ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/304742376

A Comparative Study of Different Machine Learning Methods for Landslide Susceptibility Assessment: A Case Study of Uttarakh....

Article *in* Environmental Modelling and Software · July 2016

DOI: 10.1016/j.envsoft.2016.07.005

CITATIONS
44

READS
656

5 authors, including:

Binh Thai Pham
University Of Transport Technology
40 PUBLICATIONS **344** CITATIONS

SEE PROFILE

Biswajeet Pradhan
University of Technology Sydney
492 PUBLICATIONS **11,967** CITATIONS

SEE PROFILE

Dieu Tien Bui
University College of Southeast Norway
104 PUBLICATIONS **1,855** CITATIONS

SEE PROFILE

Indra Prakash
Bhaskarachrya Institue for Space Application...
121 PUBLICATIONS **284** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project A geospatial model for the optimization grazing management in semi-arid rangeland of Iran View project

Project Geotechnical Problems View project

All content following this page was uploaded by Binh Thai Pham on 08 October 2017.

The user has requested enhancement of the downloaded file.

Environmental Modelling & Software 84 (2016) 240–250

Contents lists available at ScienceDirect

Environmental Modelling & Software

journal homepage: www.elsevier.com/locate/envsoft

A comparative study of different machine learning methods for landslide susceptibility assessment: A case study of Uttarakhand area (India)

Binh Thai Pham [a, b, *], Biswajeet Pradhan [c], Dieu Tien Bui [d], Indra Prakash [e], M.B. Dholakia [f]

[a] *Department of Civil Engineering, Gujarat Technological University, Nr.Visat Three Roads, Visat, Gandhinagar Highway, Chandkheda, Ahmedabad, 382424, Gujarat, India*
[b] *Department of Geotechnical Engineering, University of Transport Technology, 54 Trieu Khuc, Thanh Xuan, Ha Noi, Viet Nam*
[c] *Faculty of Engineering, Geospatial Information Science Research Centre (GIS RC), Department of Civil Engineering, University Putra Malaysia, 43400, Serdang, Selangor Darul Ehsan, Malaysia*
[d] *Geographic Information System Group, Department of Business Administration and Computer Science, Telemark University College, Hallvard Eikas Plass 1, N-3800, Bø i Telemark, Norway*
[e] *Department of Science & Technology, Bhaskarcharya Institute for Space Applications and Geo-Informatics (BISAG), Government of Gujarat, Gandhinagar, India*
[f] *Department of Civil Engineering, LDCE, Gujarat Technological University, Ahmedabad, 380015, Gujarat, India*

ARTICLE INFO

Article history:
Received 20 March 2016
Received in revised form
23 May 2016
Accepted 2 July 2016

Keywords:
Landslides susceptibility assessment
Machine learning
Uttarakhand
India

ABSTRACT

Landslide susceptibility assessment of Uttarakhand area of India has been done by applying five machine learning methods namely Support Vector Machines (SVM), Logistic Regression (LR), Fisher's Linear Discriminant Analysis (FLDA), Bayesian Network (BN), and Naïve Bayes (NB). Performance of these methods has been evaluated using the ROC curve and statistical index based methods. Analysis and comparison of the results show that all five landslide models performed well for landslide susceptibility assessment (AUC = 0.910–0.950). However, it has been observed that the SVM model (AUC = 0.950) has the best performance in comparison to other landslide models, followed by the LR model (AUC = 0.922), the FLDA model (AUC = 0.921), the BN model (AUC = 0.915), and the NB model (AUC = 0.910), respectively.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

Landslide is one of the most serious geo-hazards causing the loss of life and property all over the world, therefore, landslide susceptibility assessment and mitigation of its harmful impactions has been turning into urgent tasks to government and non-government agencies (Althuwaynee et al., 2012). Landslide susceptibility map is known as a useful tool for landslide hazard management through land use planning and better decision making in landslide prone areas (Akgun, 2012). Machine learning approaches are considered more efficient than other approaches such as expert's opinion based methods and analytic methods for spatial prediction of landslides (Pradhan, 2013). Main principle of these approaches is that landslide susceptibility is assessed using machine learning algorithms to analyze the spatial relationship between past landslide events and a set of conditioning factors from which the potential probability of landslide occurrence is determined (Chen et al., 2015; Guzzetti, 2006).

Many types of machine learning algorithms have been developed and applied for producing landslide susceptibility maps in many regions of the world. Zare et al. (2013), Pradhan and Lee (2010b), and Conforti et al. (2014) utilized artificial neural networks which are based on the biological neural networks to predict spatially landslide distributions. Whereas, Tien Bui et al. (2012b), Lee et al. (2013), and Jebur et al. (2015) applied evidential belief functions to generate landslide susceptibility maps. On the other hand, fuzzy logic algorithms have also been employed to assess the spatial distribution of landslides. In addition, other algorithms such as Support Vector Machines (SVM) (Vapnik, 1995), Logistic Regression (LR) (Cabrera, 1994), Fisher's Linear Discriminant

* Corresponding author. Department of Civil Engineering, Gujarat Technological University, Nr.Visat Three Roads, Visat, Gandhinagar Highway, Chandkheda, Ahmedabad, 382424, Gujarat, India.
E-mail addresses: phambinhgtvt@gmail.com, binhpt@utt.edu.vn (B.T. Pham).

http://dx.doi.org/10.1016/j.envsoft.2016.07.005
1364-8152/© 2016 Elsevier Ltd. All rights reserved.

Analysis (FLDA) (Fisher, 1936), Bayesian Network (BN) (Friedman et al., 1997), and Naïve Bayes (NB) (Soria et al., 2011) are known to be applicable for binary classification problems that could be used for landslide susceptibility assessment.

Literature review shows that SVM has been applied efficiently and widely in landslide prediction (Kavzoglu et al., 2014; Peng et al., 2014; Pourghasemi et al., 2013). Marjanović et al. (2011) has carried out the comparison study of SVM with other methods, and concluded that SVM outperforms decision tree, logistic regression, and analytical hierarchy process in producing landslide susceptibility map. In another landslide study, Kavzoglu et al. (2014) also stated that the performance of SVM is better than the conventional logistic regression. In addition, Tien Bui et al. (2012a) reported that results derived from SVM can produce better landslide susceptibility map compared to decision tree, and naïve bayes methods.

LR has also been applied widely in landslide studies (Van Den Eeckhaut et al., 2006a; Yesilnacar and Topal, 2005). Das et al. (2010) proved that LR is a promising method for spatial prediction of landslides. Likewise, Akgun (2012) stated that LR has the best performance compared with other methods namely likelihood ratio, and multi-criteria decision analysis for landslide susceptibility map. In another comparison study, LR outperforms artificial neural network and likelihood ratio methods for landslide susceptibility analysis (Lee et al., 2007).

FLDA is one of the most used methods for complex data classification because it is simple enough to be tractable to itemized formal analysis in the projected area (Durrant and Kabán, 2010). FLDA has been applied as an efficient classifier in many fields such as pattern recognition (Cooke, 2002), medicine (Ambroise and McLachlan, 2002; Asamoah-Boaheng, 2014; Dudoit et al., 2002). For landslide problems, FLDA has been applied in very few studies (Murillo-García and Alcántara-Ayala, 2015; Rossi et al., 2010).

BN is considered as a promising method for hazard assessment (Liang et al., 2012). However, it is still rarely applied for the assessment of landslide hazard. Song et al. (2012a) used this method to assess susceptibility of earthquake-induced landslides, and stated that it shows high probability of landslide detection capability and it is a good alternative tool for landslide prediction. Liang et al. (2012) also stated that BN is useful for assessment of debris flow hazard.

NB method has also been applied successfully in some of the landslide assessment studies (Tien Bui et al., 2012a; Venkatesan et al., 2013). Pham et al. (2015b) applied this method for spatial prediction of landslides, and stated that it is an efficient machine learning method for landslide susceptibility assessment.

Overall, each above method has been applied successfully and efficiently for solving many real problems in many individual studies. Out of these methods, SVM and LR have been applied widely for landslide susceptibility assessment. However, FLDA, BN, and NB methods have been applied rarely for landslide prediction. Moreover, their performance has not also been compared in literature. Therefore, main objective of the present study is to evaluate and compare the performance of these machine learning methods for landslide susceptibility assessment. For this, part of Uttarakhand area of India which is prone to frequent landslides has been selected for landslide modeling. In this study, Receiver Operating Characteristic (ROC) curve and statistic index-based evaluations methods have been adopted to evaluate and compare landslide models.

