

A Novel Hybrid Intelligent Approach of Random Subspace Ensemble and Reduced Error Pruning Trees for Landslide Susceptibility Modeling: A Case Study at Mu Cang Chai District, Yen Bai Province, Viet Nam

Binh Thai Pham¹  and Indra Prakash²

¹ Department of Geotechnical Engineering, University of Transport Technology, 54 Trieu Khuc, Thanh Xuan, Ha Noi, Viet Nam

binhpt@utt.edu.vn

² Department of Science and Technology, Bhaskarcharya Institute for Space Applications and Geo-Informatics (BISAG), Government of Gujarat, Gandhinagar, India

indra52prakash@gmail.com

Abstract. In the present study, a hybrid approach of Random Subspace Ensemble (RSS) and Reduced Error Pruning Trees (REPT) has been proposed to create a novel hybrid model namely RSS-REPT for landslide susceptibility modeling of the Mu Cang Chai district, Yen Bai province of Vietnam where is affected by a number of landslides every year. For the development of model, a spatial database consisting of 248 historic landslide events and 15 affecting factors (slope, aspect, curvature, plan curvature, profile curvature, elevation, lithology, land use, distance to faults, fault density, distance to roads, road density, distance to rivers, river density, and rainfall), was constructed to generate training and testing datasets. The novel hybrid model was then constructed using training dataset for landslide susceptibility assessment, and its predictive capability was validated using Receiver Operating Characteristic (ROC) curve and Statistical Indexes (SI) analysis. Performance of this novel model has been compared with another popular model namely Support Vector Machines (SVM). Results indicate that its performance (AUC = 0.835) is higher in comparison to the SVM model (AUC = 0.804). Thus the RSS-REPT can be considered as one of the promising methods for better landslide susceptibility assessment of landslide prone areas.

Keywords: Landslide susceptibility modeling · Random subspace ensemble · Reduced error pruning trees · Viet Nam

1 Introduction

Landslides frequently occur in hilly region of Viet Nam causing loss of life and damage to property every year [1–3]. Most of these landslides are triggered by heavy rains during monsoon (from May to October) [4]. It is well known that rainfall-induced

landslides cannot be prevented but can be predicted in regional scale by identifying high susceptible areas using GIS technology to minimize damages caused by future landslides [5].

In recent decades, many algorithms have been developed and applied to predict landslides which can be grouped broadly into three categories namely qualitative, semi-quantitative, and quantitative methods [6, 7]. Qualitative methods, which are based on expert's perspective, often use landside inventories to predict locations having identical geomorphological and geological characteristics [8]. Semi-quantitative methods use the expert's perspective in conjunction with the idea of weighting and ranking for predicting potential landslides [9]. Quantitative methods are based on the numerical analysis of the spatial relationship between affecting factors and historical landslides [10]. Out of these, quantitative methods are more objective and popular for landslide susceptibility modeling [11].

In literature, Support Vector Machines [12], Artificial Neural Networks [13], Decision Trees [14] are some of the most popular quantitative methods for landslide susceptibility modeling. However, in recent years, the hybrid models, which are a combination of ensemble techniques and individual classifiers, are considered more efficient than single models such SVM [15]. Therefore, in this study, we have proposed a novel hybrid model namely RSS-REPT which is a combination of two machine learning methods of Random Subspace (RSS) ensemble and Reduced Error Pruning Trees (REPT) classifier for landslide susceptibility modeling at the Mu Cang Chai district, Yen Bai province (Viet Nam). Predictive capability of this hybrid model has been validated using the Receiver Operating Characteristic (ROC) curve and Statistical Indexes (SI) analysis and its performance has been compared with SVM.

2 Methods Used

2.1 Random Subspace

Random Subspace (RSS) was proposed by Ho [16] to generate multiple feature spaces for constructing multiple decision trees in classification. Main objective of the RSS is to sample a feature set from the original high dimensional feature space into low dimensional subspaces on which classifiers are then constructed for classifying the classes, and, finally, the final decision is obtained by applying a majority voting rule for combining the generated classifiers [17].

More specifically, let each training sample set $X = (X_1, X_2, \dots, X_m)$ is an n -dimensional vector $X_i = (x_{i1}, x_{i2}, \dots, x_{in})$ ($i = 1, 2, \dots, n$), described by n features (landslide affecting factors). In the RSS, we randomly select $r < n$ features from the n -dimensional dataset of the original p -dimensional feature space X to obtain the r -dimensional random subspace. Thus, the modified training dataset $X^e = (X_1^e, X_2^e, \dots, X_m^e)$ includes r -dimensional training objects $X_i^e = (x_{i1}^e, x_{i2}^e, \dots, x_{ir}^e)$.

Thereafter, the classifiers are constructed in the random subspace X^e , and combined by applying simple majority voting in the final decision rule as following [18]:

$$\alpha(x) = \arg \max_{y \in \{0,1\}} \sum_a \delta_{\text{sgn}(C^a(x)), y} \quad (1)$$

where $y \in (0, 1)$ is a decision of the classifier, δ_{ij} is the Kronecker symbol, $C^a(x)$ are the generated classifiers ($a = 1, 2, \dots, A$) [18].

