Journal of Geomatics Vol 10 No. 1 April 2016

Evaluation of predictive ability of support vector machines and naive Bayes trees methods for spatial prediction of landslides in Uttarakhand state (India) using GIS

Binh Thai Pham[1,2], DieuTien Bui[3], Indra Prakash[4] and M.B. Dholakia[1]

[1]Department of Civil Engineering, Gujarat Technological University, Ahmedabad - 382424, India
[2]Department of Geotechnical Engineering, University of Transport Technology, Hanoi, Vietnam
[3]Geographic Information System Group, Department of Business Administration and Computer Science, Telemark University College, HallvardEikasPlass 1, N-3800 BøiTelemark, Norway
[4]Bhaskarcharya Institute for Space Applications and Geo-Informatics (BISAG), Gandhinagar, India
Email: phambinhgtvt@gmail.com or binhpt@utt.edu.vn

(Received: Feb 11, 2016; in final form: Mar 29, 2016)

Abstract: The main objective of this study is to apply and evaluate the predictive capability of the Support Vector Machines (SVM) and Naïve Bayes Trees (NBT) methods for spatial prediction of landslides in a part of Uttarakhand state (India). SVM is one of the most efficient machine learning methods that has been applied widely in landslide prediction whereas NBT has not been applied for landslide problems. In these models, a total of 430 historical landslide locations have been first identified to construct landslide inventory map. Landslide locations have been split randomly into two parts to generate training dataset (70% landslide locations) and testing dataset (30% landslide locations). Secondly, landslide affecting factors such as slope angle, slope aspect, elevation, plan curvature, lithology, soil, land cover, distance to roads, distance to rivers, distance to lineaments, and rainfall have been selected to assess the spatial relationship with landslide occurrences. The predictive capability of these factors has been evaluated using the Gain Ratio method. Using training dataset, the SVM and NBT models have been constructed to assess the susceptibility of landslide occurrences. Finally, the performance of the SVM and NBT models have been validated and compared using receiver operating characteristic curve technique and statistical index-based evaluations. The results show that both the SVM and NBT models perform well for spatial prediction of landslides. Out of these, the SVM model (AUC = 0.881) outperforms the NBT model (AUC = 0.832). Overall, SVM and NBT indicate promising methods which could be used for spatial prediction of landslides in landslide prone areas. Moreover, the results obtained from this study could be helpful for planning and decision making in landslide hazard management.

Keywords: Landslides, GIS, Support vector machines, Naïve Bayes trees

1. Introduction

Landslide is one of the most devastating natural disasters causing loss of human life and properties all over the world. India is known as one of the most affected countries by landslides in Asia (Guha-Sapir et al., 2014). Approximately 300 people die and 46 USD millions in properties loss every year in India (GSI, 2009). Most of landslides (about 80%) have occurred in Himalayan area (Onagh et al., 2012). Landslide studies have been turning into urgent tasks not only in India but also all over the world in order to reduce their harmful impaction to human life.

Spatial prediction of landslides is the probability of potential instability of slopes related to a set of casual factors (Guzzetti et al., 2005). It can be carried out by analyzing the spatial relationship between past landslide events and a set of geo-environmental factors. It is based on an assumption that future landslides will occur under same conditions with previous landslides (Ermini et al., 2005). Landslide susceptibility map is a final outcome of spatial prediction of landslides. It helps in land use planning and decision making for landslide hazard management (Pham et al., 2015a; Wang et al., 2009). To produce this map, Geographic Information System (GIS) is known as standard tool for integration of different types of data collected from various sources.

Many methods have been applied for spatial prediction of landslides using GIS in recent decades. These methods are based on main approaches (i) expert opinion-based approach and (ii) data mining based approach (Song et al., 2012). Expert opinion-based approach is subjective because it is based on the perspective of experts in selecting variables and assigning weights to variables. On the other hand, data mining based approach is objective as it uses machine learning algorithms to determine factors leading to landslide occurrences and calculate weights of the factors during learning of models. Out of these approaches, data mining based approach is more commonly utilized for spatial prediction of landslides. Common machine learning methods are logistic regression (Devkota et al., 2013; Lucà et al., 2011), decision tree (Pradhan, 2013; Yeon et al., 2010), artificial neural network (Choi et al., 2010; Zare et al., 2013) and support vector machines (Kavzoglu et al., 2014; Pradhan, 2013). In addition, Naïve Bayes Trees (NBT) is also an efficient machine learning technique that has been applied successfully in other fields but landslide problems.

In the present study, Support Vector Machines (SVM) and NBT methods have been applied and compared for spatial prediction of landslides. A small portion of Uttarakhand state, India had been selected as the study area. Receiver Operating Characteristic (ROC) curve method and statistic index-based evaluations have been used to validate and compare these landslide models. The analysis process has been done by using GIS application and Weka 3.7.12 software.