2. Description of the study area

The study area is located between longitudes 78°37′40′E to 79°00′50″E and latitudes 30°23′15″N to 30°03′58″N in part of the Uttarakhand state of India, covering about 1325 km^2 (Fig 1). This area is situated in a subtropical monsoon region having annual average rainfall of 600 mm. Heavy rainfalls usually take place in monsoon season (June to September). Temperature during the year ranges from below 0 °C during winter to 45 °C during summer. Mean relative humidity varies from 25% to 85%.

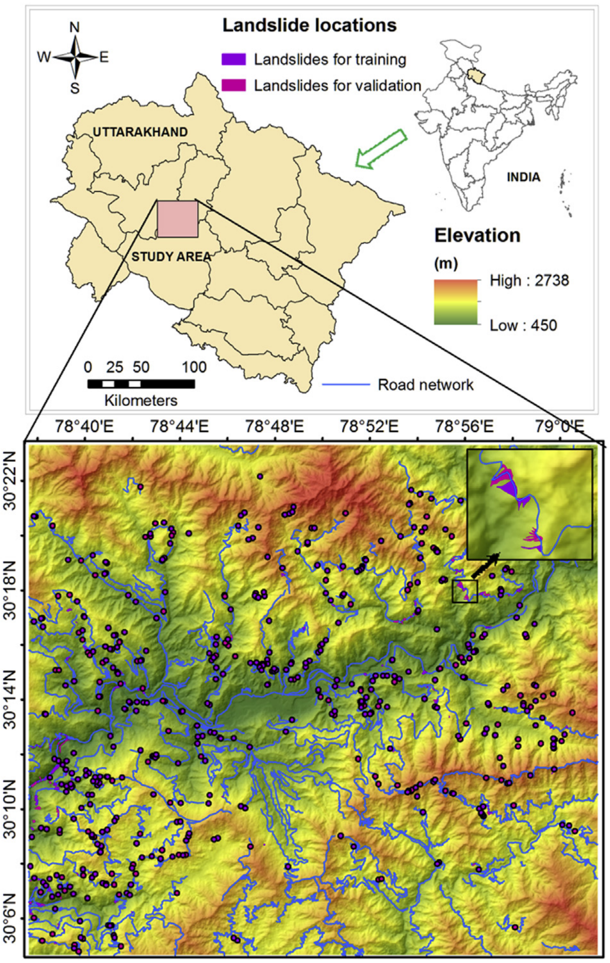

Fig. 1. Location of landslides in the study area.

Topographically, the study area belongs to mountainous region with intervening valleys. Elevation in the area varies from 450 m to 2738 m above sea level. Slope in this area is steep having slope angles up to 70°. Geologically, the area is structurally complex having folded igneous, sedimentary and metamorphic rocks (Agarwal and Kumar, 1973). Metamorphic rocks of Jaunsar group (phyllite and quartzite) are occupying major part of the area. Predominant soil types in the area are silty and loamy.

3. Methodology

Landslide susceptibility analysis in the present study has been carried out in six main steps (Fig 2): (1) preparing the geospatial database, (2) using feature selection method to select suitable landslide affecting factors for landslide analysis, (3) preparing training and testing data sets, (4) constructing landslide models, (5) validating and comparing landslide models, (6) developing landslide susceptibility map (LSM).

3.1. Preparation of geospatial database

Landslide inventory contains very important and indispensable

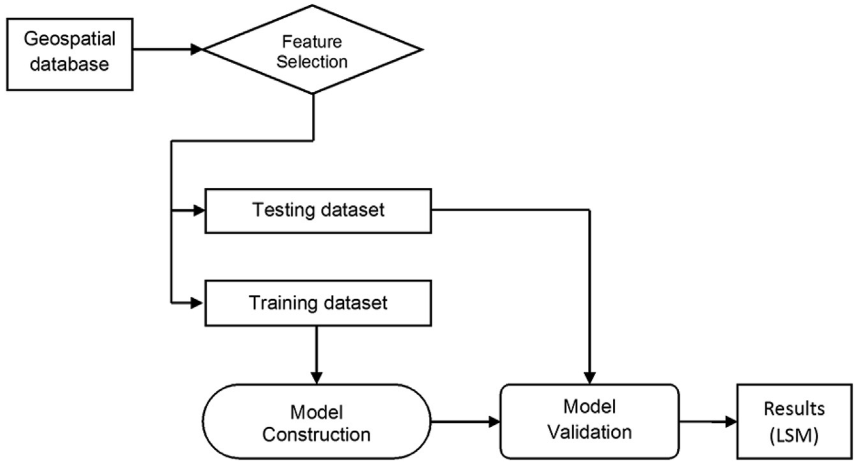

Fig. 2. Methodology adopted in the present study.

data for spatial prediction of landslides (Pham et al., 2016b). In the present study, landslide polygons have been identified considering morphology and texture of past landslides with the help of Google Earth images and tools in Google Earth pro 7.0 software (Pham et al., 2015b). These landslide locations have been validated from historical landslide reports and maps. In general, there are three main types of landslides present in the study area. These are translational (730 locations), rotational (70 locations), and debris flows (130 locations). Each type of landslide has different mechanism requiring separate studies for spatial prediction of landslides (Guo et al., 2005). Therefore, in the present study, only translational landslides have been used for landslide modeling analysis. Out of total 730 translational landslides identified, 200 locations with area larger than 20×20 m^2 (corresponding to size of a pixel in DEM) have been shown as polygons, and other 530 locations are smaller than pixel size have been shown as points (Fig 1).

For landslide prediction, the spatial relationship between historical landslide events and geo-environmental factors must be carried out (Pham et al., 2015a, 2016b). Based on the mechanism of landslide occurrences and characteristics of geo-environment of the study area, a set of fifteen landslide conditioning factors (slope angle, slope aspect, elevation curvature, plan curvature, profile curvature, soil type, land cover, annual rainfall, distance to lineaments, distance to roads, distance to rivers, lineament density, road density, and river density) have been taken into account for initial landslide analysis. The spatial distribution of metamorphic rocks in the area is almost uniform; therefore, lithology is not considered to predict spatially landslides in the present study.

Maps of landslide affecting factors have been constructed using available data of the study area. Specifically, geomorphological factors namely slope angle (Fig 3), slope aspect, elevation curvature, plan curvature, profile curvature have been extracted from Digital Elevation Model (DEM) with 20×20 m grid size that has been constructed from ASTER Global DEM collected from United States Geological Survey (http://earthexplorer.usgs.gov). Maps of soil type (Fig 3) and land cover have been extracted from state soil map and state land cover map at 1,000,000 scale collected from the Water for Welfare site (http://www.ahec.org.in/wfw/maps.htm). Annual rainfall map has been constructed using 30 years meteorological data (1984–2013) from Global Weather data for SWAT (NCEP, 2014). Road network has been digitalized from Google Earth images and co-registered with other maps; and then distance to road map has been constructed by buffering road sections on steep slopes in the study area. River network has been generated from DEM, and then distance to river map has been constructed by buffering river sections on steep slopes in the study area. Lineaments have been extracted from Landsat 8 OLI imagery using Geomatica software; and then distance to lineaments map has been constructed by buffering these lineaments in the study area. Maps of road density, river density, and lineament density have been constructed using kernel density function and "Quantile" classification method (Brewer, 2006). All these maps have been constructed as the raster format with the size of 20×20 m for analysing.

Landslide affecting factors have been classified into different classes (Table 1). This classification is based on the analysis of susceptibility of each factor class to landslide occurrences carried out by Pham et al. (2015b) in adjacent area and also in other landslide studies (Dai and Lee, 2002; Varnes, 1984).

