Spatial modelling of shallow landslide susceptibility: a study from the southern Western Ghats region of Kerala, India.

A. L. Achu, C. D. Aju and Rajesh Reghunath

*International and Inter University Centre for Natural Resources Management, University of Kerala, Thiruvananthapuram, India; †Department of Geology, University of Kerala, Thiruvananthapuram, India

ABSTRACT
The study was undertaken to produce the landslide susceptibility maps by using Dempster–Shafer, Bayesian probability and logistic regression methods for the southern Western Ghats, Kerala, India. A landslide inventory database of 82 landslides is prepared and used for landslide susceptibility modelling. Twelve landslide conditioning factors including lithology, geomorphological features, slope angle, soil texture, distance from stream, distance from road, distance from lineaments, land use/land cover, slope curvature, rainfall, topographic wetness index and relative relief are extracted from the spatial database and used for modelling. Multi-collinearity among the independent variables were tested and landslide susceptibility maps are constructed. The constructed models were validated with sensitivity, specificity, classification accuracy, ROC-AUC, root mean square error (RMSE) and kappa index. The Bayesian probability model obtained highest ROC-AUC (0.833), sensitivity (0.870), specificity (0.800) and kappa index (0.667) with least RMSE (0.4550) in validation phase. In addition, the study reveals that the agricultural areas have 10°–40° slopes falling on the denudational structural hills are extremely susceptible to landslide occurrence with extended influence from distance from roads, distance from streams and soil texture. The predicted model is trustworthy for future land use planning in the southern Western Ghats to mitigate the risk from landslide hazard.

1. Introduction
Landslides are calamitous events occurring throughout the world, specifically in mountainous and hilly areas. On an account, 75% of the global non-seismic landslides (2004–2016) are reported in seven Asian countries such as India, China, Laos, Bangladesh, Myanmar, Philippines and Indonesia (Froude and Petley 2018). Nearly, the 15% of the Indian Territory (covering about 0.49 million km²), is prone to landslide hazard (Surya 2012), affecting human life, livelihood, infrastructure and natural resources in a big way. It is estimated that on average about 500 lives are lost and costs approximately rupees 300 crore (approx. 3 billion US$) annually (Surya 2012). The most landslide-prone areas are the Himalayan belt and the Western Ghats. The Western Ghats are the broad westerly escarpment in the Indian peninsula, running for approximately 1500 km from Gulf of Cambay in the north to Cape Comorin in the south. The Western Ghats region is characterized by rugged hills with long steep-sided slopes, where loose unconsolidated soil and earth materials rest over Precambrian crystalline rocks (Sajinkumar et al. 2011). Furthermore, situated in a humid and tropical climatic region, the Western Ghats profoundly influences the regional climate on its western slopes (i.e. humid to per-humid) and the inland areas (semi-arid to arid). The western slopes receive an annual average rainfall of 3000 mm/yr⁻¹ with a mean temperature of 30°C and enhances the deformation of the saturated overburden in the form of shallow debris flow during the episodes of heavy downpour (Thampi et al. 1997; Vijith et al. 2014). Therefore, accurate and reliable landslide susceptibility maps are essential for future land use planning and hazard mitigation in Western Ghats region.

Landslide susceptibility reflects the probability of spatial variation of landslides that can answer the question of where a landslide is likely to occur (Ilia and Tsangaratos 2016; Li and Chen 2020). Landslide susceptibility mapping involves collating, handling, processing and interpreting a large amount of geographical data in a sophisticated manner (Van Westen 1994). The recent geospatial developments such as remote sensing and GIS are proven as efficient tool in handling huge volume of geographical data (Guzzetti et al. 1999; Pourghasemi, Mohammady, and Pradhan 2012; Arabameri et al., 2020).
The remote sensing techniques such as all-weather earth observation systems improve the mapping and monitoring possibilities in the high mountain chains. In addition, the spatial information related to the factors which are directly or indirectly making the terrain susceptible can be derived from remote sensing data (Van Westen, Rengers, and Soeters 2003, 2008). These data sets with an efficient geographical information system can be used to produce accurate and effective susceptibility map and development of management strategies (Vijith et al. 2014; Chen et al. 2015). However, the reliability of landslide susceptibility maps are mainly depends on the quality and quantity of the available data, selection of associated landslide influencing parameters and the appropriate methodology chosen for modelling the landslide susceptibility (Ayalew and Yamagishi 2005; Ozdemir and Altural 2013; Sameen, Pradhan, and Lee 2020).

Over the decade, a number of researches on landslide analyses using GIS and statistical techniques were carried out (Van Westen 1994; Pradhan, Singh, and Buchroithner 2006; Pradhan and Youssef 2010; Akgun, Kincal, and Pradhan 2012; Nampak, Pradhan, and Manap 2014; Pradhan and Kim 2017; Chen et al. 2018; Can et al. 2019; Arabameri et al., 2020; Nhu et al. 2020; Van Dao et al. 2020) and most landslide susceptibility researches can be grouped under three broad categories: qualitative models, physical-based models and probabilistic models. Qualitative or knowledge-driven approaches utilize prior experience and knowledge of experts, including analytical hierarchical process, fuzzy logic and other heuristic methods (Kayastha, Dhital, and De Smedt 2013; Roodposhti, Rahimi, and Beglou 2014; Zhu et al. 2014). Being subjective or partially subjective, the results of these methods vary depending upon the knowledge of experts. Physically based models predict landslide susceptibility considering the rock characteristics and failure mechanisms (e.g. Baum, Savage, and Godt 2002; Formetta et al. 2014). They mostly depend on engineering principles of slope instability expressed in terms of a factor of safety. Because of the need for comprehensive data from individual slopes, these models are often useful for mapping only smaller extents (Guzzetti et al. 1999). Probabilistic models include statistical and machine learning algorithms, which are data feeding methods and using the training data to learn and predict the information through the learning (Pourghasemi and Rahmati 2018). The main advantage of probabilistic methods is its ability to deal with large quantity of heterogeneous data and predicting engineering complex problems more accurately (Qiao and Yang 2019; Fan et al. 2019; Zhang et al. 2019; Liu et al. 2019). Prominent machine learning methods include artificial neural network (Harmouzi et al. 2019; Huang et al. 2020; Bragagnolo, da Silva, and Grzybowski 2020), support vector machines (Zhu et al. 2019; Fang et al. 2020; Pandey, Pourghasemi, and Sharma 2020; Sameen, Pradhan, and Lee 2020), random forest (Zhu et al. 2019; Arabameri et al., 2020; Fang et al. 2020; Chen et al. 2020a), classification and regression trees (Pham, Prakash, and Bui 2018; Ghasemain et al. 2020) and deep learning neural networks (Fang et al. 2020; Sameen, Pradhan, and Lee 2020; Nhu et al. 2020; Van Dao et al. 2020). Statistical methods include Dempster–Shafer model (Althuwaynee, Pradhan, and Lee 2012; Li and Chen 2020), Bayesian probability model (Vijith et al. 2014; Gadtaula and Dhakal 2019; Chen et al. 2020b), Certainty factor (Devkota et al. 2013; Chen et al. 2020b), Information value (Juliev et al. 2019; Chen et al. 2020a) and multivariate statistical model such as logistic regression (Vijith et al. 2014; Mondal and Mandal 2018; Zhu et al. 2019; Sahin, Colkesen, and Kavzoglu 2020; Fang et al. 2020) take into account the spatial association between landslides and causative factors. Bivariate statistical approaches use the idea of the spatial distribution of previous landslides with a set of selected causative factors and give weighting accordingly. The bivariate statistical models such as Dempster–Shafer models (DSM), Bayesian probability model (BPM) and multivariate method such as multivariate logistic regression (MLR) differ each other. BPM is ideal for solving decision-making problems under uncertainties, for instance, the given uncertainty is associated with the landslide phenomenon and its relationship between the landscapes and associated environmental variables (Pourghasemi, Mohammady, and Pradhan 2012). DSM is a modified form of theory of evidence and the main advantage is that the belief, disbelief, uncertainty and plausibility associated with landslide occurrence can be quantified statistically (Althuwaynee, Pradhan, and Lee 2012). Logistic regression is well known for its ability to predict the presence or absence of a phenomenon or outcome based on the values of the input predictor variables (Krishnan et al. 2015). Hence, a robust comparison of these techniques will be useful in choosing the best method for future landslide susceptibility modeling. The previous studies, in this regard (Pourghasemi, Mohammady, and Pradhan 2012; Pourghasemi et al. 2013; Vijith et al. 2014; Krishnan et al. 2015; Chen et al. 2015; Ding, Chen, and Hong 2016; Mondal and Mandal 2018; Bera, Guru, and Ramesh 2019), are mainly focused on explaining the overall accuracy of the models with receiver operating curve and area under curve value (ROC-AUC), which is difficult to interpret and unable to find the prediction accuracy of both landslide pixels and
non-landslide pixels. We address the issue by using matrix-based sensitivity, specificity and classification accuracy methods and quantifying the prediction error through RMSE.

