study area showing digital elevation model prepared from ASTER-GDEM displaying major locations in Mandakini River basin. The major Himalayan thrusts passing from Uttarakhand are shown in inset.
2.3. Data used

The base map has been prepared from Landsat satellite image, of October 2008 having 30 m spatial resolution, along with Survey of India Toposheet number 53 N and J at 1:50,000 scale. Band combination of 5, 4, 2, i.e. SWIR (short wave infrared), NIR (near infrared) and green bands of Landsat-5 satellite series was taken for preparation of geological and structural map (Figure 2) of the study area (Sati et al. 1998).

Since band 5 has high contrast for rocky outcrops, hence its combination along with pseudo-NIR combination of band 7, 5, 2 was used for demarcation of lithological units based on published geological maps of this region. ASTER-GDEM (Advance Space borne Thermal Emission and reflection Radiometer, Global Digital Elevation Model) data is used for preparation of elevation and its derived maps, which has an accuracy of ±10 m at 30 m spatial resolution (Aster-GDEM 2009). Various thematic layers such as elevation, slope, aspect, drainage, geology/lithology, buffer of thrusts/faults and

Figure 2. Geological map of the study area showing various formations and structures mainly Main Central Thrust (MCT-I, MCT-II) and Ramgarh thrust (modified after Shukla et al. 2014).
buffer of streams (Figure 3) are prepared using these data-sets in Arc GIS 10.2 software. Soil Map (Figure 3(d)) is prepared from the data provided by NATMO (National Atlas and Thematic Mapping Organization)/NBSS-LUP (National Bureau of Soil Survey and Land Use Planning) to Uttarakhand government for the detail profile of the districts and from various other sources.
Drainage (Figure 3(e)) was buffered at certain intervals. Since the effect of drainage on occurrences of landslides varies with the order of stream and the distance from the main trunk. All the streams from third order onwards were considered in this study and were buffered at 100, 200, 500, 1000, 1500 and 2000 m from the main stream. Higher categorical values were given to the higher order streams. Thus seventh-order stream, which is the main stream, will have more impact on landslides due to the power it has for toe cut erosion than the second or third-order streams. Hence, chances of occurrence of landslides on larger streams are more as compared to smaller ones. Furthermore, most landslides occur near the streams, so chances of occurrences of landslides in 200 m buffer will be more than 2 km buffer distance. Hence, the third-order stream when buffered at 100 m is given a code of 300 while the fifth-order stream buffered at 500 m is given the code of 2500. The code of 14000 was given to seventh-order stream buffered at 2000 m. However, inverse values were given in the code. That means the value of 500 m buffer on fifth order, i.e. 2500 will be taken as 1/2500 and it will be more than 1000 m buffer on same order, i.e. 5000 which will be taken as 1/5000. Similarly, data was prepared for other order streams.

The thrusts present in the study area are MCT-I, MCT-II and Ramgarh thrusts which are buffered at 200, 500, 1000, 2000, 4000, 6000, 8000 and 10,000 m to indicate the effect of thrusts in the surrounding areas. Thus, the 200 m buffer will show higher activity as compared to 10 km buffer; hence, the occurrence of landslide should be high near the thrusts as compared to areas away from the thrust, if only the thrust parameter is considered. These thrust buffers are classified into nine zones based on their proximity from landslide locations and each are given decreasing values from core to outskirts (Figure 3(f)).

Some of the major landslides have occurred in this area (Sati et al. 1998; Rautela and Thakur 1999; Chaudhary et al. 2010) which were taken into consideration for preparing the inventory map. Details of landslides were collected from the field as well as satellite images. These data were appended to the landslide inventory data obtained from Geological Survey of India (GSI) also. So, the landslide inventory was prepared by various sources covering most of the regions within the vicinity of major roads and habitation. This inventory was prepared in two batches, one until 2008 and another until 2011. Some areas have very large landslides while some have smaller landslides. So, more than one pixel represents presence of landslides. Thus, there are 2009 pixels corresponding to landslides, which are spread in the whole study area having 1,80,5636 pixels for inventory until 2008. The inventory data till 2008 was used for preparation of LSZ while landslide inventory data till 2011 was used for validation of the map. While preparing the inventory data, the extreme event case of Kedarnath disaster 2013 (Dobhal et al. 2013; Dubey et al. 2013) was not included. Extreme climatic events of cloud burst and flash flood increase the instability of the terrain with example including the Kedarnath disaster, which created many landslides even in those areas which had no history of landslide occurrences. Such areas if included in the inventory will tend to yield false positive results. Since such events occur once in a blue moon, so incorporating these data will deteriorate the model by over fitting. Hence, these extreme events have not been considered in this study.