2. Description of study area

The study area lies between TehriGarhwal district and PauriGarhwal district of Uttarakhand state in India (longitudes 78º29'01'E to 78º37'06''E and latitudes 29º56'38''N to 30º09'37''N) covering an area of about 323.815 km² (Fig. 1). The study area is situated in subtropical monsoon region. The highest temperature is about 45°C in summer season whereas the lowest temperature is around 1.3°C in winter season. The humidity varies from 25% to 85%. Heavy rainfall often occurs in monsoon season (June to September) with annual average rainfall ranging from 770mm to 1684mm.

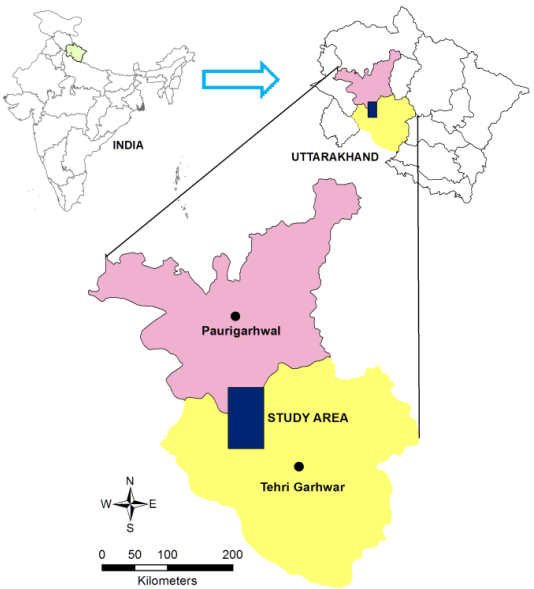

Figure 1: Location of the study area

Topographically, the study area is occupied by high mountains and intervening deep valleys (Pham et al., 2015b). Elevation ranges from 380m to 2180m (above mean sea level) with average elevation of 1081m. Slope angles are relatively steep up to 70 degrees. Slope angles of 15 to 45 degrees occupy the largest area (85.45%).Geologically, six lithological groups have presented in this study area namely Amri group (quartzite, phyllite),Blaini and Krol group (boulder bed and limestone), Bijni group (quartzite, phyllite), Jaunsar group (phyllite and quartzite), Manikot shell limestone (limestone), Tal group (sandstone, shale, quartzite, phyllite, and limestone) (Pham et al., 2015b). Baliana and Krol group and Bijni group are dominant with 30.1% and 28.1% of the study area, respectively. There are two types of soils in this study area viz., silty and loamy. Loamy soil occupies 73.73% of area.

Land cover in this study includes four categories such as dense forest, open-forest, non-forest, and scrub land. Non-forest is the dominant land cover (39.02%).

3. Methodology

Methodology of this study includes four main steps (i) constructing database for spatial prediction of landslides, (ii) evaluating predictive capability of landslide affecting factors, (ii) constructing landslide models (SVM, NBT) to assess landslide susceptibility, (iii) evaluating landslide models, (iv) constructing landslide susceptibility maps using Geo-informatics technology.

3.1. Data collection and interpretation

3.1.1. Landslide inventory map: Landslide inventory map is a compilation of landslide locations that have been occurred in the past and present. It is one of the most important data for spatial prediction of landslides. In the study area, many landslides occur every year (Onagh et al., 2012). By interpretation of Google Earth images using Google Earth pro 7.0, a total of 430 landslide location has been identified based on morphology and texture of past landslides to construct landslide inventory map (Fig. 2). These landslide locations have been evaluated in comparison with historical landslide reports, newspaper records and extensive field investigation. Subsequently, 70% landslide inventory (301 locations) has been used to construct training dataset for building landslide models whereas 30% remaining landslide inventory (129 locations) has been utilized to build testing dataset for validating landslide models.

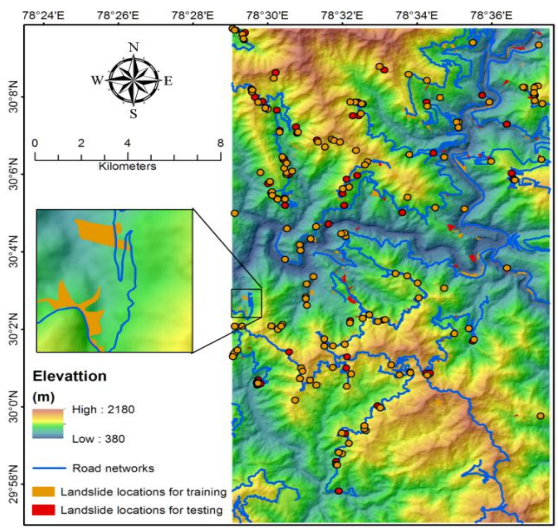

Figure 2: Landslide locations and elevation map of the study area

3.1.2. Landslide affecting factors: Landslides in the study area usually occur under various conditions such as geological activities, rainfall, geomorphological

characteristics, vegetation, human activities (Sarkar et al., 1995; Sengupta et al., 2010). Based on the mechanism of landslide occurrences and geo-environmental characteristics of the study area, eleven landslide affecting factors (slope angle, slope aspect, elevation, plan curvature, lithology, soil, land cover, distance to roads, distance to rivers, distance to lineaments, and rainfall) have been selected to assess the spatial relationship between them and landslide occurrences for spatial prediction of landslides.