2.2 Reduced Error Pruning Trees

Reduced Error Pruning Trees (REPT) introduced by Quinlan [19] is a combination of Decision Trees (DT) method and Reduced Error Pruning (REP) technique. The DT is one of the most popular classification tree methods for solving a lot of real world problems including landslide prediction [20]. However, it has a problem of producing the outcome of a large decision tree in classification. The REP is an effective machine learning technique in reducing the size of decision trees so that it can reduce the complexity of the final classifier [21]. It can also improve the predictive accuracy of the classifier by reducing the over-fitting and removing the tree sections which might cause the noisy or erroneous data [21]. Thus, the hybrid REPT method can give the desirable outcomes for classification problems. In this study, the hybrid REPT method has been used as a base classifier in ensemble framework for landslide spatial prediction.

2.3 Evaluation and Comparison Methods

In this study, the Receiver Operating Characteristic (ROC) curve and several statistical indexes have been selected for validating the performance of the novel hybrid model. Out of these methods, the ROC curve is one of the most popular quantitative evaluation methods in landslide susceptibility modeling. It is based on the graph which is plotted using two statistical values (“sensitivity” and “100-specificity”) [11, 22]. To validate quantitatively the predictive capability of the model, the area under the ROC curve called AUC has been used. Model has better performance as the AUC values closer to “1” [10, 23]. Models with high AUC values are better than those with lower AUC values. In landslide modeling studies, statistical indexes namely positive and negative predictive values, sensitivity, specificity, accuracy, kappa, and root mean squared error have been widely used for validating the performance of models [24]. These statistical indexes are calculated on the basis of calculation of the probability of pixels (landslide pixels and non-landslide pixels) in the relation with variable classes (landslide and non-landslide). Models with higher values of positive and negative predictive values, sensitivity, specificity, accuracy, and kappa have better performance than those with lower values. In contrast, models with low root mean squared error values have better predictive capability than those with high values. Models are perfect when the root mean squared error values are equal to zero.

In addition, the SVM, which is known as a benchmark single landslide model [12], has been selected to compare with the proposed hybrid RSS-REPT model. The SVM is a binary classifier which is based on the statistical algorithm to separate two classes

(landslide and non-landslide) by hyper-plane in the original space [25]. This method has already been applied widely and efficiently in many landslide studies [26, 27].

3 Description of the Study Area

The Mu Cang Chai district (long. 103°56'00"E to 104°23'00"E and lat. of 21°39'00"N to 21°50'00"N), the study area, is located in Yen Bai province in the northern part of Viet Nam (Fig. 1). It covers an area approximately 1196.47 km². The study area falls into the humid subtropical climate region having annual average humidity of 81% and annual average temperature of 25°C. There are two separate seasons in a year namely dry season (November to April) and rainy season (May to October). Heavy rainfall usually happens in the rainy season generally during August and September. The total annual rainfall varies from 3700 mm to 5490 mm. Excessive rains is the main triggering factor causing many landslides in this area.

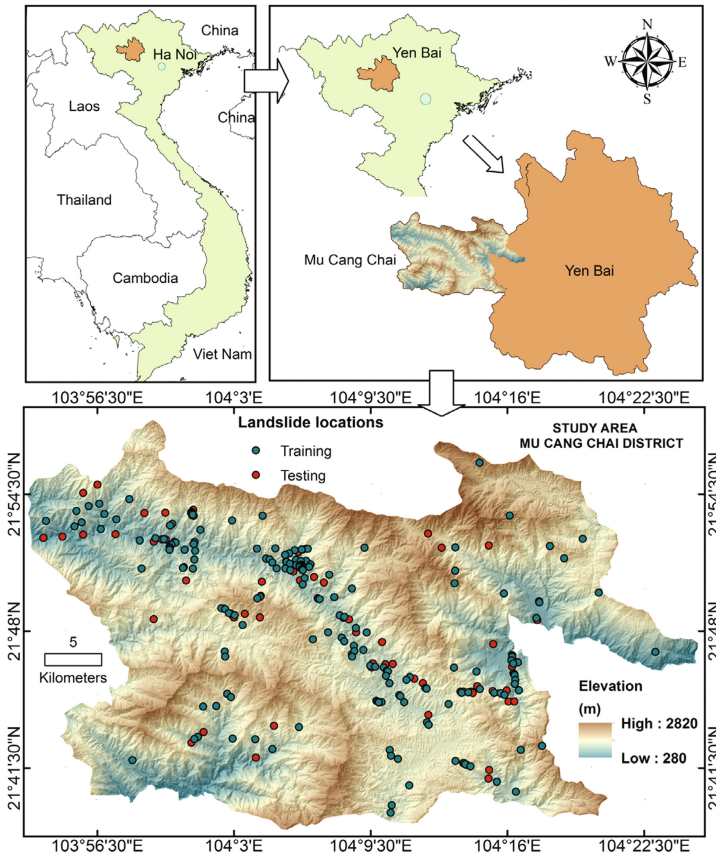


Fig. 1. Landslide inventory and location map of the study area

Topography of the study area is adulatory. The study area is occupied by high mountain ranges and deep intervening valleys. Elevation varies in this area from 280 m to 2820 m. About eighty percent of the ground slope in the area varies from 10 to 40 degrees. Some of the hills are having very steep slopes up to 88 degrees.