3. Methodology

Zonation depends on the availability of data layers. In this work we have used aspect, DEM, geology, slope, soil, stream buffer and thrust buffer as the causative factors. Most of these layers are first-order derivatives, while other second-order derivatives such as topographic wetness index, sediment transport index, curvature, etc. are not used in this study. Inclusion of more factors increases data complexity and size. In this paper we have used two methods, FDA and LR for removing human intervention in weightages assignment. Susceptibility zonation map is just the representative of the outcome of weightage assignment procedure. Other parameters such as precipitation, seismic data, which are time dependent, have not been used here since the objective of the paper deals with landslide susceptibility zonation rather than hazard zonation. Information value method has been used for assigning the weightage of a thematic class inside thematic layers or factors.
3.1. Information value (InfoVal)

This method considers the probability of landslide occurrence within a certain area of each class of a thematic layer. In this method, the combination of the landslide location map with each thematic layer permits us to determine the weight of influence on terrain instability for each parameter class with the following equations \( W_i \) (Saha et al. 2005; Vijith et al. 2009; Sarkar et al. 2013):

\[
W_i = \ln \left( \frac{\text{Densclas}}{\text{Densmap}} \right)
\]

\[
W_i = \ln \left( \frac{\sum_{i=1}^{n} N_{\text{pix}}(S_i)}{\sum_{i=1}^{n} N_{\text{pix}}(N_i)} \right)
\]

where Densclas is the landslide density within the thematic class; \( W_i \) is the weight given to the \( i \)th class of a particular thematic layer (e.g. - Vaikrita or Munsiari in the thematic layer ‘geology’). Densmap is the landslide density within the entire thematic layer; \( N_{\text{pix}}(N_i) \) is the total number of pixels in a certain thematic class. \( N_{\text{pix}}(S_i) \) is the number of landslide pixels in a certain thematic class and \( n \) is the number of classes in a thematic map. Various classes of each thematic layers are assigned the rating values as attribute information in GIS and an attribute map is generated for each thematic layer.

We can calculate Landslide Susceptibility Index (LSI) using the weighted linear combination as given by the expression (3) in the following (Malczewski 2000; Pandey et al. 2008; Michael and Samanta 2016):

\[
\text{LSI} = \sum \text{ranking} \times \text{data layer (attribute)}
\]

LSI can be categorized into five classes yielding zones of different susceptibility (i.e. very high, high, moderate, low and very low). The major problem of LSZ is the assignment of weights to all thematic layers. What weight should be assigned to which layer is the provocative question? One such method of weight assignment called as direct method, in which the mapping geomorphologist, based on their experience and knowledge of the terrain conditions, determines the weights for all thematic layers. This approach is very subjective because geomorphologists are specific to an area. For eliminating subjectivity in weights assignment, for these seven layers, \( 7! = 5040 \) (weightages) are possible; hence, total 5040 LSZ maps are prepared using all the layers. Not all the maps are representative of actual landslide conditions. Hence, selecting best weight out of these 5040 possible weights is of great concern. This problem has been solved by framing this problem as feature selection problem, where the main objective is to reduce the dimensionality of the input data.

3.2. Fisher discriminant analysis (FDA)

FDA is one of the most popular dimensionality reduction methods based on L2-norm (Wang et al. 2014). Similar to other supervised algorithms, FDA seeks an embedding transformation which increases intra class (between classes) scatter and decreases inter class (within class) scatter (Figure 4) (Ji et al. 2013). For binary (two class) problem \((1 = \text{landslide}, 0 = \text{no landslide})\), let \( n \), \( d \)-dimensional input samples, \( x = [x_1, x_2, x_3, \ldots x_n], x_i \in R^d \), is transformed to one dimension using (4) (Duda et al. 2001; Bishop 2006)

\[
y = w'x
\]