In this paper, an attempt is made to produce landslide susceptibility map in a highland segment of southern Western Ghats by comparing the bivariate and multivariate statistical methods. In addition, there is an interest to check how these models vary in their prediction efficiency in extreme climatic events or are they capable of predicting the landslide incidents even in extreme events. During the time of the present study, a major deluge was witnessed in the state of Kerala causing hundreds of landslides together with heavy flooding and numerous slope failures. Hence, the details of the landslides associated with the 2018 deluge were collected and used for validation purposes in order to assess the usefulness of the models in the future.

2. Study area and spatial database

The study area is situated in the westerly slopes of the Western Ghats between 11°17’ 32.98”N to 11°40’ 51.45”N latitude and 75°53’ 35.83”E to 76°15’ 10.13”E longitude and spread over 720.22 km² (Figure 1). The region belongs to the western edge of Wayanad plateau and spread over the Kozhikode, and Wayanad districts of Kerala state. The terrain is generally high in north and northwest and low in south. Elevation ranges between 20 m and 2325 m above mean sea level. Like other parts of the state, the area experiences tropical monsoon climate with two distinct seasons such as south-west monsoon and north-east monsoon. The south-west monsoon (June–September) alone contributes to 80% of the total rainfall in the study area with 3000 mm annual average rainfall and a mean temperature of 27°C.

A landslide inventory map is prepared through a series of field visits with the aid of high-resolution satellite images. The prevalent and disastrous type of landslides is ‘debris flows’ (hereafter referred as landslide) (Figure 2). The characteristic of this phenomenon is the swift and sudden down slope movement of highly water saturated overburden containing a varied assemblage of debris material ranging in size from soil particles to huge boulders destroying and carrying with it everything that is lying in its path (Kuriakose, Sankar, and Muraleedharan 2009; Vijith et al. 2014). A total of 82 landslides are identified and mapped in the field. The surface area of the landslides varied from 837 m² to 118,000 m² (Figure 2(a,b)). The centroid of the landslides

![Figure 1. Location map of the study area with mapped landslide locations.](image-url)
are converted into points and divided into 70%–30% proportion for training and testing the models.

As a part of this study, the relationship between shallow landslides and associated environmental variables were examined. Twelve landslide causative factors (lithology, geomorphological features, land use/land cover, soil texture, slope angle, slope curvature, distance from stream, distance from lineaments, distance from road, relative relief, topographic wetness index (TWI) and spatial variation of annual average rainfall) are selected based on expert opinion, availability of data at desired scale and from previous studies in the Ghats region as well as other parts of the world with similar environmental conditions (e.g. Vijith and Madhu 2008; Prasannakumar and Vijith 2012; Kayastha, Dhital, and De Smedt 2013; Vijith et al. 2014; Ghorbanzadeh et al. 2018; Can et al. 2019; Li and Chen 2020). The lithology data were generated from district resources maps (scale 1:250,000) produced by Geological Survey of India (GSI). The geomorphology map of the study area is gathered from Kerala State Remote Sensing and Environment Centre (scale 1:50,000). Land use/land cover data of the study area are collected from the Kerala State Land Use Board (scale 1:50,000). Terrain parameters, such as slope angle, slope curvature, relative relief and TWI are calculated from the SRTM DEM (1 arc second). Soil texture data of the study area are collected from the Department of Soil Survey & Soil Conservation, Kerala and rainfall data (2007–2017) are collected from the Indian Meteorological Department (IMD).

The study area forms a part of the Precambrian metamorphic shield with rocks of the Wayanad Group consisting of basic rocks, peninsular gneissic complex, charnockite group of rocks, high-grade metamorphic rocks and intrusive of migmatite complex (Soman 1987). Charnokite is the dominant lithology present in the study covering an area of 298 km² (Figure 3(a)). Six geomorphological features such as denudational structural hills, pediplain, piedmont zone, plateau, residual hill and rock exposure are identified and demarcated in the study area. About 51% of the study area is covered by denudational structural hill (Figure 3(b)). Major land uses land covers present in the study area are grassland, forest plantations, evergreen forest, deciduous forest, double crop, settlements, wetlands, wastelands, agricultural areas and water bodies (Figure 3(c)). It is also noted in the field visits that major landslides in the study area are associated or nearby agricultural areas in steep slopes where small bunds or check dams are present. Clay and gravelly clay are the dominated soil textures found in the study area, but gravelly loam and loam are also present in the study area (Figure 3(d)). During field visits, it is observed that majority of previous landslides occurred at medium slopes (i.e. slope between 10° and 30°) (Figure 3(e)). The slope angle is further divided into 0°–10°, 10°–20°, 20°–30°, 30°–40° and 40°–75.36° classes to assess the significance of slope angle in the occurrence of landslides. Slope curvature represents the slope morphology which influence surface runoff and flow divergence; therefore, slope curvature is considered an important terrain parameter which influences slope stability (Devkota et al. 2013; Ding, Chen, and Hong 2016). In the present study, standard curvature (i.e. combines both the profile and planform curvature) is derived from SRTM DEM using ArcGIS (Figure 3(f)). Generally, flat and convex slopes are referred as stable, whereas concave slopes are potentially unstable as they accumulate water on the lower reaches and lead to adverse hydraulic effects (Stocking 1972). Relative relief of an area represents the ranges between highest and lowest points in a unit area (Vijith et al. 2014). Relative relief of the study area ranges from 16 to 948 m/km² which is
reclassified into five classes such as 16–150, 150–300, 300–450, 450–600 and 600–948 m/km² (Figure 3(g)). The TWI is used to quantify the topographic controls in hydrological processes and it is a function of both slope and upslope contributing area (Devkota et al. 2013). TWI values of the study area range from 1.75 to 23.35 and are regrouped into three classes viz., <5, 5–10 and >10 (Figure 3(h)). The proximity parameters such as distance from streams, roads and lineaments are prepared from various
sources and used in this study. Drainage lines are generated from SRTM DEM using the Arc Hydro tool in ArcGIS and buffer zones of <100, 100–200, 200–300, 300–500 and >500 m are used to evaluate the spatial association between streams and past landslides (Figure 3(i)). Lineaments map is prepared from the combination of Landsat 8 OLI image, SRTM DEM and drainage lines. The relationship between lineament distance and landslides is find out using <200, 200–400, 400–600, 600–800 and >800 m buffer zones (Figure 3(j)). Construction of roads in mountainous areas causes loss of toe support of slopes and the change of topography leads to stress on the back of slopes leading to slope instability (Ding, Chen, and Hong 2016). Road network of the study area is generated from Survey of India topographical sheets and updated using Google Earth. In addition, <100, 100–200, 200–300, 300–500 and >500 m buffers were prepared to assess the relation between road distance and landslide occurrence (Figure 3(k)). Being located in the humid tropics, rainfall is the major triggering factor of shallow landslides; therefore, the rainfall data of the study area are collected for the available rain gauge stations (three) and IDW interpolation method is used to generate the continuous raster surface of the rainfall (Figure 3(l)).

3. Methods

The general methodology followed in the study is given in Figure 4, includes six major steps, starting from the data collection (previous landslides) through a series of field surveys with the aid of high resolution satellite images and once the landslide inventory is completed the next step is the preparation of the landslide-conditioning factors which is discussed in spatial database section. The third step is to find out the correlation between the predictor variables which was succeeded by using multi-collinearity check using VIF, tolerance and Pearson’s correlation coefficient. Thereafter, landslide susceptibility modelling is carried out using the both multivariate and bivariate methods. Next step is the validation of the predicted models, for this, a twofold methodology is adopted. First, the validation is done using matrix-based statistical measures and ROC-AUC value. Second, validation is by using recent landslides associated with a deluge in August 2018. The sixth and final step is finalizing the model and its applicability in the Western Ghats region.

3.1. Bayesian probability modelling

The BPM is a log-linear version of Bayes general theorem to estimate the relative importance of piece of evidence by statistical means (Bonham-Carter 1994; Vijith et al. 2014). The BPM works on the basic premise that the probability of a landslide occurrence at a particular location can be calculated by updating the event’s prior probability of occurrence in the study area using measures of spatial association between known event occurrences and evidential or predictive maps (Bonham-Carter 1994). The BPM calculates the weight for each landslide predictive factor based on the presence (positive) or absence (negative) of the known landslides within the area of each binary predictor theme. The weighted values of the classes of landslide causative factors are calculated using the following equations (Regmi, Giardino, and Vitek 2010; Ozdemir and Altural 2013):

$$W^+ = \ln \frac{A_1}{A_1 + A_2}$$

$$W^- = \ln \frac{A_2}{A_1 + A_2}$$

where $A_1$ is the number of landslide pixels present in the given factor class, $A_2$ is the number of landslide pixels absent in the same factor class, $A_3$ is the number of...
pixels in the given factor class with no landslides and \( A_4 \) is the number of pixels in the given factor class when neither a landslide nor a given factor is present.

A positive weight \( W^+ \) indicates the importance of the presence of a factor for the occurrence of landslides. If the \( W^+ \) is positive, the presence of the factor is favourable to the occurrence of landslides, whereas a negative value of \( W^- \) indicates the non-favourability. \( W^- \) is used to evaluate whether the absence of a factor is favourable for the occurrence of landslides. A positive value of \( W^- \) indicates the absence of such a factor is favourable for landslide occurrence and when it is negative the factor is non-favourable. The difference between \( W^+ \) and \( W^- \) is denoted as weight contrast which can be expressed as:

\[
C = (W^+ - W^-) \tag{3}
\]

where \( C \) reflects the overall spatial association between the prediction variable and landslide occurrence. A contrast value equal to zero indicates that the class is not significant to the analysis, whereas a positive value indicates a positive spatial association (Ozdemir and Altural 2013; Vijith et al. 2014).