The study area is occupied mainly by extrusive volcanic rocks of Ngoi Thia and Tu Le complexes. The intrusive magmatic rocks of Tram Tau formation and Phu Sa Phin complex occupy the smaller part of the study area. Terrigenous sedimentary rocks and Quaternary sediments are also present in the area (Fig. 3). Tectonically, the area is disturbed and dissected by various faults namely Nghia Lo, Nam Co- Minh An, and Phong Tho- Van Yen.

4 Preparation of GIS Database

4.1 Landslide Inventory Map

Landslide inventory map has been constructed with 248 historical landslide locations which have been identified using Remote Sensing techniques based on the interpretation of 1:30000 scale air photos (Year 2013) and Google Earth images. Field investigation has been carried out to check and validate these landslide locations under the national project of Viet Nam named “Survey, assessment and zoning of landslide warning in the mountainous region of Vietnam” at Vietnam Institute of Geosciences and Mineral Resources. In the study area, the largest size of landslides observed is of 100000 m³ volume at the Che Cu Na commune (2011).

4.2 Landslide Affecting Factors

In the study area, fifteen landslide affecting factors namely slope, aspect, profile curvature, curvature, plan curvature, elevation, distance to rivers, distance to faults, distance to roads, river density, fault density, road density, lithology, land use, and rainfall have been selected for landslide susceptibility modeling (Table 1). Thematic data has been obtained from the Vietnam Institute of Geosciences and Mineral Resources for the development of raster maps of 20 × 20 m size for the GIS analysis in conjunction with Digital Elevation Model (DEM).

Lithology and faults have been extracted from the available geological and mineral resources maps of the study area (1:50,000 scale) (Fig. 2). Distance to faults and fault density maps have been generated by using buffer and density functions in ArcGIS application. Similarly, distance to roads, road density, distance to rivers, and river density maps have been generated for the analysis.

Land use map (Fig. 4) has been generated from the interpretation of 1:30000 scale air photos (Year 2013) and updated from Google images. Rainfall data has been extracted from the Global Weather data for SWAT [28] of 30 years (1984 to 2014), and used to generated rainfall map of the study.