3.2. Model construction using machine learning algorithms

3.2.1. Support Vector Machines (SVM)

SVM was first proposed by Vapnik (1995). It is based on the statistical approach in order to find an optimal hyper-plane for separating two classes (Kavzoglu et al., 2014). Let $X = (X_1, X_2, ..., X_n)$ is the vector of fifteen landslide affecting factors, and $Y_j = (Y_1, Y_2)$ is the vector of classified variables (landslide and non-landslide). The optimal separating hyper-plane can be obtained by solving the classification function as below:

$$f(X) = sign\left[\sum_{i=1}^{n} \alpha_i Y_j k(X, X_i) + c\right] \tag{1}$$

where c is the offset from the origin of the hyper-plane, n is the number of landslide affecting factors, α_i are positive real constants,$k(X,X_i)$ is the kernel function that can be linear, polynomial, radial basis function, or sigmoid (Dixon and Candade, 2008; Suykens and Vandewalle, 1999). For binary classification (landslide, non-landslide), the condition for solving Equation (1) is assumed as below:

$$Y_j\left[\omega^T\varphi(X_i)+c\right] \geq 1 \Leftrightarrow \begin{cases} \omega^T\varphi(X_i)+c \geq 1, & if\ Y_j=+1(landslide) \\ \omega^T\varphi(X_i)+c \leq 1, & if\ Y_j=-1(non-landslide) \end{cases} \tag{2}$$

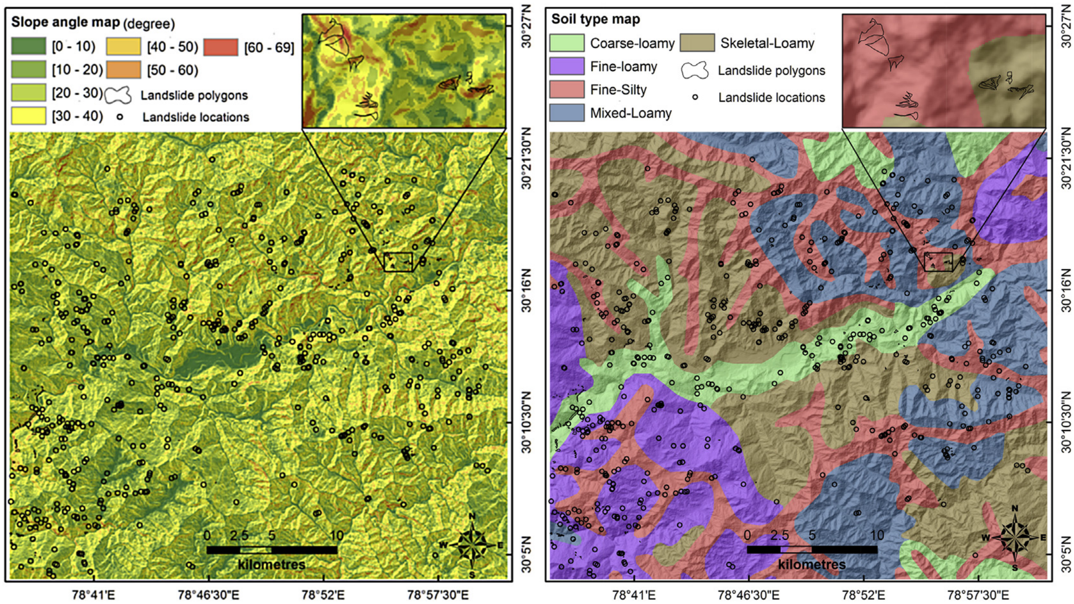

Fig. 3. Slope angle map and soil type map of the study area.

Table 1
Landside affecting factors and their classes.

No.	Landslide affecting factors	Class
1	Slope angle (degree)	(1) [0,10); (2) [10,20); (3)[20,30); (4) [30,40); (5) [40,50); (6) [50,60); (7) [60,69]
2	Slope aspect	1) flat; (2) north; (3) northeast; (4) east; (5) southeast; (6) south; (7) southwest; (8) west; (9) northwest
3	Elevation (m)	(1) [450,700); (2) [700,900); (3) [900,1100); (4) [1100,1300); (5) [1300,1500); (6) [1500,1700); (7) [1700,1900); (8) [1900,2100); (9) [2100,2300); (10) [2300, 2738]
4	Curvature	(1) high concave [−18.25,−2); (2) concave [−2,−0.05); (3) flat [−0.05,0.05); (4) convex [0.05−2); (5) high convex [2,24.25]
5	Plan curvature	(1) [−8.559,−1.078); (2) [−1.078,−0.344); (3) [−0.3444,0.257); (4) [0.257,0.992); (5) [0.992−8.472]
6	Profile curvature	(1) [−16.25,−1.385); (2) [−1.385,−0.433); (3) [−0.433,0.280); (4) [0.280,1.232); (5) [1.232−13.956]
7	Soil type	(1) coarse-loamy; (2) fine-loamy; (3) skeletal-loamy; (4) mixed-loamy; (5) fine-silt
8	Land cover	(1) dense forest; (2) open forest; (3) scrub land; (4) arable land
9	Annual rainfall (mm)	(1) [200,300); (2) [300,400); (3) [400,500); (4) [500,600); (5) [600,700); (6) [700,800); (7) [800,1002]
10	Distance to lineaments (m)	(1) [0,100); (2) [100,200); (3) [200,300); (4) [300,400); (5) [400,500); (6) [500,600); (7) [600,700); (8) $\geq$ 700
11	Distance to roads (m)	(1) [0,50); (2) [50,100); (3) [100,150); (4) [150,200); (5) [200,250); (6) $\geq$ 250
12	Distance to rivers (m)	(1) [0,50); (2) [50,100); (3) [100,150); (4) [150,200); (5) [200,250); (6) $\geq$ 250
13	Lineament density ($km/km^{2)}$)	(1) very low [0,0.673); (2) low [0.673,0.939); (3) moderate [0.939,1.184); (4) high [1.184,1.469); (5) very high [1.469,2.602]
14	Road density ($km/km^{2)}$)	(1) very low [0,0.05); (2) Low [0.05,0.154); (3) moderate [0.154,0.268); (4) high [0.268,0.423); (5) very high [0.423,1.268]
15	River density ($km/km^{2)}$)	(1) very low [0,0.086); (2) low [0.086,0.589); (3) moderate [0.589,1.049); (4) high [1.049,1.623); (5) very high [1.623,3.663]

where $\varphi(X_i)$ is a nonlinear function that divides the input space into higher dimension space, ω represents the weight vector.

3.2.2. Logistic regression (LR)

LR was introduced in the late 1960s and early 1970s (Cabrera, 1994). It is based on the central mathematical concept of the logit-natural logarithm (Peng et al., 2002). LR is well suited in describing and validating the relationship between categorical outcome variables (landslide or non-landslide) and categorical or continuous predictor variables (landslide affecting factors). In the case of landslide prediction, the main purpose of LR is to find the best fitting algorithm to analyze the spatial relationship between the presence or absence of a landslide event and a set of affecting factors (Pradhan and Lee, 2010a). The logit-natural logarithm of LR is indicated by an equation of the form:

$$Y = f(P) = \ln\left(\frac{P}{1-P}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n \tag{3}$$

Therefore, the probability of a landslide event (P) can be determined from following equation:

$$P = P(Y|X) = \frac{e^{\beta_0+\beta_1X_1+\beta_2X_2+\ldots+\beta_nX_n}}{1+e^{\beta_0+\beta_1X_1+\beta_2X_2+\ldots+\beta_nX_n}} \tag{4}$$

where Y represents outcome variables (landslide or non-landslide), $X = X_1, X_2, \ldots, X_n$ represents predictor variables of landslide affecting factors, n is the number of landslide affecting factors, β_0 is the intercept condition, $\beta_1, \beta_2, \ldots, \beta_n$ are the regression coefficients (Tu, 1996).

3.2.3. Fisher's Linear Discriminant Analysis (FLDA)

FLDA is known as the first method which used discriminant analysis (Bal and Örkcu, 2007), it was first proposed by Fisher (1936). It is based on the linear projections to find the eigenvectors of scatter matrices that discriminate the training classes for classification (Smith and Chang, 1994). Regarding landslide prediction, let (X_i, Y_j) represent the training dataset whereas $X_i, i = 1,2,\ldots,n$ represents the landslide affecting factors and $Y_j, j = 1,2$ represents landslide class or non-landslide class. Fisher's linear discriminant for classification is given by the parameter vector ω which maximizes following equation (Durrant and Kabán, 2010):

$$J(\omega) = \frac{\omega^T S_b \omega}{\omega^T S_W \omega} \tag{5}$$

where S_W is defined as "within classes scatter matrix" and S_b"between classes scatter matrix" which are expressed as below (Scholkopft and Mullert, 1999):

$$S_W = (\alpha_1 - \alpha_2)(\alpha_1 - \alpha_2)^T$$
$$S_b = \sum_{i=1,2} \sum_{X \in X_i} (X - \alpha_i)(X - \alpha_i)^T; \quad \alpha_i = \frac{1}{n_i} \sum_{X \in X_i} X \tag{6}$$

where α_i is the sample mean of the respective classes. Because$J(\omega)$ is invariant, therefore, the problem of maximizing J could be transformed into the following limited optimization problem (Cooke, 2002; Durrant and Kabán, 2010):

$$\min_\omega -\frac{1}{2}\omega^T S_b \omega \ \textit{subject to} \ \omega^T S_W \omega = 1 \tag{7}$$

3.2.4. Bayesian Network (BN)

BN was first introduced by Friedman et al. (1997) which is known as an effective method for knowledge representation under the influence of uncertainty (Pearl, 2014). It is based on Bayes' theorem to represent graphically and probabilistically correlative and causal relationships among variables (Marcot et al., 2006). BN is usually used widely in modeling of complex systems (Song et al., 2012b). For landslide prediction, BN is utilized to analyze the affecting factors contributing to rainfall-induced landslides from which assessing the susceptibility of landslide occurrences. The unique joint probability of a landslide event in relation with a set of affecting factors using BN can be estimated by equation of below form:

$$P_B(X_1, X_2, \ldots, X_n) = \prod_{i=1}^{n} P_B\left(X_1 \middle| \prod\nolimits_{X_i}\right) = \prod_{i=1}^{n} \theta_{X_i|\prod_{X_i}} \tag{8}$$

where $X = (X_1, X_2, \ldots, X_n)$ represents landslide affecting factors, $P_B(X|\prod_{X_i}) = \theta_{X_i|\prod_{X_i}}$ is a joint probability distribution over affecting factor X_i, n is the number of landslide affecting factors.