where \( y = [y_1, y_2, y_3, \ldots y_n] \) are \( n \), one-dimensional samples divided into subset \( Y_1 \) and \( Y_2 \). Geometrically, if \( ||w|| = 1 \), each \( y_i \) is the projection of the corresponding \( x_i \) onto a line in the direction
of \( w \). Suppose there are \( n_1 \) samples in \( D_1 \subset X \), labelled as 1 (landslide) and \( n_2 \) samples in \( D_2 \subset X \), labelled as 0 (no landslide). The projection of points onto line should be well separated and least overlapped. A measure of separation between the projected points is the difference of the sample means (5) of both classes:

\[
m_i = \frac{1}{n_i} \sum_{x \in D_i} x
\]

where \( m_i \) is the sample mean for both the classes. Sample mean \( \bar{m}_i \) for projected point is defined as (6)

\[
\bar{m}_i = \frac{1}{n_i} \sum_{y \in Y_i} y
\]

Scatter \( \bar{s}_i^2 \) for the projected samples is defined as (7)

\[
\bar{s}_i^2 = \sum_{y \in Y_i} (y - \bar{m}_i)^2
\]

Fisher analysis uses the linear function \( w^T x \) for which criterion is defined as the ratio of the between-class scatter to the within-class scatter. \( w \) is obtained by maximizing the criterion function (8) (Prasad and Bruce 2008). Criterion function or Fisher’s ratio is given by \( J(w) \)

\[
J(w) = \frac{(\bar{m}_1 - \bar{m}_2)^2}{\bar{s}_1^2 + \bar{s}_2^2} \text{ or } \frac{w^T S_B w}{w^T S_w w}
\]

where \( S_B \) (9) is between class scatter matrix and \( S_w \) (10) is within class scatter matrix. Both of these matrices are symmetric and positive semidefinite. These can be calculated as

\[
S_B = (m_1 - m_2)(m_1 - m_2)^T
\]

\[
S_w = \sum_{x \in D_1} (x - m_1)(x - m_1)^T + \sum_{x \in D_2} (x - m_2)(x - m_2)^T
\]

Using Eigen analysis, \( J(w) \) is maximum when \( w = S_w^{-1}(m_1 - m_2) \). Hence, \( w \) gives the direction as well as weights of all the thematic layers after projection.
In FDA, all the factors are projected in one dimension corresponding to landslide occurrences. For projecting $n$-dimensional vector into one dimension, we require multiplication factor ($w$). That, multiplication factor is nothing but the weightage vector. These weightages have been used for giving ranking to all the thematic layers.

### 3.3. Logistic Regression

LR models the data with binary responses, i.e. it predicts the presence or absence of an outcome based on the values of a set of predictor variables (Shukla et al. 2016). The dependent variable can be binary or multinomial, whereas the independent variable can be categorical, dichotomous or interval (Devkota et al. 2013). Qualitative data of landslide occurrences has binary responses ($1$ = landslide and $0$ = no landslide). Instead of predicting these two values, we have modelled the probabilities associated with these values.

Let $\pi(x)$ be the probability of success of observed values for predictor $x$. The odds of success is the ratio of the probability of success to the probability of failure. Odds vary between zero and infinity as the probability varies between zero and one. Logit is the log of odds of success and represented as (11) (Chatterjee and Simonoff 2013)

$$l(x) = \log \left( \frac{\pi(x)}{1 - \pi(x)} \right)$$

LR hypothesizes that the logit is linearly related to the predictors (12) and (13); that is,

$$l(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n = \log \left( \frac{\pi(x)}{1 - \pi(x)} \right)$$

$$\pi(x) = \frac{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n}}{1 + e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n}}$$

where $\beta_1$, $\beta_2$, $\ldots$, $\beta_n$ are LR coefficients and $\beta_0 = \text{intercept}$, $x_1$, $x_2$, $\ldots$, $x_n$ are independent variables. This equation is called LR function. Coefficients $\beta_1$, $\beta_2$, $\ldots$, $\beta_n$ have been used for giving ranking to all the thematic layers.

In LR, we find the probability of occurrence of landslide because of all causative factors. We solve LR equation for all factors ($x_1$, $x_2$, $\ldots$, $x_n$ on right hand side of (13) which are considered as landslide causative factors such as aspect, DEM, geology, etc.) and probability of landslide occurrences ($\pi(x)$). LR coefficients are nothing but the weightage for all the causative factors of landslide. These weightages are used for ranking all the factors.