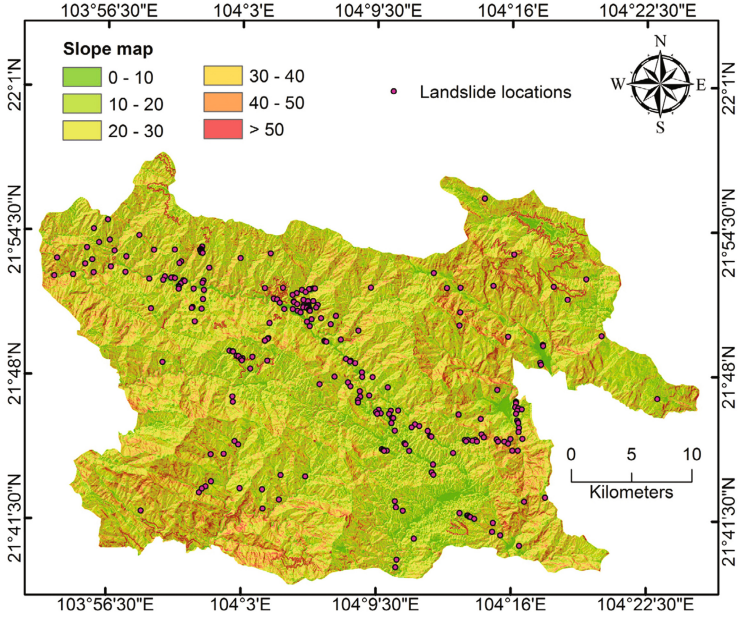


Fig. 2. Slope map of the study area

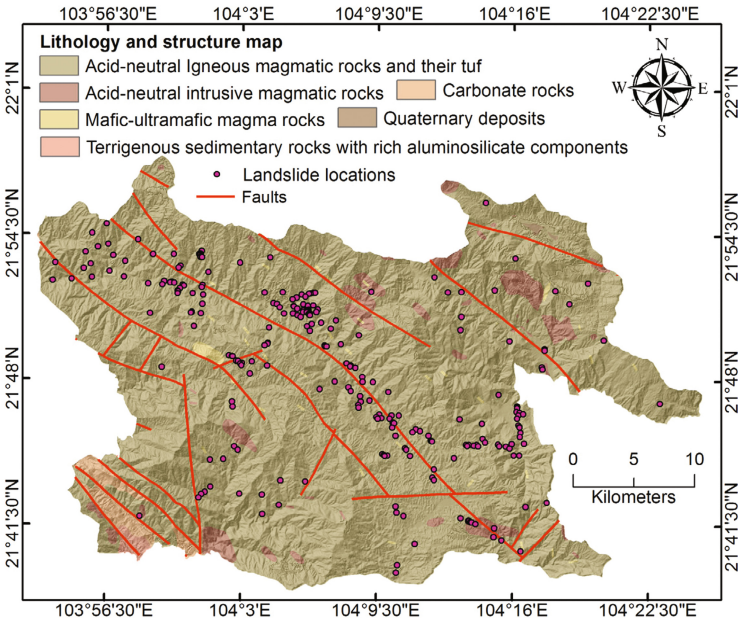


Fig. 3. Lithology and structure map of the study area

Table 1. Landslide affecting geo-environmental factors and their classes

No	Geo-environmental factors	Classes
1	Slope (degrees)	(1) 0–10, (2) 10–20, (3) 20–30, (4) 30–40, (5) 40–50, and (6) >50
2	Elevation (m)	(1) 280–700, (2) 700–900, (3) 900–1100, (4) 1100–1300, (5) 1300–1500, (6) 1500–1700, (7) 1700–1900, (8) 1900–2100, (9) 2100–2300, and (10) >2300
3	Aspect	(1) flat, (2) north, (3) northeast, (4) east, (5) southeast, (6) south, (7) southwest, (8) west, and (9) northwest
4	Plan curvature	(1) [(-334.189)–(-69.843)], (2) [(-69.843)–(-13.507)], (3) [(-13.507)–(-2.673)], (4) 4 [(-2.673)–3.827], and (5) [3.827–218.338]
5	Curvature	(1) concave (<-0.05), (2) flat (-0.05–0.05), and (3) convex (>0.05)
6	Profile curvature	(1) [(-255.398)–(-52.003)], (2) [(-52.003)–(-9.183)], (3) [(-9.183)–5.804], (4) [5.804–48.624], and (5) [48.624–290.557]
7	Land use	(1) barren land, (2) cultivated land, (3) forest land, (4) residential area, (5) scrub land, and (6) water bodies
8	Lithology	(1) acid-neutral igneous magmatic rocks and their tuff, (2) acid-neutral intrusive magmatic rocks, (3) terrigenous sedimentary rocks with rich aluminosilicate components, (4) mafic-ultramafic magma rocks, (5) carbonate rocks, and (6) quaternary deposits
9	Distance to faults (m)	(1) 0–100, (2) 100–200, (3) 200–300, (4) 300–400, (5) 400–500, (6) 500–600, (7) 600–700, and (8) >700
10	Fault density (km/km ²)	(1) very low (0–0.319), (2) low (0.319–0.861), (3) moderate (0.861–1.436), (4) high (1.436–2.169), and (5) very high (2.169–4.068)
11	Distance to roads (m)	(1) 0–50, (2) 50–100, (3) 100–150, (4) 150–200, (5) 200–250, and (6) > 250
12	Road density (km/km ²)	(1) very low (0–0.319), (2) low (0.319–0.861), (3) moderate (0.861–1.435), (4) high (1.435–2.169), and (5) very high (2.169–4.0676)
13	Distance to rivers (m)	(1) 0–50, (2) 50–100, (3) 100–150, (4) 150–200, (5) 200–250, and (6) >250
14	River density (km/km ²)	(1) very low (0–0.186), (2) low (0.186–0.428), (3) moderate (0.428– 0.703), (4) high (0.703–1.164), and (5) very high (1.1642–2.061)
15	Rainfall (mm)	(1) < 4000, (2) 4000–4250, (3) 4250–4500, (4) 4500–4750, (4) 4750–5000, (6) 5000–5250, and (7) >5250

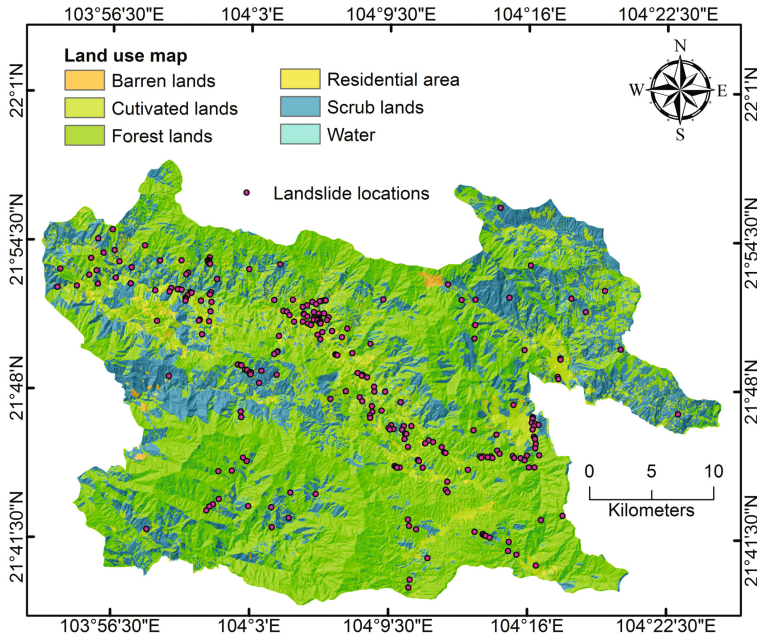


Fig. 4. Land use/ land cover map of the study area

5 Landslide Susceptibility Assessment Using a Novel Hybrid RSS-REPT Model

Landslide susceptibility assessment at the Mu Cang Chai district using a novel hybrid model RSS-REPT has been carried out in four main steps: (1) generating training and testing datasets, (2) constructing landslide model, (3) validating landslide model, (4) constructing landslide susceptibility map.