3.2.5. Naïve Bayes (NB)

NB is also based on Bayes' theorem (Pham et al., 2015b) that is a statistical approach using a conditional independence assumption that all attributes are independent with given output class (Soria et al., 2011). The main purpose of NB is to determine the prior probability of an event based on the proportion of the observed cases in relation with given output class. For spatial prediction of landslides, suppose that $X = (X_1, X_2, \ldots, X_n)$ is the vector of landslide affecting factors, $Y_j = (Y_1, Y_2)$ is the vector of categorical variables (landslide or non-landslide). The prior probability of a landslide event in relation with a set of affecting factors using NB can be estimated from following equation:

$$Y_{NB} = \underset{Y_i = [landslide, non \ \ landslide]}{\operatorname{argmax}} P(Y_i) \prod_{i=1}^{n} P(X_i/Y_i) \tag{9}$$

where, $P(Y_i)$ is the prior probability of event Y_i, $P(X_i/Y_i)$ is the conditional probability calculated as below:

$$\mathrm{P(X_i/Y_i)} = \frac{1}{\sqrt{2\pi}\delta} e^{\frac{-(X_i - \eta)^2}{2\delta^2}} \tag{10}$$

where δ is the standard deviation of X_i, η is the mean value of X_i.

3.3. Evaluation and comparison methods

According to Tien Bui et al. (2015), the performance of landslide models should be evaluated and compared for both training and testing datasets. For training dataset, the results show the degree of fit of landslide models over the data or the degree of fit between input and output using landslide models. Meanwhile, using testing dataset, the results show the predictive capability of landslide models (Tien Bui et al., 2012a). In this study, statistical index-based evaluations and ROC curve have been utilized to evaluate and compare the performance of five landslide models.

3.3.1. Statistical index based evaluations

Bennett et al. (2013) listed many qualitative and quantitative methods used to validate the performance of environmental models. In the present study, statistical indexes such as positive predictive value, negative predictive value, sensitivity, specificity, accuracy, kappa index, and root mean squared error have been selected to validate statistically the performance of landslide models (Table 2).

3.3.2. Receiver operating characteristic curve

Receiver operating characteristic (ROC) curve has been used to validate the general performance of landslide models. ROC curve is a useful technique to validate the quality of the probabilistic model (Pham et al., 2016a). It has been used as a standard method for validating landslide models in many studies (Feizizadeh et al., 2014; Tien Bui et al., 2016).

The ROC curve is constructed by plotting statistic index value pairs including "sensitivity", and "100-specificity". The AUC value is the area under the ROC curve that validates quantitatively how well the general performance of landslide models (Pradhan, 2013). Higher AUC value indicates better performance of landslide models. As the AUC value equals to 1.0, the performance of landslide models is perfect (Tien Bui et al., 2016).

4. Model study and analysis

4.1. Elimination and selection of landslide affecting factors

Geo-environmental factors (slope angle, slope aspect, elevation, curvature, plan curvature, profile curvature, soil type, land cover, mean annual rainfall, distance to lineaments, distance to roads, distance to rivers, lineament density, road density, and river density) have been considered as landslide affecting parameters in the present study. However, the contribution of these factors to landslide models might be different. Thus, it is necessary to evaluate the predictive capability of these landslide affecting factors to select suitable factors and remove irrelevant or unimportant factors for further analysis.

Correlation based Feature Selection (CFS), which is known as one of the most efficient feature selection methods (Hall, 1999), has been utilized in this study to evaluate the predictive capability of landslide affecting factors. CFS is carried out based on the correlation between each factor and each class, and inter-correlation between factors (Nguyen et al., 2010). The factors with higher Correlation Coefficient (CC) values indicate higher contribution to landslide models and vice versa. The predictive capability of fifteen landslide affecting factors using CFS method is shown in Fig 4. Based on these results, it can be concluded that all fifteen landslide affecting factors are having contribution to landslide models ($CC > 0$). However, there is no arguments to remove factors with low CC values such as curvature, plan curvature, profile curvature.

Table 2
Description of statistic index based evaluations.

No	Name	Formula	Description
1	Sensitivity (SST)	$SST = \frac{TP}{TP+FN}$	Sensitivity is the proportion of landslide pixels that are classified correctly as "landslide". It indicates how good predictive capability of landslide model is for classifying landslide pixels (Pham et al., 2016b).
2	Specificity (SPF)	$SPF = \frac{TN}{FP+TN}$	Specificity is the proportion of non-landslide pixels that are classified correctly as "non-landslide" (Pham et al., 2016b). It indicates how good predictive capability of landslide model is for classifying non-landslide pixels.
3	Accuracy (ACC)	$ACC = \frac{TP+TN}{TP+TN+FP+FN}$	Accuracy is the proportion of landslide and non-landslide pixels which are correctly classified (Bennett et al., 2013). It indicates how good performance of landslide model is.
4	Kappa (k)	$k = \frac{P_p - P_{\exp}}{1 - P_{\exp}}$	Kappa is used to evaluate the reliability of the landslide models (Bennett et al., 2013). Kappa value varies from −1 (non-reliable) to 1 (reliable).
5	Root Mean Squared Error (RMSE)	$RMSE = [(1/m)\sum_{i=1}^{m}(e_i - \bar{e}_i)^2]^{0.5}$	RMSE shows the error metric in the same units with the original data (Bennett et al., 2013). Smaller RMSE value indicates better performance of landslide model.

Note: TP (True Positive) value is the number of pixels that have been predicted correctly as "landslide", FP (False Positive) value is the number of pixel that have been predicted incorrectly as "landslide", TN (True Negative) value is the number of pixels that have been predicted correctly as "non-landslide", FN (False Negative) value is the number of pixels that have been predicted incorrectly as "non-landslide".
In addition, P_p is the proportion of number of pixels that have been classified correctly as landslide or non-landslide pixels, P_{exp} means the expected agreements, m is the number of observations, e_i is the estimated value of the ith observation, $\bar{e}_i$ is the measured value of the ith observation (Bennett et al., 2013; Chai and Draxler, 2014).

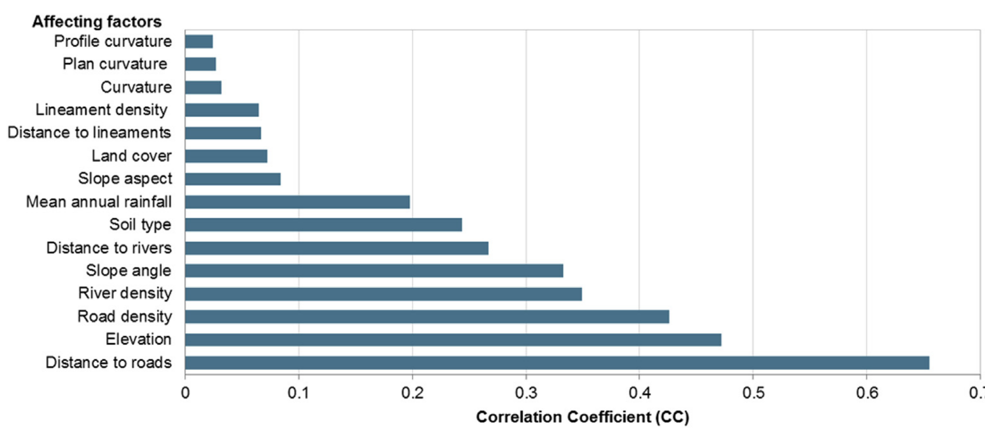

Fig. 4. Predictive capability of landslide affecting factors using CFS method.