Maps of these landslide affecting factors have been generated using GIS application. Specifically, slope angle map (Fig. 3a), slope aspect map, curvature map, and elevation map have been extracted from DEM with 20 m resolution generated from ASTER Global DEM collected from United States Geological Survey (http://earthexplorer.usgs.gov). Lithological map (Fig. 3b) has been extracted from state geological map on a scale of 1:1000000. Land cover map (Fig. 4a) has been extracted from state land cover map at a scale of 1:1000000 (http://www.ahec.org.in/wfw/maps.htm). Soil map (Fig. 4b) has been generated from state soil map at a scale of 1:1000000. Rainfall map has been constructed using meteorological data for 30 years from 1984 to 2014 (NCEP, 2014). Distance to roads map has been constructed by buffering road networks on the high slopes (larger than 15 degrees) in the study area. Distance to rivers map has been generated by buffering river networks on the high slopes (larger than 15 degrees) in the study area. Distance to lineaments map has been constructed by buffering lineament network generated from Landsat-8 sattelite image in the study area. Eleven landslide affecting factors and their classes is shown in Table 1.

3.2. Methods for spatial prediction of landslides

3.2.1. Support vector machines (SVM): SVM was first proposed by Vapnik (1995) that is one of the most effective machine learning techniques for classification (Kavzoglu et al., 2014). It is based on the statistical learning theory in order to find an optimal hyper-plane in separating two classes for classification (Tien Bui et al., 2015). SVM can be trained in two main steps (i) the original input space is first mapped into a high-dimensional feature space, (ii) the optimal hyper-plane in the feature space is determined by maximizing the margins between classes (Abe, 2005).

The performance of the SVM method depends significantly on the choice of the kernel function (Dixon and Candade, 2008). According to Tien Bui et al. (2012), Radial Basis Function (RBF) is one of the most widely used in landslide models among kernel functions. Therefore, RBF has been selected in this study in training SVM. In addition, two calculating parameter of radial basis function kernel have been optimized to obtain the best performance of the SVM model such as regularization parameter (C = 0.1) and kernel width (γ = 1).

Table 1: Landslide affecting factors and their class utilized in this study

No.	Landslide causal factors	Classes
1	Slope angle (degree)	(1) 0-8; (2); (3) 8-15; (4) 15-25; (5) 25-35; (6) 35-45; (7)> 45
2	Slope aspect	(1) flat[-1]; (2) north [0-22.5 and 337.5-360]; (3) northeast [22.5-67.5]; (4) east [67.5-112.5]; (5) southeast [112.5-157.5]; (6) south [157.5-202.5]; (7) southwest [202.5-247.5]; (8) west [247.5-292.5]; (9) northwest [292.5-337.5]
3	Elevation (m)	(1) 0-600; (2) 600-750; (3) 750-900; (4) 900-1050; (5) 1050-1200; (6) 1200-1350; (7) 1350-1500; (8) 1500-1650; (9) 1650-1800; (10) > 1,800
4	Curvature	(1) concave (<-0.05); (2) flat (-0.05-0.05); (3) and convex (> 0.05)
5	Lithology	(1) Amri group [quartzite, phyllite], (2) Blaini and Krol group [boulder bed and limestone], (3) Bijni group [quartzite, phyllite]; (4) Jaunsar group [phyllite and quartzite]; (5) Manikot shell limestone [limestone]; (6) Tal group [sandstone, shale, quartzite, phyllite, and limestone].
6	Soil	(1) coarse-loamy; (2) skeletal-loamy; (3) fine-loamy; (4) mixed-loamy; (5) fine-silt
7	Land cover	(1) non-forest; (2) dense-forest; (3) open-forest; (4) scrub-land;
8	Rainfall (mm)	(1) 0-900; (2) 900-1000; (3) 1000-1100; (4) 1100-1200; (5) 1200-1300; (6) 1300-1400; (7) 1400-1500; (8) > 1,500
9	Distance to lineaments (m)	(1) 0-50; (2) 50-100; (3) 100-150; (4) 150-200; (5) 200-250; (6) 250-300; (7) 300-350; (8) 350-400; (9) 400-450; (10) 450-500; (11) > 500
10	Distance to roads (m)	(1) 0-40; (2) 40-80;(3) 80-120; (4) 120-160; (5) 160-200; (6) > 200
11	Distance to rivers (m)	(1) 0-40; (2) 40-80;(3) 80-120; (4) 120-160; (5) 160-200; (6) > 200