### 3.2. Dempster–Shafer modelling

The Dempster–Shafer theory of evidence is first introduced by Dempster (1967) and is later modified by Shafer (1976). DSM is a bivariate statistical approach, which is useful for combining evidential maps of spatial recognition criteria that are conditionally dependent on each other and on a target variable (Walley 1987). The DSM consists of four basic propositions; belief (Bel), disbelief (Dis), uncertainty (Unc, ignorance or doubt) and plausibility (Pls) with values ranging from 0 to 1. Bel and Pls, which are the lower and upper degrees of belief, indicate the ‘pessimistic’ and ‘optimistic’, measures of spatial association of landslide, respectively (Pradhan and Kim 2017). Pls may be greater than or equal to Bel, and Unc is the difference between belief and plausibility. Dis is the belief of the proposition being false on given evidence (Dempster 1967) and is equal to 1-Pls (or 1-Unc – Bel). The values of Unc are always positive as the minimum possible value for Pls is equal to Bel. Therefore, Bel + Unc + Dis for evidence with respect to any proposition is always equal to 1 (i.e. maximum probability) (Gorum and Carranza 2015).

The Bel, Dis, Pls and Unc functions should be applied to all landslide conditioning factors in landslide susceptibility modelling. Each map represents a prediction of evidence and the integration of all factors shows the exact level of prediction (Althuwaynee, Pradhan, and Lee 2012). If the study area has \( N \) number of multiple thematic layers, where each layer is considered as evidence \( E_{ij} \) (where \( i \) = amount of layers and \( j \) = class attribute), then Bel is represented by Equations (4) and (5).

\[
WE_{ij} = \frac{N(L \cap E_{ij})}{N(L)} \tag{4}
\]

where \( W \) is the weight of \( E_{ij} \) which supports the belief a landslide exists rather than being absent, \( N(L \cap E_{ij}) \) is the number of landslide pixels in the domain, \( N(L) \) is the total number of landslides, \( N(E_{ij}) \) is the total number of pixels in the domain. \( N \) is the proportion of landslide occurrence. This equation is applied to all conditioning factors, the Bel values are calculated using Equation (5) as follows:

\[
Bel = \frac{\sum_{j=1}^{m} WE_{ij}}{\sum_{j=1}^{m} WE_{ij}} \tag{5}
\]

Similarly, the Dis values can be calculated using Equation (6).

\[
WE_{ij} = \frac{N(L) - N(L \cap E_{ij})}{N(L)} \tag{6}
\]

where \( \bar{W} \) is the weight of \( E_{ij} \) which supports the belief that a landslide is more absent than present. Then the Dis values for all conditioning factors are calculated from Equation (7)

\[
Dis = \frac{\sum_{j=1}^{m} WE_{ij}}{\sum_{j=1}^{m} WE_{ij}} \tag{7}
\]

Then the Unc = 1 – Dis – Bel, Pls = 1 – Dis. For a case of \( E_{ij} \) with no landslide occurrence (Bel\(E_{ij}0\)), then Dis\(E_{ij}\) is reset to 0.

### 3.3. Logistic regression modelling

Logistic regression performs a multivariate regression relation between dependent factor and several independent factors (Atkinson and Massari 1998; Mondal and Mandal 2018). The dependent variables are dichotomous, whereas the independent variables are integral, continuous, dichotomous or categorical (Devkota et al. 2013). In case of landslides, the dependent variable is binary depicting the presence and absence of landslides. For the selection of non-landslide pixels, several methods have been developed and tested with high precision (Zhu et al. 2019). In the present study, the presence data are collected from field and absence data are randomly created on the basis of safe slope criteria (i.e. slopes where zero landslides are selected and on the basis of expert judgement these slope pixels are converted to
To find out the best fitting model to describe the relation between the presence and absence of landslides (Akgun, Kuncal, and Pradhan 2012; Krishnan et al. 2015). The logistic regression model representing the maximum likelihood regression model can be expressed as Equation (8)

\[ P = \frac{1}{1 + e^{-z}} \]  

(8)

where \( P \) is the probability of landslide occurrence and \( e \) is the exponential. The probability values range from 0 to 1 on a S-shaped curve. The parameter \( z \) is defined as

\[ z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n, \]  

(9)

where \( \beta_0 \) represent the intercept of the model and \( n \) is the number of independent variable, \( \beta_i \) (\( i = 1, 2, 3, \ldots, n \)) is the slope coefficient and \( X_i \) (\( i = 1, 2, 3, \ldots, n \)) is the independent variable. On the basis of Equations (8) and (9), the logistic regression can be written in the following extended form.

\[ \text{Logit}(P) = \frac{1}{1 + e^{X_1 \beta_1 + X_2 \beta_2 + \ldots + X_n \beta_n}} \]  

(10)

The estimates of a regression model cannot be uniquely computed when a perfect linear combination exists among the predictor variables (Ozdemir and Altural 2013). The involvement of more than two predictors in near perfect linear combination is often known as multi-collinearity and MLR model is sensitive to collinearities among the predictor variables. Subsequently, in this study, the multi-collinearity among independent variables is estimated prior to the analysis.

### 3.4. Model validation

Regardless of the methodology applied, the model performance and validation are essential with known landslide events to assess the reliability of the model. The ROC-AUC, RMSE and matrix-based methods such as sensitivity, specificity, classification accuracy (CLA) and kappa index are used for the model performance evaluation and validation. Sensitivity is regarded as the proportion of landslide pixels that are correctly classified as landslide occurrences. Specificity is defined as the proportion of the non-landslide pixels that are correctly classified as non-landslides (Chen et al. 2018). CLA is the proportion of landslide and non-landslide pixels that are correctly classified. The ROC with AUC value is considered as a standard technique for the performance evaluation and validation of landslide susceptibility model (Althuwaynee, Pradhan, and Lee 2012; Ding, Chen, and Hong 2016; Arabameri et al., 2020). RMSE is a good measure of how accurately the model predicts the responses. Sensitivity, specificity, CLA and RMSE can be calculated by using the following Equations (11)–(14).

\[ \text{Sensitivity} = \frac{TP}{TP + FN} \]  

(11)

\[ \text{Specificity} = \frac{TN}{TN + FP} \]  

(12)

\[ \text{CLA} = \frac{TP + TN}{TP + TN + FP + FN} \]  

(13)

\[ \text{RMSE} = \sqrt{\frac{(P_1 - a_1)^2 + \ldots + (P_n - a_n)^2}{N}} \]  

(14)

where \( P_1 \) and \( a_1 \) are the predicted and actual values, respectively. The Kappa index (K) is calculated using the following Equation (15).

\[ K = \frac{P_{\text{obs}} - P_{\text{exp}}}{1 - P_{\text{exp}}} \]  

(15)

where \( P_{\text{obs}} \) is the observed agreement and \( P_{\text{exp}} \) is the expected agreement which can be explained by following equations

\[ P_{\text{obs}} = \frac{TP + TN}{N} \]  

(16)

\[ P_{\text{exp}} = \frac{(TP + FN) \times (TP + FP) + (FP + TN) \times (FN + TN))}{\sqrt{N}} \]  

(17)

where \( N \) is the number of training pixels. The kappa values are in the range of 0–1 where Kappa magnitude is almost perfect (0.8–1.0), substantial (0.6–0.8), moderate (0.4–0.6), fair (0.2–0.4), slight (0–0.2) and poor (≤0) (Landis and Koch 1977). Further, the accuracy of each susceptibility level is estimated with frequency ratio (FR) method.

### 4. Results and discussions

#### 4.1. Multi-collinearity analysis

According to the multi-collinearity analysis, all the independent factors and factor classes are below the range of critical threshold (VIF >5 and tolerance <0.2) (Bui et al. 2015), indicating the absence of multi-collinearity among the 12 landslide conditioning factors (Table 1). In the case of Pearson’s correlation coefficient, the highest correlation coefficient value (0.63) is observed between slope and relative relief. However, the value is smaller than the limit of 0.70 indicating low collinearity (Booth, Niccolucci, and Schuster 1994; Bui et al. 2015).
Table 1. Multi-collinearity analysis for the landslide conditioning factors.

| Sl. No. | Landslide conditioning factors | MLR Tolerance | MLR VIF | DSM Tolerance | DSM VIF | BPM Tolerance | BPM VIF |
|---------|--------------------------------|---------------|---------|---------------|---------|---------------|---------|
| 1       | Lithology                      | 0.67          | 1.50    | 0.69          | 1.45    | 0.67          | 1.48    |
| 2       | Geomorphological features      | 0.86          | 1.14    | 0.63          | 1.58    | 0.52          | 1.91    |
| 3       | Land use/land cover            | 0.63          | 1.58    | 0.67          | 1.50    | 0.60          | 1.67    |
| 4       | Distance from lineaments       | 0.88          | 1.14    | 0.92          | 1.09    | 0.90          | 1.12    |
| 5       | Distance from roads            | 0.79          | 1.27    | 0.87          | 1.15    | 0.88          | 1.13    |
| 6       | Distance from streams          | 0.74          | 1.36    | 0.80          | 1.26    | 0.81          | 1.23    |
| 7       | Rainfall                       | 0.46          | 2.16    | 0.74          | 1.35    | 0.73          | 1.37    |
| 8       | Relative relief                | 0.32          | 3.16    | 0.51          | 1.96    | 0.49          | 2.05    |
| 9       | Slope angle                    | 0.25          | 4.02    | 0.62          | 1.61    | 0.64          | 1.57    |
| 10      | Slope curvature                | 0.78          | 1.28    | 0.70          | 1.42    | 0.73          | 1.37    |
| 11      | Soil texture                   | 0.39          | 2.59    | 0.88          | 1.14    | 0.84          | 1.19    |
| 12      | TWI                            | 0.54          | 1.86    | 0.81          | 1.24    | 0.80          | 1.25    |

Subsequently, the correlation between factors classes are shows the absence of correlation between them (Supplementary Tables S1–S3). Therefore, all the 12 factors were included in the final susceptibility analysis.

