



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)

Binh Thai Pham^{1,2} · Dieu Tien Bui³ · Ha Viet Pham⁴ · Hung Quoc Le⁴ ·
Indra Prakash⁵ · M. B. Dholakia⁶

Received: 29 February 2016 / Accepted: 8 September 2016 / Published online: 17 September 2016
© Indian Society of Remote Sensing 2016

Abstract Landslide hazard assessment at the Mu Cang Chai district; Yen Bai province (Viet Nam) has been done using Random SubSpace fuzzy rules based Classifier Ensemble (RSSCE) method and probability analysis of rainfall data. RSSCE which is a novel classifier ensemble method has been applied to predict spatially landslide occurrences in the area. Prediction of temporally landslide occurrences in the present study has been done using rainfall data for the period 2008–2013. A total of fifteen landslide influencing factors namely slope, aspect, curvature, plan curvature, profile curvature, elevation, land use, lithology, rainfall, distance to faults, fault density, distance to roads, road density, distance to rivers, and river density have been utilized. The result of the analysis shows that

RSSCE and probability analysis of rainfall data are promising methods for landslide hazard assessment. Finally, landslide hazard map has been generated by integrating spatial prediction and temporal probability analysis of landslides for the land use planning and landslide hazard management.

Keywords Landslide hazard assessment · GIS · Fuzzy unordered rules induction algorithm · Random SubSpace · Viet Nam

Introduction

Landslide is a geo-environmental hazard of mainly hilly and mountainous regions. It is often triggered by natural causes such as heavy and prolonged rainfall or snowmelt, earthquakes or volcanic eruptions, and anthropogenic factors such as ground excavation, deforestation and land use changes (Guzzetti et al. 2005). Out of these factors, rainfall is considered one of the most prominent triggering factors of landslide occurrences in Viet Nam (Kjekstad and Highland 2009). Landslides in this area often cause hundreds of fatalities and loss of millions of US dollar every year (Pham et al. 2016a). Therefore, now landslide studies have been turning into urgent tasks requiring landslide hazard assessment.

Landslide hazard is defined as the probability of occurrences of potential landslide in specified period of time (Varnes 1984). Therefore, in the landslide hazard assessment, spatial prediction and temporal probability of landslide occurrences are considered. Most of these studies are related to spatial prediction of landslides (Alkhasawneh et al. 2014; Pham et al. 2016a). However, only few attempts have been made to establish temporal probability

✉ Binh Thai Pham
phambinhgtvt@gmail.com; binhpt@utt.edu.vn

¹ Department of Civil Engineering, Gujarat Technological University, Nr. Visat Three Roads, Visat-Gandhinagar Highway, Chandkheda, Ahmedabad, Gujarat 382424, India

² Department of Geotechnical Engineering, University of Transport Technology, 54 Trieu Khuc, Thanh Xuan, Hanoi, Vietnam

³ Geographic Information System Group, Department of Business Administration and Computer Science, Telemark University College, Hallvard Eikas Plass 1, 3800 Bø i Telemark, Norway

⁴ Vietnam Institute of Geosciences and Mineral Resources, Thanh Xuan, Hanoi, Vietnam

⁵ Department of Science and Technology, Bhaskarcharya Institute for Space Applications and Geo-Informatics (BISAG), Government of Gujarat, Gandhinagar, India

⁶ Department of Civil Engineering, LDCE, Gujarat Technological University, Ahmedabad, Gujarat 380015, India

of landslides (Guzzetti et al. 2005; Tien Bui et al. 2013) mainly due to limitation of availability of site specific data such as exact time, magnitude and velocity of mass movements.

For spatial prediction of landslides, three approaches are generally considered namely analytical methods, expert's opinion based methods, and machine learning methods (Pradhan 2013). Out of these, machine learning methods which are based on statistical analysis of the spatial relationship between a set of geo-environmental factors and landslide occurrences are known more effective for spatial prediction of large regions (Pradhan 2013). Kavzoglu et al. (2014), and Pradhan (2013) used support vector machines as an effective machine learning method for landslide prediction. Moreover, Pham et al. (2015) and Choi et al. (2012) applied successfully artificial neural networks machine learning method for landslide susceptibility assessment. Other machine learning methods such as decision trees (Alkhasawneh et al. 2014), logistic regression (Akgun 2012) have also been applied widely for spatial prediction of landslides. In general, performance of these methods is good; however, it can be further improved using ensemble techniques as it uses machine learning algorithms to combine multiple classifiers (Pham et al. 2016b).