- (1) Generating training and testing datasets: Data collected from the study area has been processed to generate the training dataset and testing dataset. Training data set has been used to construct landslide models, and the testing dataset for validating models. For generating the training dataset, 70% of landslide locations (174 locations) have been selected randomly and remaining 30% landslide locations (74 locations) have been used to generate the testing dataset. These landslide locations have also been used to sample with landslide affecting factors for generating the final datasets for machine learning models.
- (2) Constructing landslide model: In this step, the training dataset has been used. The RSS ensemble has been applied to divide the training datasets into different sub-training datasets in different ways with different iterations. Simultaneously, the REPT classifiers have been applied using the generated sub-training datasets for classification. Finally, the RSS ensemble has been used to combine all generated REPT classifiers to construct the RSS-REPT model. For constructing the

hybrid model RSS-REPT, the parameters of algorithms have been set using trial-and-error process as follows: the number of iterations to be performed is 24; the amount of data used for pruning (folds) is 10; the minimum total weight of the instances in a leaf is 2; and tree pruning is applied.

- (3) Validating landslide model: In this step, the testing dataset has been used. Predictive capability of the proposed model RSS-REPT has been validated using different evaluation criteria such as the ROC curve and SI analysis. In addition, the performance of the RSS-REPT has been validated by comparing with another well-known landslide model namely SVM.
- (4) Constructing landslide susceptibility map: Landslide susceptibility map has been constructed by generating and classifying landslide susceptibility indexes using the results of proposed RSS-REPT model.

6 Results and Discussions

6.1 Training and Validating the RSS-REPT Model

Results of training of the proposed RSS-REPT model are shown in Table 2 and Fig. 5. The results show that the values of statistical indexes and the ROC curve analysis using training dataset are high. More specifically, the positive predictive value is 79.89% indicating that the probability of classification of pixels to landslide class is high; the negative predictive value is 75.86% indicating that probability of classification of pixels to non-landslide class is high; the sensitivity value is 76.80% indicating that the probability of classification of landslide pixels to landslide class is high; the specificity value is 79.04% indicating that the probability of classification of non-landslide pixels to non-landslide class is high; the accuracy value is 77.87% indicating that the classification accuracy of the model is high; the Kappa value is 0.558 indicating that the results of the proposed model is reliable; the RMSE value is 0.384 indicating that the error of training model is relative low; and the AUC value is 0.896 indicating that the general performance of the model is very good.

Table 2. Model performance using training dataset

No	Statistical indexes	RSS-REPT	SVM
1	Positive predictive value (%)	79.89	79.31
2	Negative predictive value (%)	75.86	75.29
3	Sensitivity (%)	76.80	76.24
4	Specificity (%)	79.04	78.44
5	Accuracy (%)	77.87	77.30
6	Kappa (k)	0.558	0.546
7	Root Mean Squared Error (RMSE)	0.384	0.395

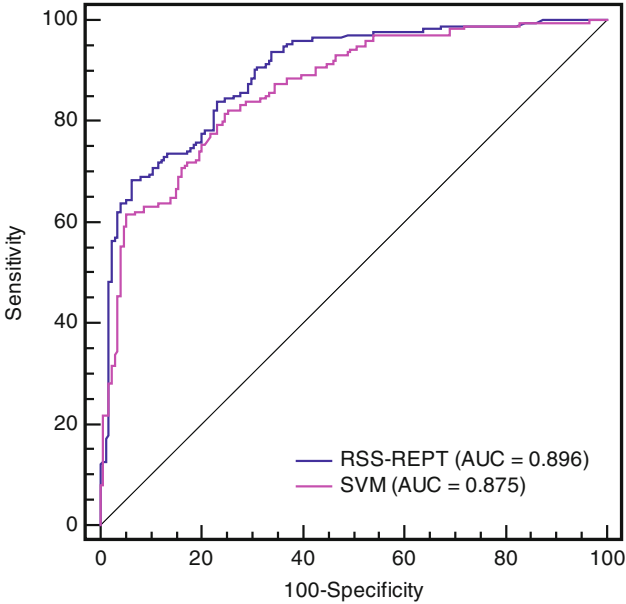


Fig. 5. The ROC curve analysis of landslide models using training dataset

Results of validating the proposed RSS-REPT model are shown in Table 3 and Fig. 6. The results show that the values of statistical indexes and the ROC curve analysis using testing dataset are high. More specifically, the positive predictive value is 79.73% indicating that the probability of classification of pixels to landslide class is high; the negative predictive value is 72.97% indicating that probability of classification of pixels to non-landslide class is high; the sensitivity value is 74.68% indicating that the probability of classification of landside pixels to landside class is high; the specificity value is 78.26% indicating that the probability of classification of non-landside pixels to non-landside class is high; the accuracy value is 76.35% indicating that the classification accuracy of the model is high; the Kappa value is 0.527 indicating that the results of the proposed model is reliable; the RMSE value is 0.419

Table 3. Model performance using testing dataset

No	Statistical indexes	RSS-REPT	SVM
1	Positive predictive value (%)	79.73	73.68
2	Negative predictive value (%)	72.97	70.83
3	Sensitivity (%)	74.68	72.73
4	Specificity (%)	78.26	71.83
5	Accuracy (%)	76.35	72.30
6	Kappa (k)	0.527	0.473
7	Root mean squared error (RMSE)	0.419	0.424