4.2. Landslide susceptibility mapping

The spatial association between landslides and each conditioning factor is analysed using data-driven DSM, BPM and MLR models (Tables 2 and 3). It is widely recognized that the underlying lithology has a significant influence in the occurrence of landslides. Furthermore, the lithological and structural variations often lead to instability on strength and permeability of both bedrocks and overlying burden. With a positive $\beta$ value of 0.050, the lithology is a significant factor which controls the occurrence of landslides in the study area, for MLR. Among the different lithological formations, most of the shallow landslides are associated with the Charnockite group of rocks with a maximum contrast value of 0.49 and Bel value of 0.44 (Table 2). Minimum values are retained in basic rocks, high grade metasedimentary rocks and metabasic and ultrabasic rocks since no landslides are reported. Approximately, 78% of the past landslides are reported in denudational structural hills where the probability of landslide occurrence is high with a maximum predicted C value of 1.20. Whereas in the case of DSM, maximum Bel value of 0.51 is assigned to residual hills instead of denudational structural hills. That is mainly due to lower areal extent. However, landforms such as pediplain and plateau are the stable landforms which are devoid of landslides. In the case of land use/land cover, 24 landslides are reported in the deciduous forest and 23 are reported in agricultural area which indicates the high susceptibility of landslide occurrence in these categories. Consequently, the maximum C value of 0.77 and Bel value of 0.25 is noted for the deciduous forest class. However, the lowest values are found in grasslands and wetlands which are practically devoid of landslide occurrence. Consequently, a MLR coefficient value of 0.260 indicates an overall positive association between landslides and land use/land cover. It is also noted that ~41.38% of the landslide occurrence is reported in gravely loam texture soils which indicates the very high probability of landslide occurrence in this type of soil texture. Accordingly, the maximum Bel value of 0.62 and C value of 1.24 are assigned to this category. A positive $\beta$ value of 0.406 indicates that the soil texture plays a significant role in shallow landslide initiation in the study area.

According to the logistic regression, slope is the most important landslide conditioning factor in the study area having a maximum $\beta$ value of 0.664. It is also noticed that the ~74% of the reported landslides are reported between 10° and 30° slopes. Even though the 10°–20° and 20°–30° slope categories have similar numbers of landslide occurrences (22 and 21), the maximum C value of 1.04 and Bel value of 0.42 is noticed in the 20°–30°category and this is due to lower areal extent. Furthermore, a minimum Dis value of 0.15 in this category indicates the strong belief of landslide occurrence. It is also noted that no landslides are reported the slope above 40° which is mainly due to the fact that in the study area, slope above 40° is mainly occupied by rocky cliffs without soil colour. Hence, the chances of shallow debris flow occurrence are impractical. In the case of relative relief, 21 landslides are reported in the 150–300 m/km² category, indicating the higher probability of landslide occurrence. Accordingly, the minimum disbelief value (0.15) and maximum Bel and C value (0.32 and 0.88, respectively) are associated with this class. The MLR coefficient value (0.613) indicates that relative relief has a significant role in landslide occurrence in the study area. It has been observed from the field as well as the models the majority of landslides in the study area are associated with closer road distance.
Table 2. Spatial relationship between landslide and its causative factor derived from BPM and DSM.

| Theme                                      | NPF* | Nol* | Bayesian probability model | Dempster-Shafer model |
|--------------------------------------------|------|------|---------------------------|------------------------|
|                                            |      |      | W*          | W+          | C          | Bel | Disc | Unc | Pls |
| Lithology                                  |      |      | (Continued) |
|                                            |      |      |              |              |            |      |      |     |     |
| Sedimentary rocks (SRL)                    |      |      |              |              |            |      |      |     |     |
|                                          |      |      |              |              |            |      |      |     |     |
| Soil texture                               |      |      |              |              |            |      |      |     |     |
|                                            |      |      |              |              |            |      |      |     |     |
| Land use/land cover                        |      |      |              |              |            |      |      |     |     |
|                                            |      |      |              |              |            |      |      |     |     |
| Slope angle (°)                            |      |      |              |              |            |      |      |     |     |
|                                            |      |      |              |              |            |      |      |     |     |
| Relative relief (m/km²)                    |      |      |              |              |            |      |      |     |     |
|                                            |      |      |              |              |            |      |      |     |     |
| Distance from roads (m)                    |      |      |              |              |            |      |      |     |     |
|                                            |      |      |              |              |            |      |      |     |     |
| Distance from streams (m)                  |      |      |              |              |            |      |      |     |     |
|                                            |      |      |              |              |            |      |      |     |     |
| Distance from lineaments (m)               |      |      |              |              |            |      |      |     |     |
|                                            |      |      |              |              |            |      |      |     |     |
| Rainfall (mm)                              |      |      |              |              |            |      |      |     |     |
|                                            |      |      |              |              |            |      |      |     |     |
The distance from roads indicates that extensive road construction in hilly terrain disturbs the angle of repose by removing the toe support and widening of roads also causes slope instability. Further, maximum C and Bel values (0.96 and 0.39) are observed in the less than the 100 m category. The logistic regression coefficient for distance from road is −0.084 which shows an overall negative relationship between landslides and distance from road. In the case of distance from lineaments, most of the landslides are located away from lineaments (i.e. >800 m) whereas maximum weights derived from DSM and BPM are assigned in 200–400 m class due to lower the areal extend. ~90% of previous landslides are reported near streams (i.e. >100 m and 100–200 m distances) indicates that streams affect slope stability through eroding the slope materials and saturating the lower reaches. The highest Bel (0.36) and C values (0.32) are noticed at 100–200 m distance from streams. MLR value of −0.112 indicates the negative influence at distance from streams in the study area, for MLR. More than 83% landslides of the study area are reported between the TWI value of 5 and 10, where the C and Bel values are maximum (0.73 and 0.51, respectively). With a high positive β value of 0.399, rainfall is a major triggering factor of shallow landslides in the study area. Maximum Bel values and C values (0.51 and 0.81, respectively) are noticed in the rainfall category between 2000 and 2300 mm. Convex slopes are more stable than concave slopes because concavities tend to concentrate surface water, and hence, there is increased energy of surface run-off and higher pore water pressure for land sliding. Increased pore pressure results in decreased shear strength of the soil, and the chance of occurrence of landslides increases. Accordingly, 62% landslides are observed in the concave slopes of the study area.

Dempster rule of combination is applied to obtain the four integrated models viz., including Belief, Disbelief, Uncertainty and Plausibility functions. The spatially distributed MASS functions derived by DSM are shown in Figure 5. A comparison between Belief and Disbelief maps of the study area (Figure 5(a,b)) shows that the Belief values are higher in the study area where Disbelief values are low. The low Disbelief and Belief values have inverse relationships in terms of landslide occurrence. A low disbelief value supports strong belief of landslide occurrence and vice versa. The major advantage of DSM is not only predicting the landslide susceptibility but also quantifying the uncertainty related to this (Nampak, Pradhan, and Manap 2014; Pradhan and Kim 2017). The uncertainty values are low where Bel values are high especially in the steep slope area (Figure 5(c)). Flat terrains have comparatively high uncertainty values with low bel values. The plausibility map is similar to the Bel map except that the contrast between lower and higher degree is less apparent in the belief map.

### Table 2. (Continued).

| Theme              | NPF* | NoL* | W* | W− | C   | Bel  | Disc | Unc  | Pls  |
|--------------------|------|------|----|----|-----|------|------|------|------|
| Slope curvature    |      |      | −0.03 | 0.00 | −0.03 | 0.33 | 0.34 | 0.33 | 0.66 |
| Flat               | 28,372 | 2 |     |     |     |      |      |      |      |
| Concave            | 417,898 | 36 | 0.17 | −0.23 | 0.40 | 0.41 | 0.27 | 0.33 | 0.73 |
| Convex             | 353,994 | 20 | −0.25 | 0.16 | −0.41 | 0.27 | 0.40 | 0.34 | 0.60 |

NPF* = number of pixels in the factor class, NoL* = number of landslides in the factor class.