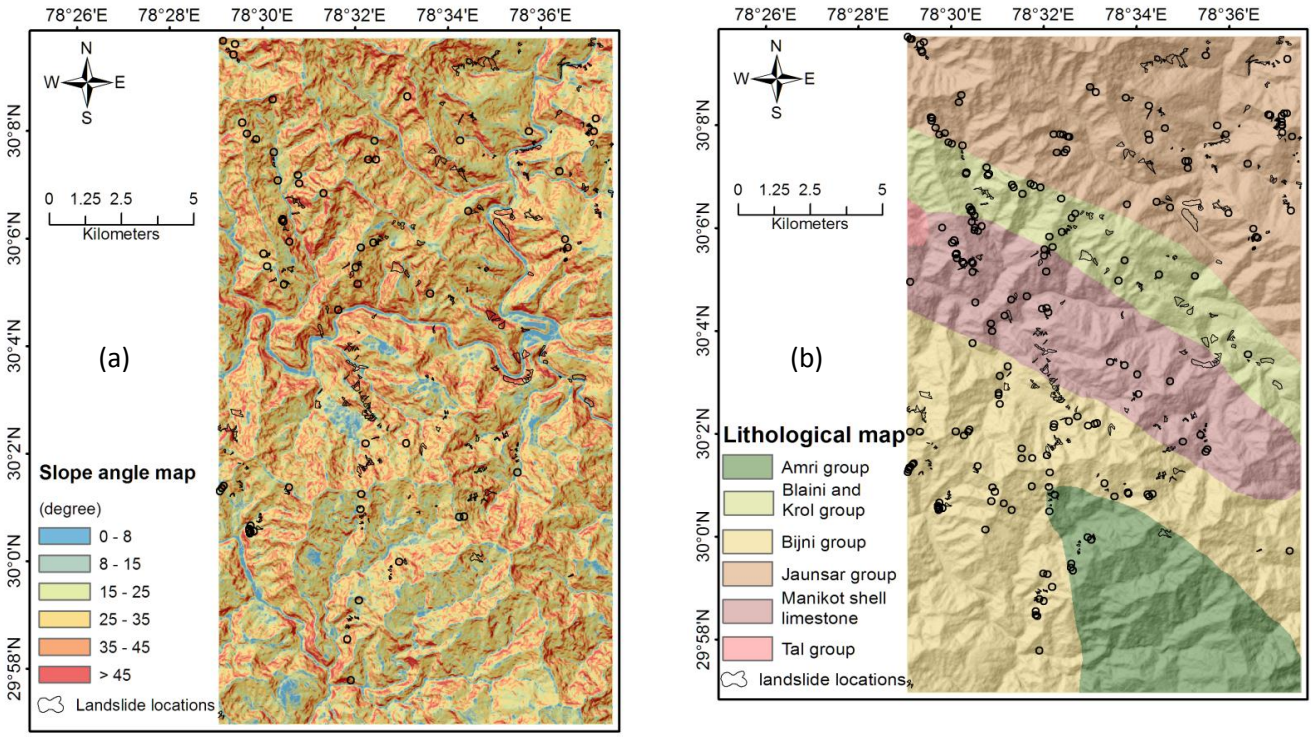

Figure 3: (a) Slope map and (b) Lithological map

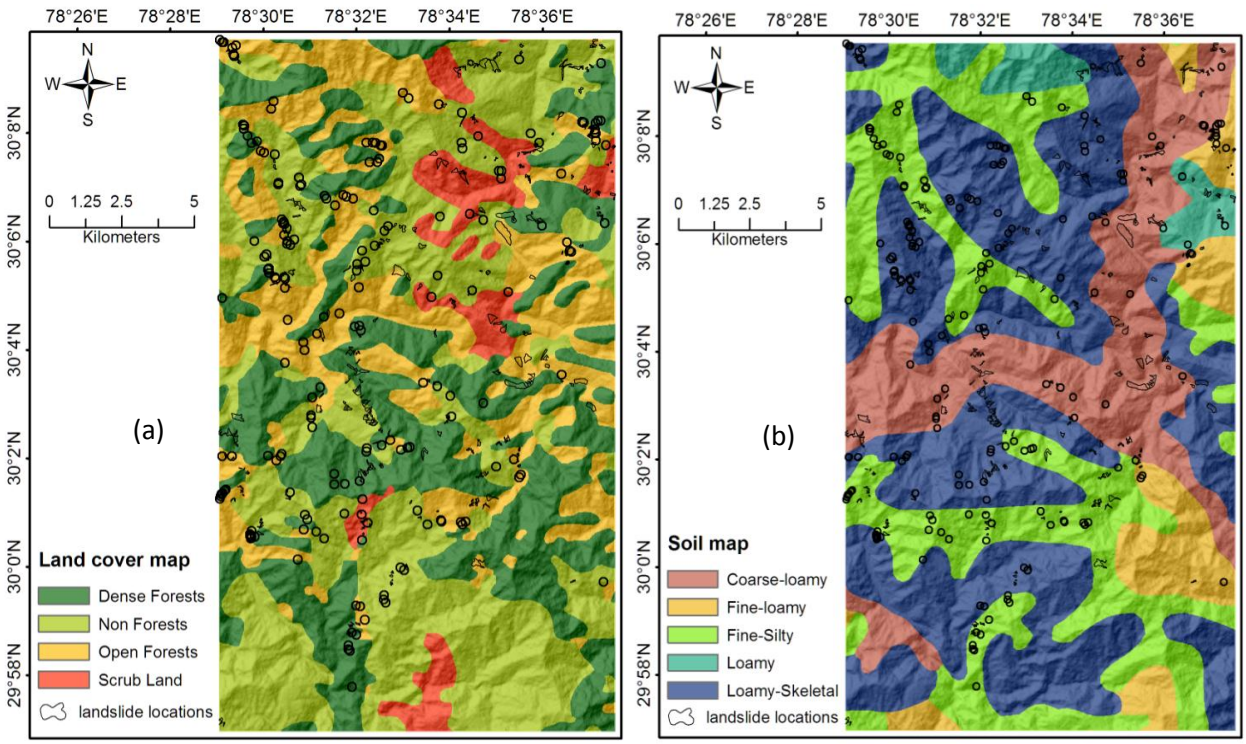

Figure 4: (a) Land cover map and (b) Soil map

3.2.2. Naïve Bayes Tree (NBT): NBT was first introduced by Kohavi (1996) which is a hybrid approach of Naïve Bayes and decision tree classifiers. Decision tree is one of the most commonly used in machine learning techniques that is based on a tree-like hierarchy for classification (Zhao and Zhang, 2008). Naïve Bayes is based on Bayes' theorem that considers all attributes are independent to maximize the posterior probability in determination the classified classes (Pham et al., 2015b; Soni et al., 2011). NBT is also a classification tree based method; however, it contains both nodes and leaves. The nodes consist of univariate splits as normal decision tree. The leaves include Naïve Bayes classifier (Kohavi, 1996). NBT is trained in two main steps (i) decision tree classifier is utilized to segment the data (ii) naïve Bayes classifier uses each segment of the data to create leaves for classifying variables. The performance of NBT is considered better than naïve Bayes and decision tree (Kohavi, 1996).

In this study, NBT has been applied at the first time for spatial prediction of landslides. Its performance has been compared with the performance of the SVM method.

4. Results

4. 1. Predictive capability of landslide affecting factors:

Evaluation of predictive capability of landslide affecting factors to landslide models in the study area has been carried out using Gain Ratio (GR) technique. It is known as an efficient feature selection method in evaluation the importance of input data to models (Quinlan, 1986). Factors with higher GR values have better predictive capability than those with lower GR values. Factors with zero GR value have no contribution to landslide models, thus it must be removed in analyzing process.

Predictive capability of eleven landslide affecting factors to landslide models in this study is shown in Fig. 5. It can be observed that distance to roads has the highest predictive capability to landslide models (GR = 0.149), followed by rainfall (GR = 0.058), curvature (GR = 0.042), elevation and slope angle (GR = 0.037), soil types (GR = 0.025), lithology (GR = 0.021), land cover (GR = 0.01), slope aspect (GR = 0.008), distance to rivers (GR = 0.006), and distance to lineaments (GR = 0.003), respectively. In general, all eleven landslide affecting factors are having contribution to landslide models in the study area (GR > 0). Therefore, they all have been selected for landslide analysis in the present study.