### 3.4. Heidke-Skill-Score

For assessing the accuracy of generated landslide susceptible maps with respect to landslide occurrences, Heidke-Skill-Score (HSS) has been used. It is the measure of skill of prediction (NDFD Verification Help 2017). It gives a score between 0 and 1. HSS can be defined as given in the following (14):

$$\text{Score} = \frac{NC - E}{T - E}$$

where NC is the number of correct predictions, i.e. number of times the prediction and observation match, $E$ is the number of predictions expected to verify based on chance, i.e. incorrect predictions, $T$ is total number of observations.
4. Experimental data preparation and processing

All the computations are performed on computer system utilizing 3.6 GHz Intel Core i7 processor with 16-gigabytes of primary memory (RAM). The required data are generated using ArcGIS 10.5 software at 30-m spatial resolution. Vector layers such as stream buffer, thrusts, geology and soil were converted to raster format. Digital elevation model, slope and aspect were in raster formats. All the layers are of size 1869 $\times$ 1734 pixels. The gaps in the data were filled by using nearest neighbour interpolation technique. All these raster data-sets were converted to ASCII format for processing in MATLAB. These layers were converted to column vector and appended to form input data of dimension, 1,805,636 $\times$ 7.

Landslide inventory used for preparation of landslide susceptibility maps in this study contains all the landslide occurred until 2008. The study area contains 1,805,636 pixels, out of which 2009 pixels are associated with landslide occurrences. The pixels representing the landslides are mere 0.11% of the whole study area indicating the high amount of non-landslide pixels. Landslides that occurred until 2011, excluding the landslides that were used for training the model, have been used for validation of the results. FDA and LR were applied on the data and weights were obtained for all layers. For finding the similarity between weight assignments by both methods, correlation analysis was performed. These weights were used to select one out of 5040 permutations as discussed earlier.

5. Results and discussions

Calculations of the weights given to each factor have a crucial role in the production of landslide susceptibility maps. For eliminating subjectivity in the weights assignment, we have used FDA and LR methods. Weights obtained by both these are given in Table 1. Since there are seven thematic layers used in the study, all the layers have been assigned 1–7 rankings based on weights obtained. Higher ranking is assigned for higher weights. Higher ranking shows more significance in determining susceptible zones. LSZ map generated using ranking obtained from FDA and LR is shown in Figures 5 and 6, respectively.

Both algorithms have given similar ranking to DEM-1, Soil-6, Stream-7 and Thrust-5 (numbers indicate the ranks obtained for these layers). Both FDA and LR methods have given highest ranking has been given to streams which is clearly justified by the location of most landslides in the study area, which have occurred near to river or streams. As we know that some particular type of soils are more susceptible to landslides hence second most significant ranking by the algorithms has been given to soil.

Correlation analysis between weights obtained by both methods yielded correlation value ($r$) as 0.9605 or $r^2$ as 0.9226, which shows higher correlation between weights obtained. This shows that any of the two methods can be used for ranking. These rankings are then multiplied by the corresponding thematic layers to yield LSI. The index values thus obtained are normalized between zero and one. This index has been classified using natural break of ArcGIS 10.5, where the intervals of (0–0.33) are assigned to very low susceptible, (0.33–0.44) for low susceptible, (0.44–0.71) for high susceptible and (0.71–1) for very high susceptible zones.

Table 1. Weightages and ranking assigned to various thematic layers by FDA and LR. Higher ranking corresponds to higher weightages.

| Thematic layers | Weights by FDA | Ranking by FDA | Weights by logistic regression | Ranking by logistic regression |
|-----------------|----------------|----------------|-------------------------------|-------------------------------|
| Aspect          | 0.4978         | 2              | 1.2709                        | 3                             |
| DEM             | 0.3173         | 1              | 0.4790                        | 1                             |
| Geology         | 0.8780         | 4              | 1.0858                        | 2                             |
| Slope           | 0.5911         | 3              | 1.2818                        | 4                             |
| Soil            | 2.1132         | 6              | 3.4464                        | 6                             |
| Stream          | 2.5829         | 7              | 3.5640                        | 7                             |
| Thrust          | 1.7770         | 5              | 3.3257                        | 5                             |
For validation of LSZ maps, landslide occurred until 2011, excluding those used for training the model, has been used. There are 223 landslides selected randomly for validation in study area. Number of landslides occurring in a particular zone using FDA and LR are given in Table 2.