Regarding temporal probability of landslides, two approaches can be applied such as analysis of potential slope instability and statistical analysis of past landslide events (Saez et al. 2012). The first approach is based on evaluation of impact of the current slope conditions to the potential instability of slopes; however, this approach is not applicable for large investigated regions (Tien Bui et al. 2013). The second approach is based on the probability analysis of past landslide events based on historical landslides records especially of rainfall induced landslide events (Corominas and Moya 2008). Out of these, statistical analysis approach based on information of rainfall induced landslide events is considered more suitable for temporal probability of landslides (Tien Bui et al. 2013).

In term of landslide hazard assessment, Guzzetti et al. (2005) employed probabilistic model, Terlien et al. (1995) used the deterministic model. In another study, Tien Bui et al. (2013) utilized machine learning methods such as support vector machines, logistic regression, evidential belief functions, bayesian neural networks, and neuro-fuzzy integrated with probability analysis of rainfall data to assess landslide hazard in Hoa Binh province (Viet Nam), and stated that machine learning methods integrated with probability analysis of rainfall data is a promising approach for landslide hazard assessment in landslide prone areas. Therefore, main objective of the present study is to assess landslide hazard at the Mu Cang Chai district, Yen Bai

province (Viet Nam) using a novel machine learning method of Random SubSpace fuzzy rules based Classifier Ensemble (RSSCE) in conjunction with probability analysis of rainfall data. RSSCE based on Fuzzy Unordered Rules Induction Algorithm (FURIA) classifier and Random SubSpace (RSS) ensemble has been proposed to predict spatially landslide occurrences. Probability analysis of available rainfall data for the period 2008–2013 has been considered for the prediction of temporally landslide occurrences.

Random SubSpace Fuzzy Rules Based Classifier Ensemble (RSSCE) Method for Landslide Spatial Prediction

Spatial prediction of landslide hazards in the present study has been carried out by RSSCE method. It is a combination of Fuzzy Unordered Rules Induction Algorithm (FURIA) classifier and Random SubSpace (RSS) ensemble. FURIA was first introduced by Hühn and Hüllermeier (2009) which is an extension of well-known RIPPER algorithm (Cohen 1995). It uses fuzzy rules and unordered rule sets to learn the classified algorithm. In addition, FURIA also uses rule stretching method to solve uncovered cases (Hühn and Hüllermeier 2009). Therefore, it usually results in higher accuracy than the RIPPER algorithm, and C4.5 classifier (Hühn and Hüllermeier 2009). Meanwhile, RSS is known as one of the most efficient ensemble techniques that have been utilized to improve performance of the individual classifiers (Onan 2015). RSS was first proposed by Ho (1998) that could combine multiple classifiers for training in modified feature space. It also creates optimally the number of training subsets that are employed to train base classifiers (Ho 1998). Therefore, RSS is known as an efficient ensemble method in dealing with datasets of many redundant features and over-fitting problems (Onan 2015). The proposed RSSCE method in the present study takes advantages of these two techniques that could result desired outcomes for spatial prediction of landslide hazards which involves three main steps (1) initiation, (2) optimization, and (3) classification.

Initiation The imitative step is to generate the input data which has been generated from data collected from the Mu Cang Chai district. It included training dataset and testing dataset. Meanwhile, training dataset has been created using 174 landslide pixels (70 % historical events) and 174 non-landslide pixels, testing dataset has been generated using 74 landslide pixels (30 % other historical events) and 74 non-landslide pixels. All fifteen landslide influencing factors have been utilized to

overlay with these landslide and non-landslide pixels for creating final datasets. Basically, training dataset is used to construct the RSSCE model whereas testing dataset is employed to validate predictive capability of the RSSCE model.

Optimization In this step, the RSS method has been applied to divide training dataset into optimal sub-training datasets that are then used to train the base classifier. Main theory of the RSS method is based on stochastic discrimination (Kleinberg 1990) that is applied to partition of the feature spaces, and then to construct classifiers based on the combination of many components that have weak discriminative capability but good penalization (Gao and Wang 2006). In final step, the RSS method has been then used to combine results of all classifiers that use sub-training datasets to give final outcomes of the RSSCE model.