4. 2. Landslide susceptibility map

Landslide models such as SVM and NBT have been constructed using training dataset. Thereafter, landslide susceptibility mapsusing these landslide models have been constructed. Firstly, landslide susceptibility indices(LSIs) have been generated for all pixels in whole study area. These susceptible indices have been then reclassified into five intervals using geometrical interval method (Frye, 2007). Five landslide susceptibility classes have been named corresponding to five susceptible index intervals such as very low, low, moderate, high, and very high (Table 2 and Fig. 6).

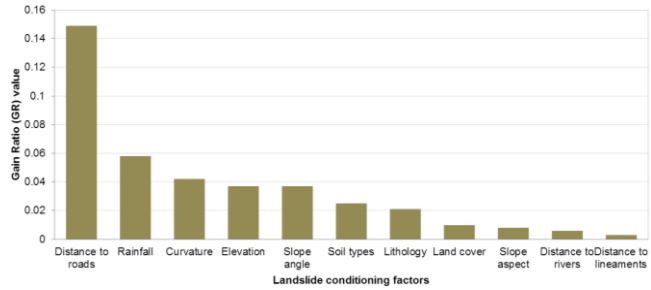

Fig 5: The Gain Ratio (GR) values of landslide affecting factors using

The performance of landslide susceptibility maps has been validated by overlaying them with landslide inventory map to calculate landslide density for each susceptible class. Landslide density is a ratio of the percentage of landslide pixels and the percentage of all pixels on each susceptible class. It can be observed from Table 2 that both landslide susceptibility maps perform well due to the fact that most of landslides have occurred on high and very high classes.

4. 3. Evaluation and comparison of landslide models

The performance of landslide models (SVM, NBT) has been validated using statistical index-based evaluations and ROC technique. Statistical indices such as sensitivity and specificity have been taken into account to evaluate the performance of landslide models. Sensitivity is the probability of the landslide pixels that have been classified correctly as “landslide” class (Pham et al. 2016a). It indicates how good landslide models for classification of landslide pixels. On the other hand, specificity is the probability of the non-landslide pixels that have been classified correctly as “non-landslide” class (Pham et al. 2016a). It shows how good landslide models for classification of non-landslide pixels.