Table 2 shows that area falling under very-high and high susceptibility zone is very less as compared to other three zones. However, number of landslide occurrences in these two zones (very-high and high susceptibility) are highest. Low and very low susceptible zone have lesser number of landslides, i.e. 22 out of 223. The very high and high susceptible zones prepared using LR have less number of landslides as compared to those falling in zones prepared by FDA. Number of landslides in low and very low susceptibility zones have also decreased in the zones prepared using LR. For quantitative assessment of the accuracy of generated susceptible maps, HSS is used. HSS gives a value between 0 and 1, higher means higher accuracy of prediction. Summary for both the methods is given in Table 3.
Defining the threshold values for LSI is required for preparation of LSZ. Depending on the threshold values, the HSS score varies for both (LR and FDA) models. However in this case when the threshold values are kept more than median of 0.5, the HSS score for both the models is similar. However when the threshold is decreased, LR method tends to have higher HSS score as compared to FDA method. This suggests that for lower threshold values, i.e. landslides occurring in low and

Figure 6. Landslide susceptibility zonation map generated using ranking obtained from logistic regression. Landslide locations used for training (landslides until 2008) marked as black circle and testing (landslides until 2011) marked as red triangle are shown on the map.

Table 2. Zonal landslides analysis of maps generated using FDA and LR. Five susceptibility zones, there landslide susceptibility index range, area and number of landslide occurring in the corresponding zones are given in the table.

| Zones               | LSI range | Area (sq. km) | No of landslides | FDA Area (sq. km) | No of landslides | LR Area (sq. km) | No of landslides |
|---------------------|-----------|---------------|------------------|-------------------|-----------------|------------------|-----------------|
| Very low susceptible| 0.00–0.33 | 226.6         | 6                | 217.6             | 6               | 562.6            | 12              |
| Low susceptible     | 0.33–0.44 | 609.9         | 16               |                   |                 | 493.2            | 42              |
| Moderate susceptible| 0.44–0.57 | 453.9         | 35               |                   |                 | 236.5            | 81              |
| High susceptible    | 0.57–0.71 | 220.6         | 87               |                   |                 | 114.4            | 82              |
| Very high susceptible| 0.71–1.00| 113.1         | 79               |                   |                 |                  |                 |
very low susceptible zones can be mapped more accurately using LR as compared to FDA. However for high and very high susceptible zones, both these methods have similar accuracies. Hence, LR can be selected for better determination of rankings in these geo-environmental scenarios. While the results obtained by FDA cannot be rejected out rightly.

Some areas such as Kedarnath, Sonprayag, etc., which fall under, moderate to least susceptible zone of LSZ, even though the extreme event of 2013 levelled up this whole area. This one extreme event created 3472 new landslides and reactivated 1401 landslides out of 6013 landslides mapped in Uttarakhand by Martha et al. (2015). Most of the new landslides have occurred in the areas of Kedarnath to Son Prayag. While the Kedarnath disaster is considered as an extreme event, which has created many landslides in this region, considering these in the data-set will definitely give different results.

6. Conclusions

Landslides are a major geological hazard in any mountainous terrain posing threat to infrastructure and life. Its effects can range from being negligible to detrimental depending on the mechanism and rate of movement. Their occurrences are inevitable but appropriate studies of landslide susceptibility zonation (LSZ) can play a significant role for planning and management of public infrastructures development and mitigation of its detrimental effects. Preparation of LSZ is governed by ranking or weightages to various causative factors considered for landslide occurrence. Most of the times geomorphologists give this ranking, which tends to incorporate the subjectivity in the output maps. Hence in this paper, assignment of rank or weightage is obtained by using FDA and LR. The FDA model converts data into 1D, so it has advantage that even a large amount of data can be processed quickly and easily. In contrast, the LR model is cumbersome and takes more time to process the large amount of data. Both FDA and LR gave highest rank to streams, followed by soil then thrust. Thus, the areas near rivers/streams where thrust passes by are more susceptible. However, the elevation does not seem to have much effect on LSZ prepared by both FDA and LR. These results are validated using Heidke-Skill-Score (HSS), a statistical measure for prediction accuracy. Both very high and high susceptible zones have highest incidences of landslides that occurred after 2008 till 2011. Out of 223 occurrences of landslides, 87 occurred in high susceptible zone followed by 79 landslides that fall under very high susceptible zone according to FDA ranking. Similar results are obtained for LR as well. Median value of LSI range, i.e. 0.5 has been considered for final HSS, since for threshold >0.50, both methods show same score. For threshold <0.50, LR performs better than FDA. The results and findings of the present study can help the developers, planners and engineers to give proper weights for various factors while managing slope and land-use planning. New infrastructure development may be restricted in higher susceptibility zones. However, proper care must be taken while using the models for specific site development.