Therefore, Backward Elimination (BE) method has been utilized further in the present study to provide more evidence in removing irrelevant or unimportant factors to improve the performance of landslide models as the elimination method helps in finding the most effective features for landslide modeling (Ballabio and Sterlacchini, 2012).

BE is a common method in feature subset selection (Chong and Jun 2005). The principle of BE is to find and remove the features that have the most decreases or least increases of classification accuracy of models (Kwon et al., 2003). Features with ρ value exceeding given removal criteria must be removed for further analysis from which the classification accuracy of models is improved (Cotter et al., 2001). In the present study, entry and removal criteria has been set 0.05 and 0.1, respectively as proposed

Table 3
Coefficients of landslide affecting factors using BE method.

No	Landslide affecting factors	Unstandardized coefficients		Standardized coefficients	t	ρ	Significance
		B	Std. Error	Beta			
1	Slope angle	0.091	0.004	0.203	21.959	0.000	Yes
2	Slope aspect	0.013	0.002	0.054	5.974	0.000	Yes
3	Elevation	−0.036	0.003	−0.156	−12.513	0.000	Yes
4	Curvature	−0.010	0.008	−0.022	−1.138	0.255	No
5	Plan curvature	−0.008	0.007	−0.018	−1.265	0.206	No
6	Profile curvature	−0.008	0.007	−0.016	−1.120	0.263	No
7	Soil type	−0.023	0.003	−0.065	−6.598	0.000	Yes
8	Land cover	0.000	0.004	0.000	−0.056	0.955	No
9	Annual rainfall	0.003	0.003	0.012	1.147	0.252	No
10	Distance to Lineaments	0.000	0.003	−0.002	−0.197	0.844	No
11	Distance to Roads	−0.122	0.003	−0.537	−47.116	0.000	Yes
12	Distance to Rivers	−0.021	0.003	−0.069	−6.527	0.000	Yes
13	Lineament density	−0.020	0.004	−0.052	−5.178	0.000	Yes
14	Road density	−0.009	0.005	−0.025	−1.956	0.050	Yes
15	River density	0.018	0.004	0.051	4.498	0.000	Yes

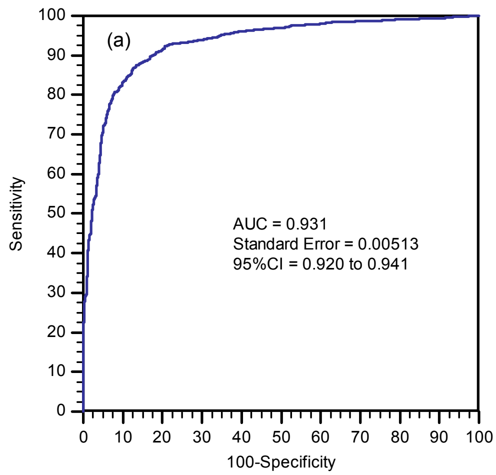

Fig. 5. Predictive capability of the SVM model: (a) before elimination (using 15 factors), (b) after elimination (using 9 proposed factors).

Table 4
Model performance using training dataset.

No.	Parameters	SVM	LR	FLDA	BN	NB
1	TP	2594	2457	2416	2413	2380
2	TN	2650	2565	2602	2531	2526
3	FP	322	459	500	503	536
4	FN	274	359	322	393	398
5	SST (%)	90.45	87.25	88.24	85.99	85.67
6	SPF (%)	89.17	84.82	83.88	83.42	82.50
7	ACC (%)	89.79	85.99	85.92	84.66	84.01
8	k	0.796	0.720	0.719	0.693	0.680
9	RMSE	0.279	0.325	0.329	0.343	0.355

by Chong and Jun (2005). The coefficients of landslide affecting factors using BE method are shown in Table 3 in which B are the unstandardized regression coefficients, Std. Error are the standard errors of the unstandardized regression coefficients, Beta is the standardized coefficient of regression, t is value of statistic test (http://www.jerrydallal.com/lhsp/slrout.htm), and ρ represents retention threshold (Wang et al., 2006). It can be observed that factors such as curvature ($\rho = 0.255$), plan curvature ($\rho = 0.206$), profile curvature ($\rho = 0.263$), land cover ($\rho = 0.955$), annual rainfall ($\rho = 0.252$), distance to lineaments ($\rho = 0.844$) have exceeded removal criteria (0.1). Therefore, these factors have been removed for spatial prediction of landslide in this study.

SVM has been utilized to test predictive capability before and after elimination of landslide affecting factors (Fig 5). It can be observed that the predictive capability of SVM has been improved significantly (1.9%) after eliminating unimportant factors using the BE method.

Based on above analysis of selection of affecting factors, out of

Table 5
Model predictive capability using testing dataset.

No.	Parameters	SVM	LR	FLDA	BN	NB
1	TP	1089	1039	1028	1035	1002
2	TN	1130	1100	1108	1075	1096
3	FP	160	210	221	214	247
4	FN	122	152	144	177	156
5	SST (%)	89.93	87.24	87.71	85.40	86.53
6	SPF (%)	87.60	83.97	83.37	83.40	81.61
7	ACC (%)	88.72	85.53	85.41	84.37	83.89
8	k	0.775	0.711	0.708	0.687	0.678
9	RMSE	0.290	0.329	0.331	0.346	0.358

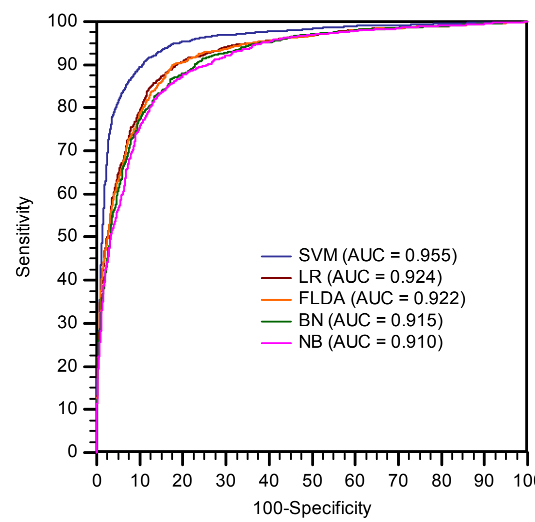

Fig. 6. Analysis of the ROC curve of different landslide models using training dataset.

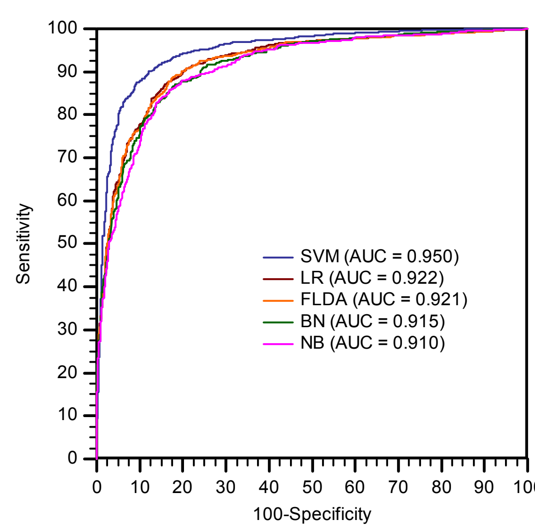

Fig. 7. Analysis of the ROC curve of different landslide models using testing dataset.