ROC curve is a standard method to evaluate general performance of landslide models. It is constructed by plotting pairs of values (“sensitivity” and “100-specificity”). Area under ROC curve (AUC) is used to evaluate quantitatively the performance of landslide models. If the AUC value is larger than 0.8, the performance of landslide models is good and acceptable (Kantardzic, 2011).

The performance of landslide models is shown in Fig. 7 and Fig. 8. It can be observed that both landslide models perform well for spatial prediction of landslides in this study (AUC > 0.8). Out of these, the SVM model (AUC = 0.881) outperforms the NBT model (AUC = 0.832). More specifically, the SVM model (sensitivity = 84.4%, specificity = 81.6%) performs better than the NBT model (sensitivity = 82.5%, specificity = 76.6%) for classification of both landslide and non-landslide pixels. For both the SVM model

and the NBT model, the classification of landslide pixels is significantly better than those of non-landslide pixels.

Table 2: Landslide density on landslide susceptibility maps

Class	LSIs		% Pixels		% Landslides		Landside Density	
	SVM	NBT	SVM	NBT	SVM	NBT	SVM	NBT
Very low	0-0.01	0-0.001	0.176	0.164	0.005	0.011	0.026	0.064
Low	0.01-0.039	0.001-0.009	0.248	0.205	0.012	0.004	0.047	0.021
Moderate	0.039 - 0.121	0.0089-0.043	0.260	0.257	0.035	0.016	0.134	0.063
High	0.121-0.352	0.043-0.207	0.130	0.203	0.061	0.078	0.472	0.384
Very high	0.352-1	0.207-1	0.186	0.171	0.887	0.891	4.772	5.205

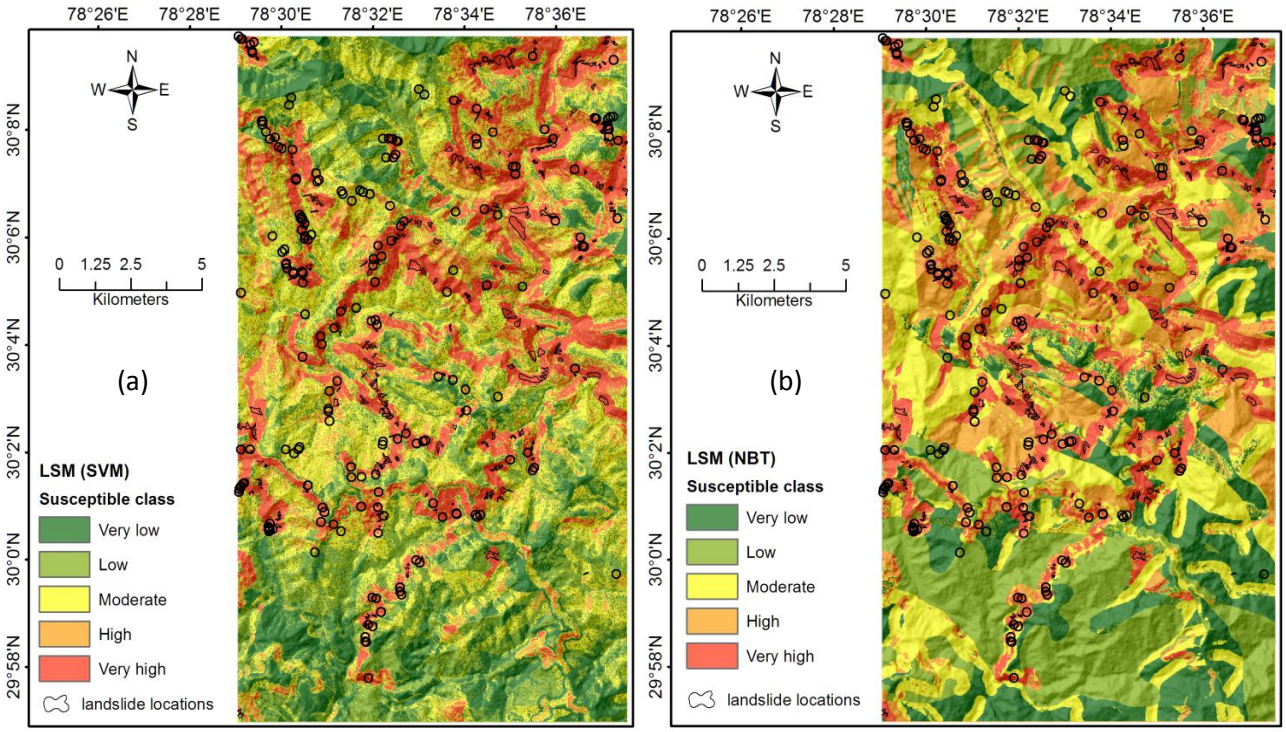

Figure 6: Landslide Susceptibility Maps (LSM) using (a) the SVM model and (b) the NBT model