### Table 3. Variables retained in the multivariate logistic regression model and their coefficients.

| Landslide conditioning factors | βa | dfb | Expl(β) | Lower | Upper |
|-------------------------------|----|-----|---------|-------|-------|
| Constant                      | −0.043 | 1 | 0.958     |       |       |
| Lithology                     | 0.050 | 1 | 1.052     | 0.612 | 1.807 |
| Geomorphological features     | 0.453 | 1 | 1.574     | 0.903 | 2.742 |
| Land use/land cover           | 0.260 | 1 | 1.297     | 0.791 | 2.126 |
| Distance from lineaments      | −0.215 | 1 | 0.807     | 0.505 | 1.288 |
| Distance from roads           | −0.310 | 1 | 0.734     | 0.465 | 1.158 |
| Distance from streams         | 0.004 | 1 | 1.004     | 0.571 | 1.765 |
| Rainfall                      | 0.399 | 1 | 1.490     | 0.895 | 2.480 |
| Relative relief               | 0.613 | 1 | 1.942     | 1.111 | 3.392 |
| Slope angle                   | 0.664 | 1 | 1.846     | 1.162 | 2.933 |
| Slope curvature               | 0.334 | 1 | 1.397     | 0.836 | 2.333 |
| Soil texture                  | 0.406 | 1 | 1.501     | 0.896 | 2.515 |
| TWI                            | 0.283 | 1 | 1.328     | 0.844 | 2.087 |

aLogistic coefficient, bdegrees of freedom.

ANNALS OF GIS 123
Finally, the C, Bel and MLR coefficient values are used to prepare the final susceptibility maps using Map Algebra tool in ArcGIS on a pixel to pixel level. In the literature, it is possible to see different types of classification schemes such as equal interval, quantile, standard deviation, natural break and geometrical interval. In the present study, the best results are achieved through the natural break interval method. The final susceptibility maps are further classified into five distinct classes (Figure 6), (1) stable area, (2) low susceptible area, (3) moderately susceptible, (4) highly susceptible area and (5) extremely susceptible zones based on natural break classification scheme (Figure 6(a–c)).

4.3 Performance evaluation and model validation

4.3.1 Model performance evaluation

The training data set is used evaluate the model performance and the result is given in Table 4 and Figure 7(a). BPM shows optimum performance in classifying the landslide pixels (sensitivity = 0.862) followed by DSM.

(Figure 5(d)). Finally, the C, Bel and MLR coefficient values are used to prepare the final susceptibility maps using Map Algebra tool in ArcGIS on a pixel to pixel level. In the literature, it is possible to see different types of classification schemes such as equal interval, quantile, standard deviation, natural break and geometrical interval. In the present study, the best results are achieved through the natural break interval method. The final susceptibility maps are further classified into five distinct classes (Figure 6), (1) stable area, (2) low susceptible area, (3) moderately susceptible, (4) highly susceptible area and (5) extremely susceptible zones based on natural break classification scheme (Figure 6(a–c)).

4.3 Performance evaluation and model validation

4.3.1 Model performance evaluation

The training data set is used evaluate the model performance and the result is given in Table 4 and Figure 7(a). BPM shows optimum performance in classifying the landslide pixels (sensitivity = 0.862) followed by DSM.

Figure 5. MASS functions derived by DSM: Belief (a), Disbelief (b), Uncertainty (c) and Plausibility (d).
In the case of classifying non-landslide pixels, BPM also out-
weigh other methods with a specificity value of 0.793
followed by DSM (specificity = 0.776) and MLR (specific-
ity = 0.754). In general, BPM performed well in classify-
ing the landslide pixels and non-landslide pixels
(CL = 0.828) followed by DSM (CLA = 0.784) and MLR
(CLA = 0.759). In terms of kappa index, BPM indicates
a substantial agreement with model and reality
(k = 0.655) whereas both DSM and MLR show moderate
agreement (Table 4). According to the ROC-AUC analysis,
the BPM shows the overall best performance in success
rate curve with 80% accuracy. Whereas DSM and MLR
models show 73.3% and 66.9% accuracy in success rate
curve (Figure 7(a)).

### 4.3.2. Model validation

Since the model performance is evaluated using training
data sets (which is used to construct the model) the

---

**Table 4. Model performance evaluation using statistical
measures.**

| Models | TP  | TN  | FP  | FN  | N   | Sensitivity | Specificity | CLA  | K   |
|--------|-----|-----|-----|-----|-----|-------------|-------------|------|-----|
| BPM    | 50  | 46  | 12  | 8   | 116 | 0.862       | 0.793       | 0.828 | 0.655|
| DSM    | 46  | 45  | 13  | 12  | 116 | 0.793       | 0.776       | 0.784 | 0.569|
| MLR    | 42  | 46  | 15  | 13  | 116 | 0.764       | 0.754       | 0.759 | 0.517|

---

**Figure 6.** Landslide susceptibility map produced by (a) BPM, (b) DSM, (c) MLR and (d) spatial distribution of landslides occurred in August 2018 plotted over the predicted models.
reliability of the model is estimated using validation data sets (which is not used for model construction). In the validation phase, MLR achieved the maximum accuracy in classifying the unknown landslide pixels (sensitivity = 0.882) followed by BPM (sensitivity = 0.870) and DSM (sensitivity = 0.850). In the case of classifying the non-landslide pixels, BPM shows better accuracy (specificity = 0.800) than both MLR (specificity = 0.771) and DSM (specificity = 0.813). Subsequently, the maximum classification accuracy was achieved by BPM (CLA = 0.833) followed by DSM (CLA = 0.786) and MLR (CLA = 0.771). In the case of RMSE, DSM shows maximum error (0.5970) whereas BPM achieved least error (0.4550) (Table 5). In the testing data sets, BPM also shows substantial agreement with model and reality (k = 0.655) whereas DSM and MLR shows moderate agreement (Table 5). Finally, also in the prediction rate, BPM retained its superiority over the other two methods with 83.3% prediction accuracy with least standard error. Unlike the bivariate approaches, MLR model obtained less accuracy in prediction rate (i.e. 70.1% accuracy) (Figure 7(b)).

The FR method is applied to estimate the accuracy of each susceptibility level (Table 6). MLR obtained highest FR value in extremely susceptible class followed by BPM and DSM. DSM reported its maximum FR value in highly susceptible area compared to extremely susceptible class. The prediction errors of the models are assessed quantitatively by counting the occurrence of landslide events in stable are. It is noticed that no landslides are found in the stable area derived by both the BPM and DSM. While stable area of MLR accounts for 4.88% of landslides and the FR of the least susceptible zone is 0.25 which is higher than that of the other models. In general, BPM shows a consistency in FR values of the different susceptible classes.

![Figure 7. The ROC-AUC values of the predicted landslide susceptibility models (a) success rate and (b) prediction rate.](image-url)

| Models | TP  | TN  | FP  | FN  | N   | Sensitivity | Specificity | CLA  | RMSE       | K     |
|--------|-----|-----|-----|-----|-----|-------------|-------------|------|------------|-------|
| BPM    | 20  | 20  | 5   | 3   | 48  | 0.870       | 0.800       | 0.833| 0.4550     | 0.667 |
| DSM    | 17  | 22  | 6   | 3   | 48  | 0.850       | 0.786       | 0.813| 0.5970     | 0.622 |
| MLR    | 15  | 22  | 9   | 2   | 48  | 0.882       | 0.710       | 0.771| 0.4960     | 0.542 |

| Landslide susceptibility zones | BMP Area (%) | BPM FR | DSM Area (%) | DSM FR | MLR Area (%) | MLR FR |
|-------------------------------|--------------|--------|--------------|--------|--------------|--------|
| Stable area                   | 15.37        | 0      | 9.95         | 0      | 19.11        | 0.25   |
| Low susceptible               | 24.51        | 0.2    | 24.62        | 0.4    | 28.93        | 0.76   |
| Moderately susceptible        | 17.38        | 1.12   | 23.09        | 0.79   | 26.32        | 1.07   |
| Highly susceptible            | 24.2         | 1.16   | 23.89        | 2.14   | 17.36        | 1.19   |
| Extremely susceptible         | 18.54        | 2.56   | 18.46        | 1.12   | 7.5          | 3.25   |

Figure 7. The ROC-AUC values of the predicted landslide susceptibility models (a) success rate and (b) prediction rate.

Table 5. Model validation using statistical measures.

Table 6. Computed FR values in the susceptibility levels of the predicted models.
The locations of the landslides associated with the 2018 deluge were collected by field check and a total of 49 landslides were used. The landslides in each susceptibility level are presented in Table 7. In the case of BPM, 85.71% of recent landslides occurred in the extreme and highly susceptible classes and only six landslides are reported in moderate landslide area. It is also noted that no landslides are reported in the least susceptible class. Whereas DSM and MLR models are not succeeded in predicting the landslide susceptibility compared to BPM (Table 7).

Finally, to check the statistically significant differences between the three predicted landslide susceptibility models, the Wilcoxon signed-rank test method with p and Z values is used for pairwise comparisons of the three models. When p values are smaller than the 0.05 significance level and Z values exceed critical values of Z (−1.96 to +1.96), the performances of landslide models are different (Bui et al. 2015; Dou et al. 2019). In the present study, the Wilcoxon signed-rank test shows that the performances of the three models are significantly different (Table 8).