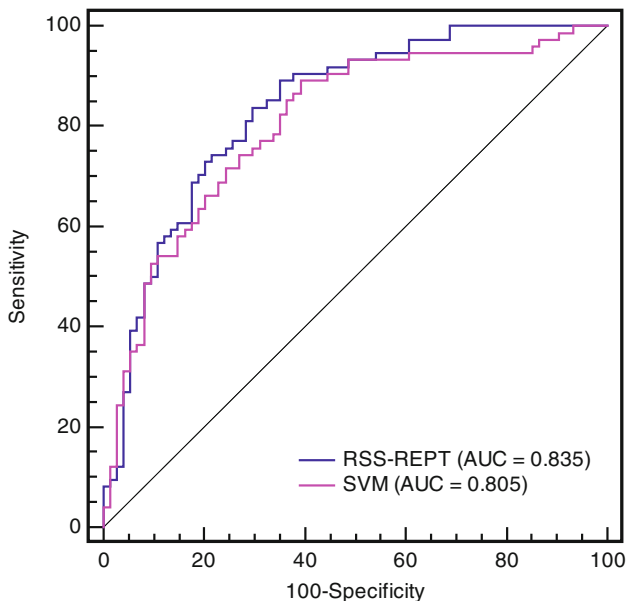


Fig. 6. The ROC curve analysis of landslide models using testing dataset

indicating that the error of training model is relative low; and the AUC value is 0.835 indicating that the general performance of the model is good.

Based on the analysis of the results, it is clearly observed that the proposed hybrid model RSS-REPT has a good performance for landslide susceptibility modeling in the present study, even better than the benchmark model SVM. Outperformance of good hybrid model in comparison to single model has also been observed by other workers [3]. In the present case, the hybrid RSS-REPT model takes advantages of both state of the art machine learning methods of the RSS and REPT. More specifically, the RSS ensemble used in the hybrid model RSS-REPT is known as one of the efficient ensemble techniques which can help in improving the performance of the single models [7, 29]. It has been observed that the RSS ensemble might get the benefits by using random subspaces for building and aggregating the classifiers, especially in the case of datasets with many redundant features [18]. Moreover, the classifiers combined by the RSS in random subspaces might be superior to a single classifier built in the original feature space [18]. In addition, the REPT classifier used in the hybrid model RSS-REPT also has many advantages, which can enhance the predictive capability of the hybrid model, such as (i) the REPT is also a hybrid tree model of the DT and REP which can be utilized efficiently for complex classification problems including landslide prediction, (ii) the sub-trees constructed in the REPT can be adapted to characteristics specific to different subsets of the data; and thus producing more accurate results [19, 30], and (iii) it uses the REP technique which can help in minimizing the error from the variance [31].

6.2 Cartographic Generation of Landslide Susceptibility Map

Landslide susceptibility map is a great tool for land use planners and decision makers in landslide hazard management [32, 33]. In the present study, the landslide susceptibility map has been constructed using the proposed hybrid model RSS-REPT through two main steps. Firstly, landslide susceptibility indexes have been generated for all pixels of the study area. Secondly, these susceptibility indexes have been classified to construct the landslide susceptibility map (Fig. 7) with different susceptibility classes (very low, low, moderate, high, and very high) using geometrical interval method [34, 35].