Acknowledgments

Authors would like thank Dr R. P. Singh; Ms A. S. Ningreichon and Ms Yogita Garbyal of Department of Geology, University of Delhi for helping in carrying out the geological and field mapping of this study area. We would like to
thank two anonymous reviewers whose comments really helped in shaping this manuscript. Special thanks to the ASTER GDEM team and USGS for the open data access. ASTER GDEM is a product of NASA and METI. Landsat-5 data available from U.S. Geological Survey. These data are distributed by the Land Processes Distributed Active Archive Center (LP DAAC), located at USGS/EROS, Sioux Falls, SD. http://lpdaac.usgs.gov.

Disclosure statement
No potential conflict of interest was reported by the authors.

Funding
The fieldwork for this research was supported by NRDMS project Landslide Dham (MANU Project) [project number NRDMS/11/3010/013(G)]. This work is partially supported by NRDMS [project number NRDMS/02/41/016] sanctioned to D.P. Shukla.

ORCID
Sharad Kumar Gupta http://orcid.org/0000-0003-3444-1333
Dericks P. Shukla http://orcid.org/0000-0001-6546-9203

References
Ahmed B. 2015. Landslide susceptibility mapping using multi-criteria evaluation techniques in Chittagong Metropolitan Area, Bangladesh. Landslides. 12:1077–1095.
Akgün A, Türk N. 2010. Landslide susceptibility mapping for Ayvalik (Western Turkey) and its vicinity by multicriteria decision analysis. Environ Earth Sci. 61:595–611.
Arora MK, Das Gupta AS, Gupta RP. 2004. An artificial neural network approach for landslide hazard zonation in the Bhagirathi (Ganga) Valley, Himalayas. Int J Remote Sens. 25:559–572.
ASTER GDEM Validation Team. ASTER global DEM Validation Summary Report (No. 1). METI. Washington: NASA USGS; 2009.
Bishop CM. 2006. Fisher’s linear discriminant. In: Pattern Recognition and Machine Learning. New York: Springer-Verlag: p. 186–192.
Bui TD, 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. 13:361–378.
Célérier J, Harrison TM, Webb AAG, Yin A. 2009. The Kumaun and Garwal Lesser Himalaya, India: Part 1. Structure and stratigraphy. Bull Geol Soc Am. 121:1262–1280.
Chatterjee S, Simonoff JS. 2013. Logistic Regression. In: Handbook of Regression Analysis. Hoboken, NJ, USA: John Wiley & Sons, Inc.; p. 147–176.
Chaudhary S, Gupta V, Sundriyal YP. 2010. Surface and sub-surface characterization of Byung landslide in Mandakini valley, Garhwal Himalaya. Himal Geol. 31:125–132.
Chauhan S, Sharma M, Arora MK, Gupta NK. 2010. Landslide susceptibility zonation through ratings derived from artificial neural network. Int J Appl Earth Obs Geoinf. 12:340–350.
Choubey VM, Bist K, Saini N, Ramola R. 1999. Relation between soil-gas radon variation and different lithotectonic units, Garhwal Himalaya, India. Appl Radiat Isot. 51:587–592.
Devkota KC, Regmi AD, Pourghasemi HR, Yoshida K, Pradhan B, Ryu IC, Dhital MR, Althuwaynee OF. 2013. Landslide susceptibility mapping using certainty factor, index of entropy and logistic regression models in GIS and their comparison at Mugling-Narayanghat road section in Nepal Himalaya. Nat Hazards. 65:135–165.
Dobhal DP, Gupta AK, Mehta M, Khandelwal DD. 2013. Kedarnath disaster: facts and plausible causes. Curr Sci. 105:171–174.
Dubey CS, Shukla DP, Ningreichen AS, Usham AL. 2013. Orographic control of the Kedarnath disaster. Curr Sci. 105:1474–1476.
Duda RO, Hart PE, Stork DG. 2001. Fisher Linear Discriminant. In: Pattern Classification. 2nd ed. University of Michigan: John Wiley & Sons, Inc.; p. 44–51..
Ghosh K, Bandyopadhyay S, De SK. 2017. A Comparative Evaluation of Weight-Rating and Analytical Hierarchical (AHP) for Landslide Susceptibility Mapping in Dhalai District, Tripura. In: Hazra S, Mukhopadhyay A, Ghosh A, Mitra D, Dadhwal V (eds) Environment and Earth Observation. Remote Sensing/Photogrammetry. Springer, Cham; p. 175–193.
Hong H, Pradhan B, Jeban MN, Bui DT, Xu C, Akgun A. 2016. Spatial prediction of landslide hazard at the Luxi area (China) using support vector machines. Environ Earth Sci. 75:1–14.