4.4. Discussion

The number of landslide occurrences in every year is steadily increasing in the Western Ghats region; therefore, it is important to identify landslide susceptible areas in order to minimize the anthropogenic disturbances on such regions. In this context, this study compares the efficiency of three common approaches to landslide susceptibility mapping by utilizing the available data. The analysis shows that the anthropogenic activities in hilly tracks of the Western Ghats have a significant role in making the land susceptible to landslides. For instance, the agricultural area with slopes between 10° and 40° accounted for 39.65% of the total reported landslides. It is also noted that nearly 10% of previous landslides have occurred in wastelands/open lands (i.e. vegetation-cleared area). Furthermore, 61% of the landslides in the August 2018 deluge in the study area were in the agricultural area where slopes were between 10° and 40°. During the field visits, it is noticed that anthropogenic disturbances such as the destruction of first- and second-order streams and construction of check dams or bunds in the steep slopes for irrigation activities are the major conditions which initiate landslides. Likewise in the case of road networks, 33 past landslides (prior to 2018) occurred with a maximum distance of 300 m from the roads and this susceptibility remains in the recent landslides also (17 recent landslides are found within the distance of 300 m). Therefore, it is likely that further road development and associated disturbances will be a serious threat to slope stability in the study area. It is also noted that ~26% of the previous landslides are occurred >500 m road distances that are mainly located in interior forest areas with little road networks.

A comparative assessment of the weights derived from BPM and DSM shows uncertainty in some of the weights calculated by DSM. For instance, 41.3% of landslides are reported in the deciduous forest category and accordingly, BPM predicted the maximum C value in this class in contrast to the DSM which shows a maximum value (0.65) in the forest plantation category where only two landslides are reported. Further, 45 landslides are reported in denudational structural hills where BPM predicted a maximum contrast value (1.21) for the same category. While DSM shows the maximum Bel value (0.51) in residual hills where only two landslides are reported, may be due to lower areal extent. Further, in the case of flat curvature, DSM predicted a value of 0.66 which is greater than the value of convex (0.60) curvature and comparable to the value of concave curvature (0.73). It is also noticed that only two previous landslides are reported in the flat curvature where BPM shows a negative value (−0.03) to the flat curvature indicating the very low probability of landslide occurrence. These kinds of uncertainty in weighting will be the reason for accuracy differences in the models. Furthermore, in the bivariate models, landslide susceptibility is represented by accumulating the weights derived by equations. Because the weights are accumulated without distinction or grading, each causative factor has the same impact and relationships. This can lead to the overestimation or underestimation of landslide susceptibility. For example, a combination of high relative relief and steep slope leads to high landslide susceptibility,
although these causative factors might be same characteristics and same tendency for occurrence of landslides.

The results of this study are comparable with the other studies conducted in Ghat's region as well as other parts of the world with similar environmental conditions. Krishnan et al. (2015) studied landslide susceptibility of upland catchment of Meenachil river in southern Western Ghats using bivariate Information value method with MLR model and reported bivariate method gave higher accuracy over the MLR. Vijith et al. (2014) used the weight of evidence model to assess the landslide susceptibility of southern Western Ghats with an accuracy of 89.2%. The study concluded that denudational structural hills, with slope in-between 16° and 45° with high relative relief are susceptible for the occurrence of landslides, which is highly correlated with the results of the present study. Different scholars across the world tried to find out efficient bivariate and multivariate models for landslide susceptibility mapping. Pourghasemi et al. (2013) compared the efficiency of BPM and Dempster–Shafer model and reported the superiority of BPM over DSM. In contrast to this result, Mohammady, Pourghasemi, and Pradhan (2012) studied the landslide susceptibility of Golestan Province, Iran using DSM, BPM and FR methods. The results show a maximum accuracy of the FR and DSM and the worst performance of BPM in both prediction and success rate. Devkota et al. (2013) reported the outperformance of bivariate statistical methods over MLR model. However, the foresaid studies indicated their prediction efficiency on the basis of ROC-AUC curve only, which is difficult to interpret or cannot answer the question which landslides are well predicted and which are poorly predicted. Further, the prediction error which is more important than prediction accuracy is absent in the previous studies. Hence, in this study, matrix-based calculations such as sensitivity, specificity and classification accuracy were used to test predictive efficiency of the foresaid models and error estimation is done with RMSE method. Among the proposed methods, it is clear that BPM is the optimum model, achieved highest accuracy in both training and testing phase. It is also noted that least RMSE is reported in BPM predicted landslide susceptibility model compared to other models. As the square root of variance, RMSE can be interpreted as the standard deviation of the unexplained variance and has the useful property of being in the same units as the response variable. Lower values of RMSE indicate better fit and higher RMSE indicates the chances of overprediction or under prediction. Hence, considering sensitivity, specificity, classification accuracy, ROC-AUC, RMSE, kappa index and FR values, the study proposes the use of BPM over DSM and MLR in landslide susceptibility mapping in southern Western Ghats.

Even though the predicted models show reasonable accuracy with statistical significance, the study may have some limitations since the selection of landslide conditioning factors are subjective and assumes the variables are conditionally independent. However, it is felt that the study could be extremely useful for improving land use management in the study area to prevent and mitigate further damage due to landslide hazards.

5. Conclusion

Landslides are catastrophic events which repeatedly occur in the Western Ghats region of India during monsoons and cause widespread damage to life, property and environment. Therefore, areas prone to landslides should be identified to reduce the damage caused by landslides. In the present study, a comparison between two bivariate statistical methods (i.e. DSM and BPM) with a multivariate statistical technique (MLR) is attempted to model the landslide susceptibility zones in the southern Western Ghats. The multi-collinearity test and Pearson correlation coefficient test reveal the absence of multicollinearity among the variables. Based on the model validation using statistical measure and ROC-AUC analysis, BPM appears to be the best method with highest ROC-AUC (0.833), sensitivity (0.870), specificity (0.800), kappa index (0.667) and lowest RMSE of 0.4550. In the present study, highest RMSE (0.5970) is reported by DSM method, indicates the inefficiency of the model in predicting the landslide susceptibility. In addition, in this study, MLR forms the least accurate method with ROC-AUC (0.701), sensitivity (0.882), specificity (0.710), kappa index (0.542) and RMSE of 0.4960. Finally, the FR method which is used to identify the accuracy of each susceptible zone suggests that the BPM is the overall best method. The bivariate models have more than 80% accuracy in predicting the unknown landslides indicating that the models are trustworthy for future land use planning and infrastructure development in the study area.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

A. L. Achu http://orcid.org/0000-0001-6821-2665
References

Akgun, A., C. Kincal, and B. Pradhan. 2012. "Application of Remote Sensing Data and GIS for Landslide Risk Assessment as an Environmental Threat to Izmir City (West Turkey)." Environmental Monitoring and Assessment 184 (9): 5453–5470. doi:10.1007/s10661-011-2352-8.

Althuwaynee, O. F., B. Pradhan, and S. Lee. 2012. "Application of an Evidential Belief Function Model in Landslide Susceptibility Mapping." Computers & Geosciences 44: 120–135. doi:10.1016/j.cageo.2012.03.003.

Arabameri, A., S. Saha, J. Roy, W. Chen, T. Blaschke, and D. Tien Bui. 2020. “Landslide Susceptibility Evaluation and Management Using Different Machine Learning Methods in the Gallicich River Watershed, Iran.” Remote Sensing 12 (3): 475. doi:10.3390/rs12030475.

Atkinson, P. M., and R. Massari. 1998. “Generalised Linear Modelling of Susceptibility to Landsliding in the Central Apennines, Italy.” Computers & Geosciences 24 (4): 373–385. doi:10.1016/S0098-3004(97)00117-9.

Ayalew, L., and H. Yamagishi. 2005. “The Application of GIS-based Logistic Regression for Landslide Susceptibility Mapping in the Kakuda-Yahiko Mountains, Central Japan.” Geomorphology 65 (1–2): 15–31. doi:10.1016/j.geomorph.2004.06.010.

Baum, R. L., W. Z. Savage, and J. W. Godt. 2002. “TRIGRS—A Fortran Program for Transient Rainfall Infiltration and Grid-based Regional Slope-Stability Analysis.” US Geological Survey Open-file Report 424: 38.

Bera, S., B. Guru, and V. Ramesh. 2019. “Evaluation of Landslide Susceptibility Models: A Comparative Study on the Part of Western Ghat Region, India.” Remote Sensing Applications: Society and Environment 13: 39–52. doi:10.1016/j.rsase.2018.10.010.

Bonham-Carter, G. F. 1994. “Geographic Information Systems for Geoscientists-modeling with GIS.” Computer Methods in the Geosciences 13: 398.

Booth, G. D., M. J. Niccolucci, and E. G. Schuster. 1994. Identifying proxy sets in multiple linear regression: an aid to better coefficient interpretation. Research paper INT (USA).

Bragagnolo, L., R. V. da Silva, and J. M. V. Grzybowski. 2020. “Landslide Susceptibility Mapping with R Landslide: A Free Open-source GIS-integrated Tool Based on Artificial Neural Networks.” Environmental Modelling and Software 123: 104565. doi:10.1016/j.envsoft.2019.104565.

Bui, D. T., T. A. Tuan, H. Klempe, B. Pradhan, and I. Revhaug. 2015. “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. doi:10.1007/s10346-015-0557-6.