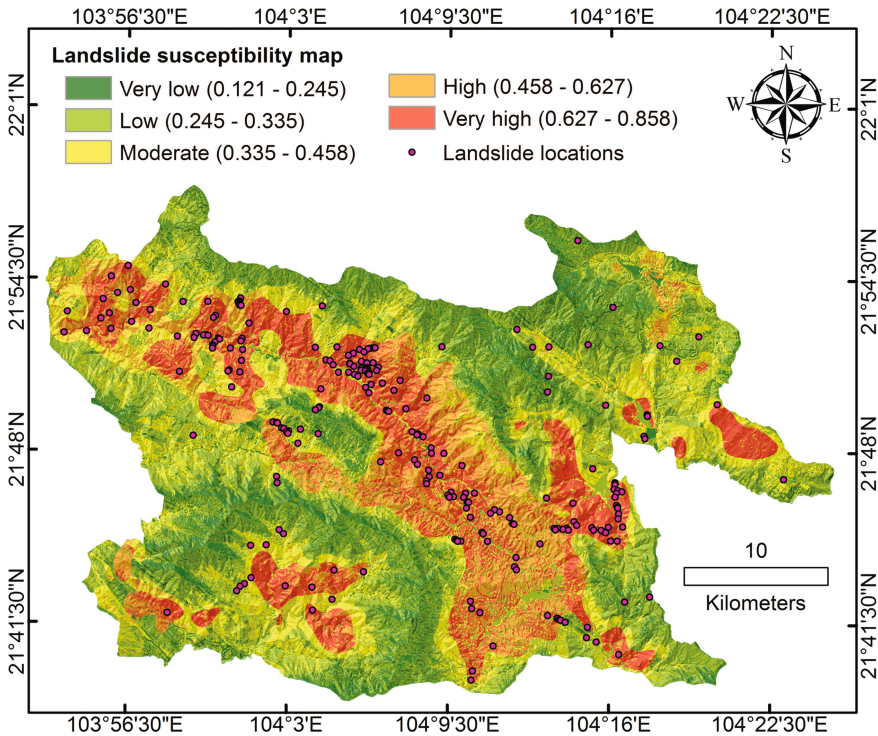


Fig. 7. Landslide susceptibility map of the study area using the RSS-REPT model

To validate the reliability of the landslide susceptibility map produced in this study, the landslide density analysis has been carried out on the basis of the analysis of the distribution of historical landslide pixels on susceptibility classes of the map. The results of landslide density analysis are shown in Table 4. It can be observed that the highest density of landslide pixels is on the very high susceptibility class (4.08), followed by the high susceptibility class (0.98), the moderate susceptibility class (0.62), the low susceptibility class (0.17), and very low susceptibility class (0.07), respectively. These results indicate the reliability of the landslide susceptibility map.

Table 4. Landslide density analysis on landslide susceptibility map

No	Classes	Class pixels	Landslide pixels	% Class pixels	% Landslide pixels	Landslide density
1	Very low	365259	2	12.21	0.81	0.07
2	Low	1107124	16	37.01	6.45	0.17
3	Moderate	660522	34	22.08	13.71	0.62
4	High	367999	30	12.30	12.1	0.98
5	Very high	490283	166	16.39	66.94	4.08

7 Concluding Remarks

Landslide susceptibility modeling has been done at the Mu Cang Chai district, Yen Bai province (Viet Nam) using a novel hybrid model namely RSS-REPT which is a combination of two state of the art machine learning methods of the RSS ensemble and the REPT classifier. For validation, the ROC curve and several statistical indexes have been used. In addition, performance of the proposed hybrid model has been compared with another benchmark single landslide model SVM.

Result analysis shows that the proposed hybrid model RSS-REPT performed well for landslide susceptibility modeling. Its predictive capability outperforms the well-known landslide model SVM. Therefore, it can be reasonably concluded that the proposed hybrid model RSS-REPT is a promising tool for landslide susceptibility modeling which can be considered as a better alternative method for landslide hazard management also in other landslide prone areas.

Acknowledgement. Authors are thankful to the Vietnam Institute of Geosciences and Mineral Resources for sharing the data. Authors are also thankful to the Director, Bhaskarcharya Institute for Space Applications and Geo-Informatics, Gujarat, India for providing facilities to carry out this research work.

References

1. Pham, B.T., Bui, D.T., Dholakia, M.B., Prakash, I., Pham, H.V., Mehmood, K., Le, H.Q.: A novel ensemble classifier of rotation forest and Naïve Bayer for landslide susceptibility assessment at the Luc Yen district, Yen Bai Province (Viet Nam) using GIS. *Geomatics, Natural Hazards and Risk*, pp. 1–23 (2016)
2. Tien Bui, D., Ho, T.-C., Pradhan, B., Pham, B.-T., Nhu, V.-H., Revhaug, I.: GIS-based modeling of rainfall-induced landslides using data mining-based functional trees classifier with AdaBoost, Bagging, and MultiBoost ensemble frameworks. *Environ. Earth Sci.* **75**, 1–22 (2016)
3. Tien Bui, D., Pham, B.T., Nguyen, Q.P., Hoang, N.-D.: 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* **9**, 1–21 (2016)