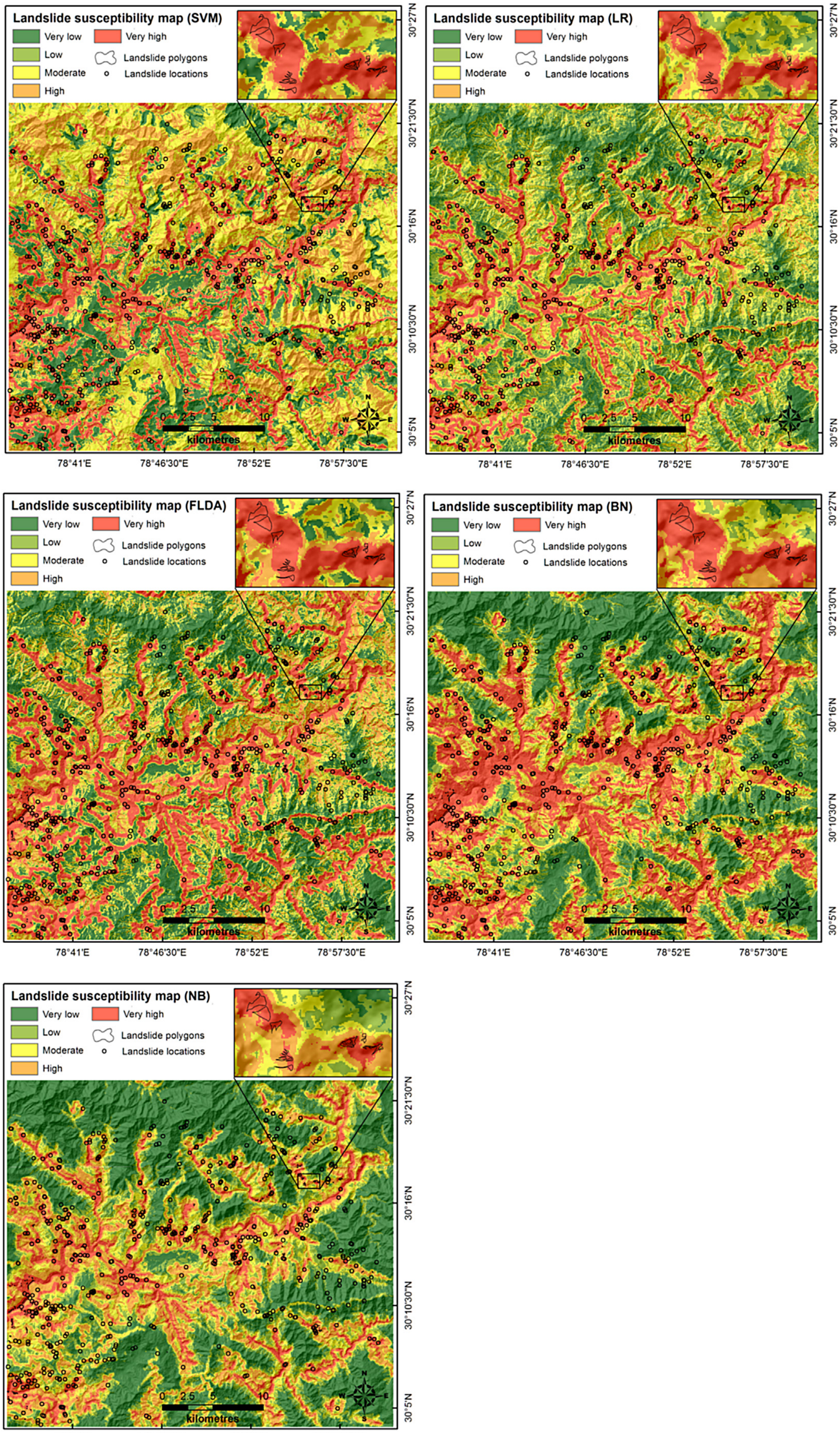

Fig. 8. Landslide susceptibility maps using different landslide models.

fifteen, only nine landslide affecting factors have been selected for landslide modeling in the present study. These factors are slope angle, slope aspect, elevation, soil type, distance to roads, distance to rivers, lineament density, road density, and river density.

4.2. Training and validating the landslide models

For constructing the proposed models for landslide prediction, training dataset and testing dataset have been generated. The training dataset is used to train landslide models whereas the testing dataset is utilized to validate the performance of the landslide models (Pham et al., 2016b).

The landslide inventory has been first divided randomly into two parts having (i) 70% of landslide inventory (2924 pixels), and (ii) 30% of remaining landslide inventory (1253 pixels). First part has been used to generate training dataset, and second part has been utilized to construct testing dataset. Similarly, non-landslide pixels have also been selected randomly from non-landslide area. For training dataset, 2924 non-landslide pixels have been used, and 1253 non-landslide pixels have been utilized to construct testing dataset. Landslide and non-landslide pixels with grid size of 20 × 20 m have been sampled with landslide affecting factors to generate the final datasets. Using training dataset, landslide models (SVM, LR, FLDA, BN, and NB) have been constructed.

Performance analysis of five landslide models using training and testing data sets has been done (Table 4 and Table 5). The results show that all applied models have shown high predictive capability for spatial prediction of landslides. Out of these, the SVM model has the highest predictive capability, followed by the LR model, the FLDA model, the BN model, and the NB model, respectively.

In addition, analysis of the ROC curve using training dataset is shown in Fig 6. The highest AUC value belongs to the SVM model (AUC = 0.955), followed by the LR model (AUC = 0.924), the FLDA model (AUC = 0.922), the FLDA model (AUC = 0.915), and the NB model (AUC = 0.910), respectively. Moreover, analysis of the ROC curve using testing dataset is shown in Fig 7. It can be observed that the SVM model has the highest AUC value (0.950), followed by the LR model (0.922), the FLDA model (0.921), the BN model (0.915), and the NB model (0.910), respectively. The analysis results of ROC curve confirm that the SVM model has the highest predictive capability compared with other landslide models (LR, FLDA, BN, and NB).

4.3. Development of landslide susceptibility maps

In the present study, all five landslide models have been used for the development of landslide susceptibility maps. Maps have been prepared in two main steps: (i) by generating landslide susceptibility indexes (LSIs), and (ii) reclassifying LSIs. In the first step, LSIs have been generated for entire study area in which each pixel has been assigned a unique susceptible index. In the second step, LSIs have been reclassified into different intervals using Geometrical Intervals (GI) method developed by ESRI (Frye, 2007). Based on the LSIs intervals, five susceptible classes have been identified namely very low, low, moderate, high, and very high for developing landslide susceptibility map (Fig 8).

Landslide susceptibility maps developed from different landslide models have been validated by calculating Landslide Density (LD) of each susceptible class on the maps (Table 6). LD is a ratio of the percentage of landslide pixels (P_{LS}) and the percentage of all pixels (P_P) of each susceptible class on the map as shown in following equation:

$$LD = \frac{P_{LS}}{P_p} \tag{11}$$

Table 6
Landslide Density (LD) on landslide susceptibility maps developed from different landslide models.

Class	LD				
	SVM	LR	FLDA	BN	NB
Very low	0.02	0.05	0.05	0.04	0.05
Low	0.04	0.06	0.05	0.06	0.08
Moderate	0.06	0.13	0.08	0.08	0.31
High	0.13	0.42	0.19	0.28	0.94
Very high	4.69	4.76	3.55	3.21	7.66

In addition, five landslide susceptibility maps have been combined with slope angle map to determine the intervals of slope angles that are more susceptible to landslide occurrences (Fig 9).

The overlay analysis of results confirms that most of landslides have occurred in very high and high classes which are associated with moderate slope of the ground varying from 20 to 40°.

5. Discussions and conclusions

In comparison to conventional methods such as expert's opinion based methods or analytic methods, machine learning methods are well known more efficient in solving many real world problems (Pham et al., 2015b; Pradhan, 2013). In the present study, five machine learning methods namely SVM, LR, FLDA, BN, and NB have been evaluated and compared for landslide susceptibility assessment. Two of these methods namely SVM and LR has already been applied widely in landslide prediction (Van Den Eeckhaut et al., 2006b; Yao et al., 2008), whereas other methods FLDA, BN, NB have rarely been applied for spatial prediction of landslides.

To improve the performance of landslide models, the predictive capability of landslide affecting factors should be tested using feature selection methods to select suitable factors and remove irrelevant or unimportant factors (Ballabio and Sterlacchini, 2012). In the present study, the CFS method has been applied to test the contribution of these affecting factors to landslide models. The CFS method is capable to show high predictive capability factors, however, it is unable to eliminate low predictive capability factors (Hall, 1999). Therefore, BE method has also been used to remove unimportant factors for improving the predictive capability of landslide models. The results show that out of fifteen landslide affecting factors, only nine factors (slope angle, slope aspect, elevation, soil type, distance to roads, distance to rivers, lineament density, road density, and river density) are having better predictive capability hence these factors have been used for improving the performance of landslide models. The results also show that all five landslide models perform really well considering the above factors

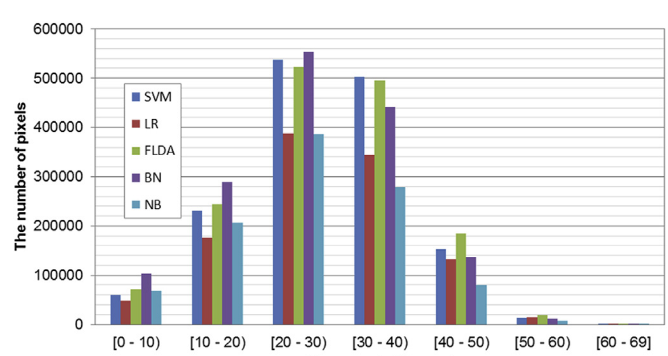

Fig. 9. Distribution of pixels of high and very high classes on slope map.

for landslide susceptibility assessment.