Can, A., G. Dagdelenler, M. Ercanoğlu, and H. Sonmez. 2019. “Landslide Susceptibility Mapping at Ovacık-Karabük (Turkey) Using Different Artificial Neural Network Models: Comparison of Training Algorithms.” Bulletin of Engineering Geology and the Environment 78 (1): 89–102. doi:10.1007/s10064-017-1034-3.

Chen, T., L. Zhu, R. Q. Niu, C. J. Trinder, L. Peng, and T. Lei. 2020a. “Mapping Landslide Susceptibility at the Three Gorges Reservoir, China, Using Gradient Boosting Decision Tree, Random Forest and Information Value Models.” Journal of Mountain Science 17 (3): 670–685. doi:10.1007/s11629-019-5839-3.

Chen, W., W. Li, E. Hou, H. Bai, H. Chai, D. Wang, X. Cui, and Q. Wang. 2015. “Application of Frequency Ratio, Statistical Index, and Index of Entropy Models and Their Comparison in Landslide Susceptibility Mapping for the Baozhong Region of Baoji, China.” Arabian Journal of Geosciences 8 (4): 1829–1841. doi:10.1007/s12517-014-1554-0.

Chen, W., X. Xie, J. Peng, H. Shahabi, H. Hong, D. T. Bui, Z. Duan, S. Li, and A. X. Zhu. 2018. “GIS-based Landslide Susceptibility Evaluation Using a Novel Hybrid Integration Approach of Bivariate Statistical Based Random Forest Method.” Catena 164: 135–149. doi:10.1016/j.catena.2018.01.012.

Chen, Z., F. Ye, W. Fu, Y. Ke, and H. Hong. 2020b. “The Influence of DEM Spatial Resolution on Landslide Susceptibility Mapping in the Baxie River Basin, NW China.” Natural Hazards 1–25. doi:10.1007/s11069-020-03899-9.

Dempster, A. P. 1967. “Upper and Lower Probabilities Induced by a Multivalued Mapping.” The Annals of Mathematical Statistics 38 (2): 325–339. doi:10.1214/aoms/1177698950.

Devkota, K. C., A. D. Regmi, H. R. Pourghasemi, K. Yoshida, B. Pradhan, I. C. Ryu, M. R. Dhitol, and O. F. Althuwaynee. 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.” Natural Hazards 65 (1): 135–165. doi:10.1007/s11069-012-0347-6.

Ding, Q., W. Chen, and H. Hong. 2016. “Application of Frequency Ratio, Weights of Evidence and Evidential Belief Function Models in Landslide Susceptibility Mapping.” Geocarto International 32 (6): 619–639. doi:10.1080/10106049.2016.1165294.

Dou, J., A. P. Yunus, D. T. Bui, A. Merghadi, M. Sahana, Z. Zhu, C. W. Chen, Z. Han, and B. T. Pham. 2019. “Improved Landslide Assessment Using Support Vector Machine with Bagging, Boosting, and Stacking Ensemble Machine Learning Framework in a Mountainous Watershed, Japan.” Landslides 17: 1–18.

Fan, J., D. Jiang, W. Liu, F. Wu, J. Chen, and J. J. K. Daemen. 2019. “Discontinuous Fatigue of Salt Rock with Low-stress Intervals.” International Journal of Rock Mechanics and Mining Sciences 115: 77–86. doi:10.1016/j.ijrmms.2019.01.013.

Fang, Z., Y. Wang, L. Peng, and H. Hong. 2020. “Integration of Convolutional Neural Network and Conventional Machine Learning Classifiers for Landslide Susceptibility Mapping.” Computers & Geosciences 104470. doi:10.1016/j.cageo.2020.104470.

Formetta, G., V. Rago, G. Capparelli, R. Rigon, F. Muto, and P. Versace. 2014. “Integrated Physically Based System for Modeling Landslide Susceptibility.” Procedia Earth and Planetary Science 9: 74–82. doi:10.1016/j.proeps.2014.06.006.

Froude, M. J., and D. N. Petley. 2018. “Global Fatal Landslide Occurrence from 2004 to 2016.” Natural Hazards & Earth System Sciences 18 (8): 2161–2181. doi:10.5194/nhess-18-2161-2018.

Gadtaula, A., and S. Dhakal. 2019. “Landslide Susceptibility Mapping Using Weight of Evidence Method in Haku, Rasuwa District, Nepal.” Journal of Nepal Geological Society 58: 163–171. doi:10.3126/jngs.v58i0.24601.
Ghasemain, B., D. T. Asl, B. T. Pham, M. Avand, H. D. Nguyen, and S. Janizadeh. 2020. “Shallow Landslide Susceptibility Mapping: A Comparison between Classification and Regression Tree and Reduced Error Pruning Tree Algorithms.” *Vietnam Journal of Earth Sciences*.

Ghorbanzadeh, O., T. Blaschke, J. Aryal, and K. Gholamnia. 2018. “A New GIS-based Technique Using an Adaptive Neuro-fuzzy Inference System for Land Subsidence Susceptibility Mapping.” *Journal of Spatial Science* 1–17. doi:10.1080/14498596.2018.1505564.

Gorum, T., and E. J. M. Carranza. 2015. “Control of Style-of-faulting on Spatial Pattern of Earthquake-triggered Landslides.” *International Journal of Environmental Science and Technology* 12 (10): 3189–3212. doi:10.1007/s13762-015-0752-y.

Guzzetti, F., A. Carrara, M. Cardinali, and P. Reichenbach. 1999. “Landslide Hazard Evaluation: A Review of Current Techniques and Their Application in a Multi-scale Study, Central Italy.” *Geomorphology* 31 (1–4): 181–216. doi:10.1016/S0169-555X(99)00078-1.

Harmouzi, H., H. A. Nefeslioglu, M. Rouai, E. A. Sezer, A. Dekayir, and C. Gökceoğlu. 2019. “Landslide Susceptibility Mapping of the Mediterranean Coastal Zone of Morocco between Oued Laou and El Jebha Using Artificial Neural Networks (ANN).” *Arabian Journal of Geosciences* 12 (22): 696. doi:10.1007/s12517-019-4922-3.

Huang, F., J. Zhang, C. Zhou, Y. Wang, J. Huang, and L. Zhu. 2020. “A Deep Learning Algorithm Using a Fully Connected Sparse Autoencoder Neural Network for Landslide Susceptibility Prediction.” *Landslides* 17 (1): 217–229. doi:10.1007/s10346-019-01274-9.

Ilia, I., and P. Tsangaratos. 2016. “Applying Weight of Evidence Method and Sensitivity Analysis to Produce a Landslide Susceptibility Map.” *Landslides* 13 (2): 379–397. doi:10.1007/s10346-015-0576-3.

Juliev, M., M. Mergili, I. Mondal, B. Nurtavaev, A. Pulatov, and J. Hübl. 2019. “Comparative Analysis of Statistical Methods for Landslide Susceptibility Mapping in the Bostanlik District, Uzbekistan.” *Science of the Total Environment* 653: 801–814. doi:10.1016/j.scitotenv.2018.10.431.

Kayastha, P., M. R. Dhital, and F. De Smedt. 2013. “Application of the Analytical Hierarchy Process (AHP) for Landslide Susceptibility Mapping: A Case Study from the Tinau Watershed, West Nepal.” *Computers & Geosciences* 52: 398–408. doi:10.1016/j.cageo.2012.11.003.

Krishnan, M. V. N., P. Pratheesh, P. G. Rejith, and H. Vijith. 2015. “Determining the Suitability of Two Different Statistical Techniques in Shallow Landslide (Debris Flow) Initiation Susceptibility Assessment in the Western Ghats.” *Environmental Research, Engineering and Management* 70 (4): 26–39.

Kuriakose, S. L., G. Sankar, and C. Muraleedharan. 2009. “History of Landslide Susceptibility and a Chorology of Landslide-prone Areas in the Western Ghats of Kerala, India.” *Environmental Geology* 57 (7): 1553–1568. doi:10.1007/s00254-008-1431-9.

Landis, J. R., and G. G. Koch. 1977. “The Measurement of Observer Agreement for Categorical Data.” *biometrics* 33 (1): 159–174. doi:10.2307/2529310.

Li, Y., and W. Chen. 2020. “Landslide Susceptibility Evaluation Using Hybrid Integration of Evidential Belief Function and Machine Learning Techniques.” *Water* 12 (1): 113. doi:10.3390/w12010113.

Liu, W., Z. Zhang, J. Chen, J. Fan, D. Jiang, D. Jjk, and Y. Li. 2019. “Physical Simulation of Construction and Control of Two Butted-well Horizontal Cavern Energy Storage Using Large Molded Rock Salt Specimens.” *Energy* 185: 682–694. doi:10.1016/j.energy.2019.07.014.

Mohammady, M., H. R. Pourghasemi, and B. Pradhan. 2012. “Landslide Susceptibility Mapping at Golestan Province, Iran: A Comparison between Frequency Ratio, Dempster–Shafer, and Weights-of-evidence Models.” *Journal of Asian Earth Sciences* 61: 221–236. doi:10.1016/j.jseaes.2012.10.005.