4. Pham, B.T., Tien Bui, D., Dholakia, M.B., Prakash, I., Pham, H.V.: 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**, 1–18 (2016)
5. Tsangaratos, P., Ilija, I.: Landslide susceptibility mapping using a modified decision tree classifier in the Xanthi Perfection. *Greece Landslides* **13**, 305–320 (2016)
6. Abella, E.A.C., Van Westen, C.J.: Qualitative landslide susceptibility assessment by multicriteria analysis: a case study from San Antonio del Sur, Guantánamo. *Cuba. Geomorphol.* **94**, 453–466 (2008)
7. Shirzadi, A., Bui, D.T., Pham, B.T., Solaimani, K., Chapi, K., Kaviani, A., Shahabi, H., Revhaug, I.: Shallow landslide susceptibility assessment using a novel hybrid intelligence approach. *Environ. Earth Sci.* **76**, 60 (2017)
8. Saha, A.K., Gupta, R.P., Sarkar, I., Arora, M.K., Csaplovics, E.: An approach for GIS-based statistical landslide susceptibility zonation—with a case study in the Himalayas. *Landslides* **2**, 61–69 (2005)
9. Mathew, J., Jha, V., Rawat, G.: Application of binary logistic regression analysis and its validation for landslide susceptibility mapping in part of Garhwal Himalaya, India. *Int. J. Remote Sens.* **28**, 2257–2275 (2007)
10. Pham, B.T., Tien Bui, D., Pourghasemi, H.R., Indra, P., Dholakia, M.B.: 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**, 1–19 (2015)
11. Mathew, J., Jha, V., Rawat, G.: Landslide susceptibility zonation mapping and its validation in part of Garhwal Lesser Himalaya, India, using binary logistic regression analysis and receiver operating characteristic curve method. *Landslides* **6**, 17–26 (2009)
12. Pham, B.T., Pradhan, B., Tien Bui, D., Prakash, I., Dholakia, M.B.: A comparative study of different machine learning methods for landslide susceptibility assessment: a case study of Uttarakhand area (India). *Environ. Model Softw.* **84**, 240–250 (2016)
13. Tsangaratos, P., Benardos, A.: Estimating landslide susceptibility through a artificial neural network classifier. *Nat. Hazards* **74**, 1489–1516 (2014)
14. Pham, B.T., Bui, D.T., Prakash, I., Dholakia, M.: Evaluation of predictive ability of support vector machines and naïve Bayes trees methods for spatial prediction of landslides in Uttarakhand state (India) using GIS. *J. Geomatics* **10**, 71–79 (2016)
15. Pham, B.T., Tien Bui, D., Prakash, I., Dholakia, M.B.: Rotation forest fuzzy rule-based classifier ensemble for spatial prediction of landslides using GIS. *Nat. Hazards* **83**, 1–31 (2016)
16. Ho, T.K.: The random subspace method for constructing decision forests. *IEEE Trans. Pattern Anal. Mach. Intell.* **20**, 832–844 (1998)
17. Xia, J., Dalla Mura, M., Chanussot, J., Du, P., He, X.: Random subspace ensembles for hyperspectral image classification with extended morphological attribute profiles. *IEEE Trans. Geosci. Remote Sens.* **53**, 4768–4786 (2015)
18. Skurichina, M., Duin, R.P.: Bagging, boosting and the random subspace method for linear classifiers. *Pattern Anal. Appl.* **5**, 121–135 (2002)
19. Quinlan, J.R.: Simplifying decision trees. *Int. J. Man Mach. Stud.* **27**, 221–234 (1987)
20. Nefeslioglu, H., Sezer, E., Gokceoglu, C., Bozkir, A., Duman, T.: Assessment of landslide susceptibility by decision trees in the metropolitan area of Istanbul, Turkey. *Mathe. Probl. Eng.* **2010** (2010)
21. Galathiya, A., Ganatra, A., Bhensdadia, C.: Improved decision tree induction algorithm with feature selection, cross validation, model complexity and reduced error pruning. *Int. J. Comput. Sci. Inf. Technol.* **3**, 3427–3431 (2012)

22. Pham, B.T., Tien Bui, D., Indra, P., Dholakia, M.: 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**, 338–344 (2015)
23. Gorsevski, P.V., Gessler, P.E., Foltz, R.B., Elliot, W.J.: Spatial prediction of landslide hazard using logistic regression and ROC analysis. *Trans. GIS* **10**, 395–415 (2006)
24. 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.: Characterising performance of environmental models. *Environ. Model Softw.* **40**, 1–20 (2013)
25. Marjanović, M., Kovačević, M., Bajat, B., Voženilek, V.: Landslide susceptibility assessment using SVM machine learning algorithm. *Eng. Geol.* **123**, 225–234 (2011)
26. Kavzoglu, T., Sahin, E.K., Colkesen, I.: Landslide susceptibility mapping using GIS-based multi-criteria decision analysis, support vector machines, and logistic regression. *Landslides* **11**, 425–439 (2014)
27. Wu, Y., Li, W.: GIS-based landslide susceptibility analysis using support vector machine model at a regional scale. *Electron. J. Geotech. Eng.* **21**, 6938–6945 (2016)
28. NCEP: Global weather data for SWAT (2014). <http://globalweather.tamu.edu/home>
29. Pham, B.T., Bui, D.T., Prakash, I.: Landslide susceptibility assessment using bagging ensemble based alternating decision trees, logistic regression and j48 decision trees methods: a comparative study. *Geotech. Geol. Eng.*, 1–15 (2017)
30. Brodley, C.E., Utgoff, P.E.: Multivariate decision trees. *Mach. Learn.* **19**, 45–77 (1995)
31. Tama, B.A., Rhee, K.-H.: Tree-based classifier ensembles for early detection method of diabetes: an exploratory study. *Artif. Intell. Rev.*, 1–16 (2017)
32. Pham, B.T., Tien Bui, D., Pham, H.V., Le, H.Q., Prakash, I., Dholakia, M.B.: Landslide hazard assessment using random subspace fuzzy rules based classifier ensemble and probability analysis of rainfall data: a case study at Mu Cang Chai District, Yen Bai Province (Viet Nam). *J. Indian Soc. Remote Sens.*, 1–11 (2016)
33. Pham, B.T., Tien Bui, D., Prakash, I., Nguyen, L.H., Dholakia, M.B.: A comparative study of sequential minimal optimization-based support vector machines, vote feature intervals, and logistic regression in landslide susceptibility assessment using GIS. *Environ. Earth Sci.* **76**, 371 (2017)
34. Frye, C.: About the geometrical interval classification method (2007). <http://blogs.esri.com/esri/arcgis>
35. Pham, B.T., Tien Bui, D., Prakash, I., Dholakia, M.B.: Hybrid integration of multilayer perceptron neural networks and machine learning ensembles for landslide susceptibility assessment at Himalayan area (India) using GIS. *CATENA* **149**(Part 1), 52–63 (2017)