Analysis of comparative results indicated that out of these five models, the SVM model is one of the best methods. This has also been observed by other landslide studies (Marjanović et al., 2009, 2011; Tien Bui et al., 2015). In the present study, the LR model using a sequence of convergence criterions in maximizing the likelihood function for classification that can improve its performance (Tu, 1996), has shown good predictive capability.

The FLDA model applied in the study area show that it has almost same predictive capability in comparison to the LR model. This result appears to be more appropriate because FLDA uses discriminant analysis which has ability to deal with over-fitting problems (Alexandre-Cortizo et al., 2005; Liu and Wechsler, 1998). However, its predictive capability might be affected by equal variance-covariance matrices (Gilbert, 1969). Therefore, it has less predictive capability than SVM model.

The results also show that the BN model outperforms the NB model because BN can also indicate uncertainty interdependencies among random predictor variables, and it provides a semantic way for handling missing data and avoiding over-fitting problems (Liang et al., 2012; Song et al., 2012b). The analysis of results shows that NB is having lowest predictive capability compared to other landslide models. It is because NB is based on the independent assumption among predictor variables that might affect its predictive accuracy (Pham et al., 2015b).

Overall, all five landslide models have given good performance for landslide susceptibility assessment but the SVM model has given comparatively best performance, thus it can be used for assessment and development of better landslide susceptibility map for proper landslide hazard management.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Acknowledgement

Authors are thankful to the Director, Bhaskarcharya Institute for Space Applications and Geo-Informatics (BISAG), Department of Science & Technology, Government of Gujarat, Gandhinagar, Gujarat, India for providing facilities to carry out this research work.

References

Agarwal, N., Kumar, G., 1973. Geology of the upper Bhagirathi and Yamuna valleys, Uttarkashi district, Kumaun Himalaya. Himal. Geol. 3, 2–23.
Akgun, A., 2012. A comparison of landslide susceptibility maps produced by logistic regression, multi-criteria decision, and likelihood ratio methods: a case study at İzmir, Turkey. Landslides 9 (1), 93–106.
Alexandre-Cortizo, E., Rosa-Zurera, M., Lopez-Ferreras, F., 2005. Application of fisher linear discriminant analysis to speech/music classification, Computer as a Tool, 2005. In: EUROCON 2005. The International Conference on. IEEE, pp. 1666–1669.
Althuwaynee, O.F., Pradhan, B., Lee, S., 2012. Application of an evidential belief function model in landslide susceptibility mapping. Comput. Geosci. 44, 120–135.
Ambroise, C., McLachlan, G.J., 2002. Selection bias in gene extraction on the basis of microarray gene-expression data. Proc. Natl. Acad. Sci. 99 (10), 6562–6566.
Asamoah-Boaheng, M., 2014. Application of discrimination and classification on diabetes mellitus data. Int. J. Appl. 4 (6).
Bal, H., Örkcu, H., 2007. Data envelopment analysis approach to two-group classification problems and an experimental comparison with some classification models. Hacettepe J. Math. Stat. 36 (2).
Ballabio, C., Sterlacchini, S., 2012. Support vector machines for landslide susceptibility mapping: the Staffora River Basin case study, Italy. Math. Geosci. 44 (1), 47–70.
Bennett, N.D., Croke, B.F., Guariso, G., Guillaume, J.H., Hamilton, S.H., Jakeman, A.J., Marsili-Libelli, S., Newham, L.T., Norton, J.P., Perrin, C., 2013. Characterising performance of environmental models. Environ. Model. Softw. 40, 1–20.
Brewer, C.A., 2006. Basic mapping principles for visualizing cancer data using geographic information systems (GIS). Am. J. Prev. Med. 30 (2), S25–S36.
Cabrera, A.F., 1994. Logistic regression analysis in higher education: an applied perspective. In: Higher Education: Handbook of Theory and Research, 10, pp. 225–256.
Chai, T., Draxler, R.R., 2014. Root mean square error (RMSE) or mean absolute error (MAE)?—Arguments against avoiding RMSE in the literature. Geosci. Model Dev. 7 (3), 1247–1250.
Chen, J., Zeng, Z., Jiang, P., Tang, H., 2015. Deformation prediction of landslide based on functional network. Neurocomputing 149 Part A (0), 151–157.
Chong, I.-G., Jun, C.-H., 2005. Performance of some variable selection methods when multicollinearity is present. Chemom. Intell. Lab. Syst. 78 (1), 103–112.
Conforti, M., Pascale, S., Robustelli, G., Sdao, F., 2014. Evaluation of prediction capability of the artificial neural networks for mapping landslide susceptibility in the Turbolo River catchment (northern Calabria, Italy). CATENA 113, 236–250.
Cooke, T., 2002. Two variations on Fisher's linear discriminant for pattern recognition. IEEE Trans. Pattern Anal. Mach. Intell. (2), 268–273.
Cotter, S.F., Kreutz-Delgado, K., Rao, B.D., 2001. Backward sequential elimination for sparse vector subset selection. Signal Process. 81 (9), 1849–1864.
Dai, F., Lee, C., 2002. Landslide characteristics and slope instability modeling using GIS, Lantau Island, Hong Kong. Geomorphology 42 (3), 213–228.
Das, I., Sahoo, S., van Westen, C., Stein, A., Hack, R., 2010. Landslide susceptibility assessment using logistic regression and its comparison with a rock mass classification system, along a road section in the northern Himalayas (India). Geomorphology 114 (4), 627–637.
Dixon, B., Candade, N., 2008. Multispectral landuse classification using neural networks and support vector machines: one or the other, or both? Int. J. Remote Sens. 29 (4), 1185–1206.
Dudoit, S., Fridlyand, J., Speed, T.P., 2002. Comparison of discrimination methods for the classification of tumors using gene expression data. J. Am. Stat. Assoc. 97 (457), 77–87.
Durrant, R.J., Kabán, A., 2010. Compressed Fisher linear discriminant analysis: classification of randomly projected data. In: Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, pp. 1119–1128.
Feizizadeh, B., Roodposhti, M.S., Jankowski, P., Blaschke, T., 2014. A GIS-based extended fuzzy multi-criteria evaluation for landslide susceptibility mapping. Comput. Geosci. 73, 208–221.
Fisher, R.A., 1936. The use of multiple measurements in taxonomic problems. Ann. Eugen. 7 (2), 179–188.
Friedman, N., Geiger, D., Goldszmidt, M., 1997. Bayesian network classifiers. Mach. Learn. 29, 131–163.
Frye, C., 2007. About the Geometrical Interval Classification Method. http://blogs.esri.com/esri/arcgis.
Gilbert, E.S., 1969. The effect of unequal variance-covariance matrices on Fisher's linear discriminant function. Biometrics 505–515.
Guo, X., Huang, X., Jia, Y., 2005. Forward modeling of different types of landslides with multi-electrode electric method. Appl. Geophys. 2 (1), 14–20.
Guzzetti, F., 2006. Landslide Hazard and Risk Assessment. University of Bonn, Bonn, Germany.
Hall, M.A., 1999. Correlation-based Feature Selection for Machine Learning. The University of Waikato.
Jebur, M.N., Pradhan, B., Tehrany, M.S., 2015. Manifestation of LiDAR-derived parameters in the spatial prediction of landslides using novel ensemble evidential belief functions and support vector machine models in GIS. Sel. Top. Appl. Earth Observ. Remote Sens. IEEE J. 8 (2), 674–690.
Kavzoglu, T., Sahin, E.K., Colkesen, I., 2014. Landslide susceptibility mapping using GIS-based multi-criteria decision analysis, support vector machines, and logistic regression. Landslides 11 (3), 425–439.
Kwon, O.-W., Chan, K., Hao, J., Lee, T.-W., 2003. Emotion Recognition by Speech Signals, INTERSPEECH. Citeseer.
Lee, S., Hwang, J., Park, I., 2013. Application of data-driven evidential belief functions to landslide susceptibility mapping in Jinbu, Korea. CATENA 100, 15–30.
Lee, S., Ryu, J.-H., Kim, I.-S., 2007. Landslide susceptibility analysis and its verification using likelihood ratio, logistic regression, and artificial neural network models: case study of Youngin, Korea. Landslides 4 (4), 327–338.
Liang, W.-j., Zhuang, D.-f., Jiang, D., Pan, J.-j., Ren, H.-y., 2012. Assessment of debris flow hazards using a Bayesian Network. Geomorphology 171, 94–100.
Liu, C., Wechsler, H., 1998. Enhanced fisher linear discriminant models for face recognition, Pattern Recognition, 1998. In: Proceedings. Fourteenth International Conference on. IEEE, pp. 1368–1372.
Marcot, B.G., Steventon, J.D., Sutherland, G.D., McCann, R.K., 2006. Guidelines for developing and updating Bayesian belief networks applied to ecological modeling and conservation. Can. J. For. Res. 36 (12), 3063–3074.
Marjanović, M., Bajat, B., Kovačević, M., 2009. Landslide susceptibility assessment with machine learning algorithms. In: Intelligent Networking and Collaborative Systems, 2009. INCOS'09. International Conference on. IEEE, pp. 273–278.
Marjanović, M., Kovačević, M., Bajat, B., Voženílek, V., 2011. Landslide susceptibility assessment using SVM machine learning algorithm. Eng. Geol. 123 (3), 225–234.
Murillo-García, F.G., Alcántara-Ayala, I., 2015. Landslide Susceptibility Analysis and Mapping Using Statistical Multivariate Techniques: Pahuatlán, Puebla, Mexico, Recent Advances in Modeling Landslides and Debris Flows. Springer, pp. 179–194.
NCEP, 2014. Global Weather Data for SWAT. http://globalweather.tamu.edu/home.