Hong Y, Adler RF, Huffman GJ. 2007. Satellite remote sensing for global landslide monitoring. EOS Trans Am Geophys Union. 88(37):357–358.

Ji Z, Jing P, Yu T, Su Y, Liu C. 2013 Ranking fisher discriminant analysis. Neurocomputing. 120:54–60.

Kanungo D, Arora M, Sarkar S, Gupta R. 2009. Landslide susceptibility zonation (LSZ) mapping – a review. J South Asia Disaster Stud. 2:81–105. http://krc.cbr.res.in/dspace/handle/123456789/786.

Kanungo DP, Sarkar S, Sharma S. 2011. Combining neural network with fuzzy, certainty factor and likelihood ratio concepts for spatial prediction of landslides. Nat Hazards. 59:1491–1512.

Kumar D, Thakur M, Dubey CS, Shukla DP. 2017. Landslide susceptibility mapping and prediction using support vector machine for Mandakini River Basin, Garhwal Himalaya, India. Geomorphology. 295:115–125.

Lee S, Pradhan B. 2006. Probabilistic landslide hazards and risk mapping on Penang Island, Malaysia. J Earth Syst Sci. 115:661–672.

Lee S, Pradhan B. 2007. Landslide hazard mapping at Selangor, Malaysia using frequency ratio and logistic regression models. Landslides. 4:33–41.

Malczewski J. 2000. On the use of weighted linear combination method in GIS: common and best practice approaches. Trans GIS. 4:5–22.

Martha TR, Roy P, Govindharaj KB, Kumar KV, Diwakar PG, Dadhwal VK. 2015. Landslides triggered by the June 2013 extreme rainfall event in parts of Uttarakhand state, India. Landslides. 12:135–146.

Mathew J, Jha VK, Rawat GS. 2009. Landslide susceptibility zonation mapping and its validation in part of Garhwal Lesser Himalaya, India, using binary logistic regression analysis and receiver operating characteristic curve method. Landslides. 6:17–26.

Michael EA, Samanta S. 2016. Landslide vulnerability mapping (LVM) using weighted linear combination (WLC) model through remote sensing and GIS techniques. Model Earth Syst Environ. 2:1–15.

 Micheletti N, Foresti L, Robert S, Leuenerberger M, Pedrazzini A, Jaboyeddof M, Kanevski M. 2014. Machine learning feature selection methods for landslide susceptibility mapping. Math Geosci. 46:33–57.

“NDFD verification help – site navigation.” [Online]. [accessed 2017 Mar 10]. http://www.nws.noaa.gov/ndfd/veriﬁcation/help/hss.htm.

Nourani V, Pradhan B, Ghaffari H, Sharifi SS. 2014. Landslide susceptibility mapping at Zonouz Plain, Iran using genetic programming and comparison with frequency ratio, logistic regression, and artificial neural network models. Nat Hazards. 71(1):523–547.

Pandey A, Dabral PP, Chowdary VM, Yadav NK. 2008. Landslide hazard zonation using remote sensing and GIS: a case study of Dikrong river basin, Arunachal Pradesh, India. Environ Geol. 54:1517–1529.

Pham BT, Pradhan B, Tien Bui D, Prakash I, Dholakia MB. 2016a. A comparative study of different machine learning methods for landslide susceptibility assessment: a case study of Uttarakhand area (India). Environ Model Softw. 84:240–250.