Mondal, S., and S. Mandal. 2018. “RS & GIS-based Landslide Susceptibility Mapping of the Balason River Basin, Darjeeling Himalaya, Using Logistic Regression (MLR) Model.” *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards* 12 (1): 29–44. doi:10.1080/17499518.2017.1347949.

Nampak, H., B. Pradhan, and M. A. Manap. 2014. “Application of GIS Based Data Driven Evidential Belief Function Model to Predict Groundwater Potential Zonation.” *Journal of Hydrology* 513: 283–300. doi:10.1016/j.jhydrol.2014.02.053.

Nhu, V. H., N. D. Hoang, H. Nguyen, P. T. T. Ngo, T. T. Bui, P. V. Hoa, P. Samui, and D. T. Bui. 2020. “Effectiveness Assessment of Keras Based Deep Learning with Different Robust Optimization Algorithms for Shallow Landslide Susceptibility Mapping at Tropical Area.” *Catena* 188: 104458. doi:10.1016/j.catena.2020.104458.

Ozdemir, A., and T. Altural. 2013. “A Comparative Study of Frequency Ratio, Weights of Evidence and Logistic Regression Methods for Landslide Susceptibility Mapping: Sultan Mountains, SW Turkey.” *Journal of Asian Earth Sciences* 64: 180–197. doi:10.1016/j.jseaes.2012.12.014.

Pandey, V. K., H. R. Pourghasemi, and M. C. Sharma. 2020. “Landslide Susceptibility Mapping Using Maximum Entropy and Support Vector Machine Models along the Highway Corridor, Garhwal Himalaya.” *Geocarto International* 35 (2): 168–187. doi:10.1080/10106049.2018.1510038.

Pham, B. T., I. Prakash, and D. T. Bui. 2018. “Spatial Prediction of Landslides Using a Hybrid Machine Learning Approach Based on Random Subspace and Classification and Regression Trees.” *Geomorphology* 303: 256–270. doi:10.1016/j.geomorph.2017.12.008.

Pourghasemi, H., B. Pradhan, C. Gokceoglu, and K. D. Moezzi. 2013. “A Comparative Assessment of Prediction Capabilities of Dempster–Shafer and Weights-of-evidence Models in Landslide Susceptibility Mapping Using GIS.” *Geomatics, Natural Hazards and Risk* 4 (2): 93–118. doi:10.1080/19475705.2012.662915.

Pourghasemi, H. R., M. Mohammady, and B. Pradhan. 2012. “Landslide Susceptibility Mapping Using Index of Entropy and Conditional Probability Models in GIS: Safarood Basin, Iran.” *Catena* 97: 71–84. doi:10.1016/j.catena.2012.05.005.

Pourghasemi, H. R., and O. Rahmati. 2018. “Prediction of the Landslide Susceptibility: Which Algorithm, Which Precision?” *Catena* 162: 177–192. doi:10.1016/j.catena.2017.11.022.

Pradhan, A. M. S., and Y. T. Kim. 2017. “Spatial Data Analysis and Application of Evidential Belief Functions to Shallow Landslide Susceptibility Mapping at Mt. Umyeon, Seoul, Korea.” *Bulletin of Engineering Geology and the Environment* 76 (4): 1263–1279. doi:10.1007/s10064-016-0919-x.
Pradhan, B., R. P. Singh, and M. F. Buchroithner. 2006. "Estimation of Stress and Its Use in Evaluation of Landslide Prone Regions Using Remote Sensing Data." Advances in Space Research 37 (4): 698–709. doi:10.1016/j.asr.2005.03.137.

Pradhan, B., and A. M. Youssef. 2010. "Manifestation of Remote Sensing Data and GIS on Landslide Hazard Analysis Using Spatial-based Statistical Models." Arabian Journal of Geosciences 3 (3): 319–326. doi:10.1007/s12517-009-0089-2.

Prasannakumar, V., and H. Vijith. 2012. "Evaluation and Validation of Landslide Spatial Susceptibility in the Western Ghats of Kerala, through GIS-based Weights of Evidence Model and Area under Curve Technique." Journal of the Geological Society of India 80 (4): 515–523. doi:10.1007/s12665-012-0217-3.

Qiao, W., and Z. Yang. 2019. “Solving Large-scale Function Optimization Problem by Using a New Metaheuristic Algorithm Based on Quantum Dolphin Swarm Algorithm." IEEE Access 7: 138972–138989. doi:10.1109/ACCESS.2019.2942169.

Regmi, N. R., J. R. Giardino, and J. D. Vitek. 2010. "Modeling Susceptibility to Landslides Using the Weight of Evidence Approach: Western Colorado, USA." Geomorphology 115 (1–2): 172–187. doi:10.1016/j.geomorph.2009.10.002.

Roodposhti, M. S., S. Rahimi, and M. J. Beglou. 2014. “PROMETHEE II and Fuzzy AHP: An Enhanced GIS-based Landslide Susceptibility Mapping." Natural Hazards 73 (1): 77–95. doi:10.1007/s11069-012-0523-8.

Sahin, E. K., I. Colkesen, and T. Kavzoglu. 2020. “A Comparative Assessment of Canonical Correlation Forest, Random Forest, Rotation Forest and Logistic Regression Methods for Landslide Susceptibility Mapping." Geocarto International 35 (4): 341–363. doi:10.1080/10106049.2018.1516248.

Sajinkumar, K. S., S. Anbazhagan, A. P. Pradeepkumar, and V. R. Rani. 2011. “Weathering and Landslide Occurrences in Parts of Western Ghats, Kerala.” Journal of the Geological Society of India 78 (3): 249. doi:10.1007/s12594-011-0089-1.

Sameen, M. I., B. Pradhan, and S. Lee. 2020. “Application of Convolutional Neural Networks Featuring Bayesian Optimization for Landslide Susceptibility Assessment.” Catena 186: 104249. doi:10.1016/j.catena.2019.104249.

Shafer, G. 1976. A Mathematical Theory of Evidence. Vol. 42. Princeton University Press.

Soman, K. 1987. “Geology of Kerala.” GSI Publications 2 (1).

Stocking, M. A. 1972. “Relief Analysis and Soil Erosion in Rhodesia Using Multivariate Techniques.” Zeitschrift fur Geomorphologie NF 16: 432–443.

Surya, P. 2012. Training Module on Comprehensive Landslides Risk Management. National Institute of Disaster Management. New Delhi-110002, 282.

Thampi, P. K., J. Mathai, G. Sankar, and S. Sidhathran. 1997. Evaluation study in terms of landslide mitigation in parts of Western Ghats, Kerala. Technical report. Trivandrum: Center for Earth Science Studies.

Van Dao, D., A. Jaafari, M. Bayat, D. Mafi-Gholami, C. Qi, H. Moayedi, T. Van Phong, et al. 2020. “A Spatially Explicit Deep Learning Neural Network Model for the Prediction of Landslide Susceptibility.” Catena 188: 104451. doi:10.1016/j.catena.2019.104451.

Van Westen, C. J. 1994. “GIS in Landslide Hazard Zonation: A Review, with Examples from the Andes of Colombia.” In Mountain Environments & Geographic Information Systems, 135–166. Taylor & Francis.

Van Westen, C. J., E. Castellanos, and S. L. Kuriakose. 2008. “Spatial Data for Landslide Susceptibility, Hazard, and Vulnerability Assessment: An Overview.” Engineering Geology 102 (3–4): 112–131. doi:10.1016/j.enggeo.2008.03.010.

Van Westen, C. J., N. Rengers, and R. Soeters. 2003. “Use of Geomorphological Information in Indirect Landslide Susceptibility Assessment.” Natural Hazards 30 (3): 399–419. doi:10.1023/B:NHAZ.000007097.42735.9e.

Vijith, H., K. N. Krishnakumar, G. S. Pradeep, M. V. Ninu Krishnan, and G. Madhu. 2014. “Shallow Landslide Initiation Susceptibility Mapping by GIS-based Weights-of-evidence Analysis of Multi-class Spatial Data-sets: A Case Study from the Natural Sloping Terrain of Western Ghats, India.” Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards 8 (1): 48–62. doi:10.1080/17499518.2013.843437.

Vijith, H., and G. Madhu. 2008. “Estimating Potential Landslide Sites of an Upland Sub-watershed in Western Ghat’s of Kerala (India) through Frequency Ratio and GIS.” Environmental Geology 55 (7): 1397–1405. doi:10.1007/s00254-007-1090-2.

Walley, P. 1987. “Belief Function Representations of Statistical Evidence.” The Annals of Statistics 15 (4): 1439–1465. doi:10.1214/aos/1176350603.

Zhang, Z., D. Jiang, W. Liu, J. Chen, E. Li, J. Fan, and K. Xie. 2019. “Study on the Mechanism of Roof Collapse and Leakage of Horizontal Cavern in Thinly Bedded Salt Rocks.” Environmental Earth Sciences 78 (10): 292. doi:10.1007/s12665-019-8292-2.

Zhu, A. X., R. Wang, J. Qiao, C. Z. Qin, Y. Chen, J. Liu, F. Du, Y. Lin, and T. Zhu. 2014. “An Expert Knowledge-based Approach to Landslide Susceptibility Mapping Using GIS and Fuzzy Logic.” Geomorphology 214: 128–138. doi:10.1016/j.geomorph.2014.02.003.