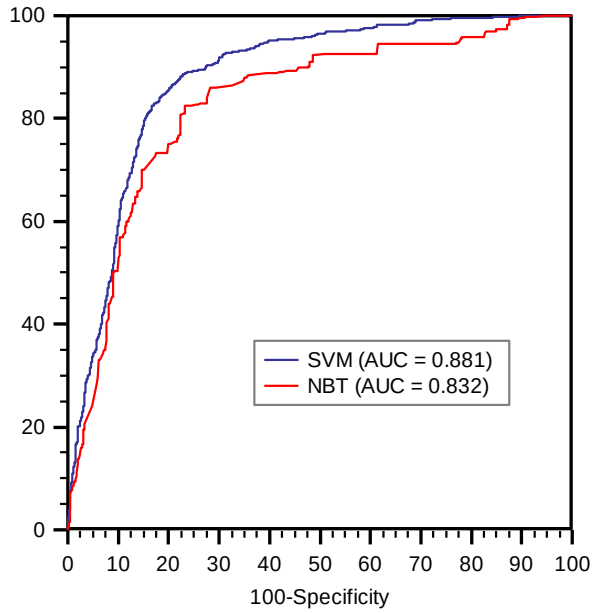

Figure 7: The performance of landslide models using ROC curve

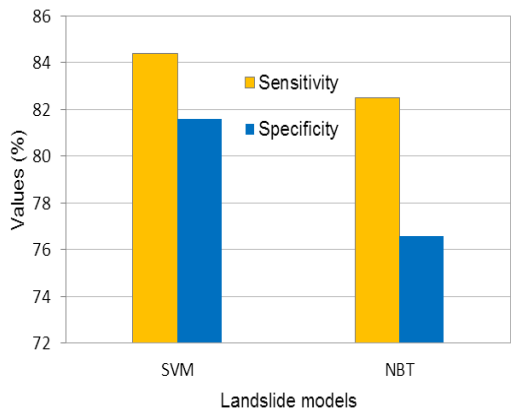

Figure 8: The performance of landslide models using statistical index-based evaluations

4.4 Discussions

The results show that both landslide models perform well for spatial prediction of landslides. Out of these, the SVM model outperforms the NBT model. This corroborates the fact of early study that the SVM model is one of the most efficient methods compared to conventional methods (Hassanien et al., 2015; Salama et al., 2013; Tien Bui et al., 2012). It might be due to the reason that NBT uses an assumption that all parameters are independent which is not really true in landslide problems (Tien Bui et al., 2015).

In addition, it is also necessary to evaluate the predictive capability of landslide affecting factors to landslide models in order to select suitable parameters to construct dataset for model learning process (Pham et al. 2016b). It is because the performance of landslide models depends significantly on the quality of input data that is constructed from landslide affecting factors (Pradhan, 2013). In the present study, the predictive capability of eleven landslide affecting factors has been evaluated using the GR method which is one of the most widely used feature selection techniques (Karegowda et al., 2010). The results show that in this area, distance to roads is important factor that has the highest predictive capability for landslide models. It is reasonable because most of identified landslides are on or adjacent to roads or highways. In addition, rainfall is also having high contribution to landslide models; this observation confirms the landslides occurrences during long-term heavy rain (Sengupta et al., 2010). Other factors are also having contribution to landslide models. Therefore, all eleven landslide affecting factors are appropriate for spatial prediction of landslides in the present study.

Conclusions

Overall, SVM and NBT indicate promising methods for spatial prediction of landslides which could be also used in other landslide prone areas. Out of these, the SVM model is having better predictive capability than NBT model. The present study would be helpful for land use planning, decision making and management of landslide hazard prone areas.

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

Abe, S. (2005). Support vector machines for pattern classification, 2. Springer.

Choi, J., H.J. Oh, J.S. Won and S. Lee (2010). Validation of an artificial neural network model for landslide susceptibility mapping. Environmental Earth Sciences, 60(3), pp 473-483.

Devkota, K.C., A.D. Regmi, H.R. Pourghasemi, K. Yoshida, B. Pradhan, I.C. Ryu, M.R. Dhital 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), pp 135-165.

Dixon, B. and N. Candade (2008). Multispectral landuse classification using neural networks and support vector machines: One or the other, or both? International Journal of Remote Sensing, 29(4), pp 1185-1206.

Ermini, L., F. Catani and N. Casagli (2005). Artificial neural networks applied to landslide susceptibility assessment. Geomorphology, 66(1), pp 327-343.

Frye, C. (2007). About the geometrical interval classification method. https://blogs.esri.com/esri/arcgis/2007/10/18/about-the-geometrical-interval-classification-method/.

GSI (2009). Geological Survey of India. www.portal.gsi.gov.in.

Guha-Sapir, D., P. Hoyois and R. Below (2014). Annual disaster statistical review 2013: The numbers and trends. Brussels: CRED.