Nguyen, H., Franke, K., Petrović, S., 2010. Improving effectiveness of intrusion detection by correlation feature selection. In: Availability, Reliability, and Security, 2010. ARES'10 International Conference on. IEEE, pp. 17–24.
Pearl, J., 2014. Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann.
Peng, C.-Y.J., Lee, K.L., Ingersoll, G.M., 2002. An introduction to logistic regression analysis and reporting. J. Educ. Res. 96 (1), 3–14.
Peng, L., Niu, R., Huang, B., Wu, X., Zhao, Y., Ye, R., 2014. Landslide susceptibility mapping based on rough set theory and support vector machines: a case of the Three Gorges area, China. Geomorphology 204 (0), 287–301.
Pham, B.T., Bui, D.T., Prakash, I., Dholakia, M., 2016a. 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 (1), 71–79.
Pham, B.T., Tien Bui, D., Dholakia, M.B., Prakash, I., Pham, H.V., 2016b. A comparative study of least square support vector machines and multiclass alternating decision trees for spatial prediction of rainfall-induced landslides in a tropical cyclones area. Geotech. Geol. Eng. 34 (1), 1–18.
Pham, B.T., Tien Bui, D., Indra, P., Dholakia, M.B., 2015a. Landslide susceptibility assessment at a part of Uttarakhand Himalaya, India using GIS – based statistical approach of frequency ratio method. Int. J. Eng. Res. Technol. 4 (11), 338–344.
Pham, B.T., Tien Bui, D., Pourghasemi, H.R., Indra, P., Dholakia, M.B., 2015b. Landslide susceptibility assessment in the Uttarakhand area (India) using GIS: a comparison study of prediction capability of naïve bayes, multilayer perceptron neural networks, and functional trees methods. Theor. Appl. Climatol. 122 (3–4), 1–19.
Pourghasemi, H.R., Jirandeh, A.G., Pradhan, B., Xu, C., Gokceoglu, C., 2013. Landslide susceptibility mapping using support vector machine and GIS at the Golestan Province, Iran. J. Earth Syst. Sci. 2, 349–369.
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 (0), 350–365.
Pradhan, B., Lee, S., 2010a. Delineation of landslide hazard areas on Penang Island, Malaysia, by using frequency ratio, logistic regression, and artificial neural network models. Environ. Earth Sci. 60, 1037–1054.
Pradhan, B., Lee, S., 2010b. Landslide susceptibility assessment and factor effect analysis: backpropagation artificial neural networks and their comparison with frequency ratio and bivariate logistic regression modelling. Environ. Model. Softw. 25 (6), 747–759.
Rossi, M., Guzzetti, F., Reichenbach, P., Mondini, A.C., Peruccacci, S., 2010. Optimal landslide susceptibility zonation based on multiple forecasts. Geomorphology 114 (3), 129–142.
Scholkopft, B., Mullert, K.-R., 1999. Fisher discriminant analysis with kernels. Neural Netw. Signal Process. IX 1 (1), 1.
Smith, J.R., Chang, S.-F., 1994. Transform features for texture classification and discrimination in large image databases, Image Processing, 1994. In: Proceedings. ICIP-94., IEEE International Conference. IEEE, pp. 407–411.
Song, Y., Gong, J., Gao, S., Wang, D., Cui, T., Li, Y., Wei, B., 2012a. Susceptibility assessment of earthquake-induced landslides using Bayesian network: a case study in Beichuan, China. Comput. Geosci. 42 (0), 189–199.
Song, Y., Gong, J., Gao, S., Wang, D., Cui, T., Li, Y., Wei, B., 2012b. Susceptibility assessment of earthquake-induced landslides using Bayesian network: a case study in Beichuan, China. Comput. Geosci. 42, 189–199.
Soria, D., Garibaldi, J.M., Ambrogi, F., Biganzoli, E.M., Ellis, I.O., 2011. A 'non-parametric'version of the naive Bayes classifier. Knowl. Based Syst. 24 (6), 775–784.
Suykens, J.A., Vandewalle, J., 1999. Least squares support vector machine classifiers. Neural Process. Lett. 9 (3), 293–300.
Tien Bui, D., Pham, B.T., Nguyen, Q.P., Hoang, N.-D., 2016. Spatial prediction of rainfall-induced shallow landslides using hybrid integration approach of Least-Squares Support Vector Machines and differential evolution optimization: a case study in Central Vietnam. Int. J. Digital Earth 1–21.
Tien Bui, D., Pradhan, B., Lofman, O., Revhaug, I., 2012a. Landslide susceptibility assessment in Vietnam using support vector machines, decision tree, and Naive Bayes models. Math. Problems Eng. 2012, 1–26.
Tien Bui, D., Pradhan, B., Lofman, O., Revhaug, I., Dick, O.B., 2012b. Spatial prediction of landslide hazards in Hoa Binh province (Vietnam): a comparative assessment of the efficacy of evidential belief functions and fuzzy logic models. CATENA 96 (0), 28–40.
Tien Bui, D., Tuan, T.A., Klempe, H., Pradhan, B., Revhaug, I., 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 1–18.
Tu, J.V., 1996. Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. J. Clin. Epidemiol. 49 (11), 1225–1231.
Van Den Eeckhaut, M., Vanwalleghem, T., Poesen, J., Govers, G., Verstraeten, G., Vandekerckhove, L., 2006a. Prediction of landslide susceptibility using rare events logistic regression: a case-study in the Flemish Ardennes (Belgium). Geomorphology 76 (3), 392–410.
Van Den Eeckhaut, M., Vanwalleghem, T., Poesen, J., Govers, G., Verstraeten, G., Vandekerckhove, L., 2006b. Prediction of landslide susceptibility using rare events logistic regression: a case-study in the Flemish Ardennes (Belgium). Geomorphology 76 (3–4), 392–410.
Vapnik, V.N., 1995. The Nature of Statistical Learning Theory. Springer-Verlag, New York.
Varnes, D.J., 1984. Landslide Hazard Zonation: a Review of Principles and Practice.
Venkatesan, M., Thangavelu, A., Prabhavathy, P., 2013. An improved Bayesian classification data mining method for early warning landslide susceptibility model using GIS. In: Proceedings of Seventh International Conference on Bio-inspired Computing: Theories and Applications (BIC-TA 2012). Springer, pp. 277–288.
Wang, T.J., Gona, P., Larson, M.G., Tofler, G.H., Levy, D., Newton-Cheh, C., Jacques, P.F., Rifai, N., Selhub, J., Robins, S.J., 2006. Multiple biomarkers for the prediction of first major cardiovascular events and death. N. Engl. J. Med. 355 (25), 2631–2639.
Yao, X., Tham, L.G., Dai, F.C., 2008. Landslide susceptibility mapping based on Support Vector Machine: a case study on natural slopes of Hong Kong, China. Geomorphology 101 (4), 572–582.
Yesilnacar, E., Topal, T., 2005. Landslide susceptibility mapping: a comparison of logistic regression and neural networks methods in a medium scale study, Hendek region (Turkey). Eng. Geol. 79 (3), 251–266.
Zare, M., Pourghasemi, H.R., Vafakhah, M., Pradhan, B., 2013. Landslide susceptibility mapping at Vaz Watershed (Iran) using an artificial neural network model: a comparison between multilayer perceptron (MLP) and radial basic function (RBF) algorithms. Arab. J. Geosci. 6 (8), 2873–2888.

View publication stats