Pham BT, Pradhan B, Tien Bui D, Prakash I, Dholakia MB. 2016b. Evaluation of predictive ability of support vector machines and naive Bayes trees methods for spatial prediction of landslides in Uttarakhand state (India) using GIS. J Geomatics. 10:71–79.

Pourghasemi HR, Pradhan B, Gokceoglu C, Mohammadi M, Moradi HR. 2013. Application of weights-of-evidence and certainty factor models and their comparison in landslide susceptibility mapping at Haraz watershed, Iran. Arab J Geosci. 6:2351–2365.

Pradhan B. 2013. A comparative study on the predictive ability of the decision tree, support vector machine and neuro-fuzzy models in landslide susceptibility mapping using GIS. Comput Geosci. 51:350–365.

Prasad S, Bruce LM. 2008. Limitations of principal components analysis for hyperspectral target recognition. Geosci Remote Sens Lett IEEE. 5(4):625–629.

Ramesh V, Mani S, Baskar M, Kavitha G, Anbazhagan S. 2017. Landslide hazard zonation mapping and cut slope stability analyses along Yercaud ghat road (Kupperanur–Yercaud) section, Tamil Nadu, India. Int J Geo-Eng. 8(1):2.

Rautela P, Thakur VC. 1999. Landslide hazard zonation in Kaliganga and Madhyamaheshwar valleys of Garhwal Himalaya: a GIS based approach. Himal Geol. 20:31–44.

Rawat MS, Uniyal DP, Dobhal R, Joshi V, Rawat BS, Bartwal A, Singh D, Aswal A. 2015. Study of landslide hazard zonation in Mandakini Valley, Rudraprayag district, Uttarakhand using remote sensing and GIS. Curr Sci. 109:158–170.

Ray Y, Srivastava P. 2010. Widespread aggradation in the mountainous catchment of the Alaknanda–Ganga River system: timescales and implications to Hinterland–foreland relationships. Quat Sci Rev. 29(17–18):2238–2260.

Saha AK, Gupta RP, Sarkar I, Arora MK, Csaplovics E. 2005. An approach for GIS-based statistical landslide susceptibility zonation - with a case study in the Himalayas. Landslides. 2:61–69.

Sarkar S, Roy AK, Martha TR. 2013. Landslide susceptibility assessment using information value method in parts of the Darjeeling Himalayas. J Geol Soc India. 82(4):351–362.
Sati SP, Naithani A, Rawat GS. 1998. Landslides in the Garhwal lesser Himalaya, UP, India. Environmentalist. 18:149–155.
Shahabi H, Hashim M. 2015. Landslide susceptibility mapping using GIS-based statistical models and remote sensing data in tropical environment. Sci Rep. 5:1–15.
Shukla DP, Dubey CS, Ningreichon AS, Singh RP, Mishra BK, Singh SK. 2014. GIS-based morpho-tectonic studies of Alaknanda river basin: a precursor for hazard zonation. Nat Hazards. 71:1433–1452.
Shukla DP, Gupta S, Dubey CS, Thakur M. 2016. Geo-spatial Technology for Landslide Hazard Zonation and Prediction. In: Marghany M (ed). Environmental Applications of Remote Sensing. Rijeka, Coratia: InTech; p. 281–308.
Soeters R, Westen CJ. 1996. Slope instability Recognition, analysis and zonation. In: Turner AK, Schuster RL, editors. Landslide: Investigations and Mitigation Special Report, vol 247. Washington, D.C.: Transportation Research Board, National Research Council, National Academy Press; p. 129-177.
Terlien MTJ, Van Westen CJ, Van Asch TJ. 1995. Deterministic modelling in GIS-based landslide hazard assessment. In: Carrara A, Guzzetti F, editors. Geographical information systems in assessing natural hazards. Dordrecht: Springer; p. 57–77.
Valdiya KS, Goel OP. 1983. Lithological subdivision and petrology of the Great Himalayan Vaikrita Group in Kumaun. India Proc Indian Acad Sci – Earth Planet Sci. 92:141–163.
Vijith H, Rejith PG, Madhu G. 2009. Using InfoVal method and GIS techniques for the spatial modelling of landslide susceptibility in the upper catchment of river Meenachil in Kerala. J Indian Soc Remote Sens. 37:241–250.
Wang H, Lu X, Hu Z, Zheng W. 2014. Fisher Discriminant Analysis with L1-Norm. IEEE Trans Cybern. 44:828–842.