Guzzetti, F., P. Reichenbach, M. Cardinali, M. Galli and F. Ardizzone (2005). Probabilistic landslide hazard assessment at the basin scale. Geomorphology, 72(1–4), pp 272-299.

Hassanien, A.E., M.A. Salama, J. Platos and V. Snasel (2015). Rough local transfer function for cardiac disorders detection using heart sounds. Logic Journal of IGP.

Kantardzic, M. (2011). Data mining: Concepts, models, methods and algorithms. John Wiley & Sons.

Karegowda, A.G., A. Manjunath and M. Jayaram (2010). Comparative study of attribute selection using gain ratio and correlation based feature selection. International Journal of Information Technology and Knowledge Management, 2(2), pp 271-277.

Kavzoglu, T., E.K. Sahin and I. Colkesen (2014). Landslide susceptibility mapping using GIS-based multi-criteria decision analysis, support vector machines and logistic regression. Landslides, 11(3), pp 425-439.

Kohavi, R., 1996. Scaling Up the Accuracy of Naive-Bayes Classifiers: A Decision-Tree Hybrid, KDD. Citeseer, pp. 202-207.

Lucà, F., M. Conforti and G. Robustelli (2011). Comparison of GIS-based gullying susceptibility mapping using bivariate and multivariate statistics. Northern Calabria, South Italy, Geomorphology, 134(3–4), pp 297-308.

NCEP (2014). Global weather data for SWAT. http://globalweather.tamu.edu/home.

Onagh, M., V. Kumra and P.K. Rai (2012). Landslide susceptibility mapping in a part of Uttarkashi district (India) by multiple linear regression method. International Journal of Geology, Earth and Environmental Sciences.

Pham, B.T., D. Tien Bui, P. Indra and M.B. Dholakia (2015a). Landslide susceptibility assessment at a part of Uttarakhand Himalaya, India using GIS – based statistical approach of frequency ratio method. International Journal of Engineering Research & Technology 4(11), pp 338-344.

Pham, B.T., D. Tien Bui, H.R. Pourghasemi, P. Indra and M.B. Dholakia (2015b). Landslide susceptibility assesssment 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, pp 1-19.

Pham, B.T., D. Tien Bui, MB. Dholakia, I. Prakash and H.V. Pham (2016a). 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. Geotechnical and Geological Engineering 34:1-18.

Pham, B.T., D. Tien Bui, I. Prakash and M.B. Dholakia (2016b). Rotation forest fuzzy rule-based classifier ensemble for spatial prediction of landslides using GIS. Natural Hazards:1-31.

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. Computers & Geosciences, 51, pp 350-365.

Quinlan, J.R. (1986). Induction of decision trees. Machine Learning, 1(1), 81-106.

Salama, M.A., O.S. Soliman, I. Maglogiannis, A.E. Hassanien and A.A. Fahmy (2013). Rough set-based identification of heart valve diseases using heart sounds. Rough Sets and Intelligent Systems-Professor Zdzisław Pawlak in Memoriam. Springer, pp 475-491.

Sarkar, S., D.P. Kanungo and G.S. Mehrotra (1995). Landslide hazard zonation: A case study in Garhwal Himalaya, India. Mountain Research and Development, 15(4), pp 301-309.

Sengupta, A., S. Gupta and K. Anbarasu (2010). Rainfall thresholds for the initiation of landslide at Lanta Khola in north Sikkim, India. Natural hazards, 52(1), pp 31-42.

Song, Y., J. Gong, S. Gao, D. Wang, T. Cui, Y. Li and B. Wei (2012). Susceptibility assessment of earthquake-induced landslides using Bayesian network: A case study in Beichuan, China. Computers & Geosciences, 42(0), pp 189-199.

Soni, J., U. Ansari, D. Sharma and S. Soni (2011). Predictive data mining for medical diagnosis: An overview of heart disease prediction. International Journal of Computer Applications, 17(8), pp 43-48.

Tien Bui, D., B. Pradhan, O. Lofman and I. Revhaug (2012). Landslide susceptibility assessment in Vietnam using support vector machines, decision tree and naive Bayes models. Mathematical Problems in Engineering.

Tien Bui, D., 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, pp 1-18.

Vapnik, V.N. (1995). The nature of statistical learning theory. Springer Verlag, New York.

Wang, W.D., C.M. Xie and X.G. Du (2009). Landslides susceptibility mapping based on geographical information system, GuiZhou, south-west China. Environmental geology, 58(1), pp 33-43.

Yeon, Y.-K., J.-G. Han and K.H. Ryu (2010). Landslide susceptibility mapping in Injae, Korea, using a decision tree. Engineering Geology, 116(3), pp 274-283.

Zare, M., H.R. Pourghasemi, M. Vafakhah and B. Pradhan (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. Arabian Journal of Geosciences, 6(8), pp 2873-2888.

Zhao, Y. and Y. Zhang (2008). Comparison of decision tree methods for finding active objects. Advances in Space Research, 41(12), pp 1955-1959.