Classification This step is carried out to classify landslide or non-landslide variables for spatial prediction of landslide hazards, it uses the FURIA method to analyze spatial relationship between landslide occurrences and a set of geo-environmental factors using optimal sub-training datasets obtained from the optimization step. In this step, fuzzy rules have been applied using trapezoidal membership function as following equation (Hühn and Hüllermeier 2009).

$$I^F(\mu) \stackrel{\text{df}}{=} \begin{cases} 1 & \psi^{c,L} \leq \mu \leq \psi^{c,U} \\ \frac{\mu - \Phi^{s,L}}{\Phi^{c,L} - \Phi^{s,L}} & \Phi^{s,L} \leq \mu \leq \Phi^{c,L} \\ \frac{\Phi^{s,U} - \mu}{\Phi^{c,U} - \Phi^{s,U}} & \Phi^{c,U} \leq \mu \leq \Phi^{s,U} \\ 0 & \text{else} \end{cases} \quad (1)$$

where $\Phi^{c,U}$ and $\Phi^{c,L}$ are the upper and lower bounds of the fuzzy set corresponding to unit elements membership, respectively. Also, $\Phi^{s,U}$ and $\Phi^{s,L}$ are the upper and lower bounds of the fuzzy set respective with elements membership bigger than zero.

Notably, while applying the RSSCE model, it can be observed that performance of the RSSCE model depends significantly on the selection of learning parameters such as number of folds that is used to determine the amount of data for reduced-error pruning of the FURIA classifier, and the number of iterations that is used to learn the RSSCE model. Therefore, optimization task of these parameters has been carried out to obtain the best performance of the RSSCE model using the trial-and-error process (Pham et al. 2015). As a result, in the present study the number of folds is set to 5, and the number of iterations is set to 13 to train the RSSCE model.

Study Area

The Mu Cang Chai district (Lat. 21°39'00"N–21°50'00"N; Long. 103°56'00"E–104°23'00"E) affected by landslides have been selected as the study area. It is located in the northwest of Yen Bai province of Viet Nam (Fig. 1), covering an area of about 1196.47 km². The area is situated in a tropical monsoon region having heavy rainfall (average 3700–5490 mm) during the months May–October. The annual mean humidity in this area is about 81 % and mean temperature 14.3 °C. Most of the study region is covered by forests (61.76 %).

Topography of the area is hilly with intervening deep valleys. Mountains in the region occupy steep slopes up to 88 degrees at places and elevation of the area varies from 280 to 2820 m with the average elevation of about 1515 m (above standard sea level). The area is occupied by igneous, sedimentary and metamorphic rocks. The volcanic extrusive rocks of Tu Le and Ngoi Thia complexes and intrusive igneous rocks of Phu Sa Phin complex and Tram Tau formation are predominant in the area. The area is tectonically active and dissected by three main faults namely Phong Tho—Van Yen, Nam Co—Minh An, and Nghia Lo.

Landslide Inventory

Data of landslide inventory is essential in landslide hazard assessment (Tien Bui et al. 2013). In this study, landslides data has been obtained from the Vietnam Institute of Geosciences and Mineral Resources under the national project namely “Survey, assessment and zoning of landslide warning in the mountainous region of Vietnam”. In all 248 landslide locations have been identified by interpreting air photos (Year 2013) on 1:33,000 scales. These landslides have been subsequently validated by field investigations. Landslides in study area are having varying sizes small (<200 m³), average (200–1000 m³), large (1000–20,000 m³), very large (20,000–100,000 m³). The biggest landslide event occurred at the Che Cu Na commune in February, 2011 having volume size of 10,000 m³. The volume of landslides has been determined through field investigation and spatial analysis. In the study area translational (35 locations), rotational (124), toppling (45), debris and mixed (36) type of landslides have been observed. Maximum number of landslides is of rotational type (124). Analysis of the landslide inventory shows that most of landslides in the study region occurred during rainy season (May–October). Specific date of land slide occurrence is available only of 42 landslides in the available record. Therefore, these landslide events have been employed for temporal prediction of landslides and 248 landslide events have been utilized for spatial prediction of landslides.

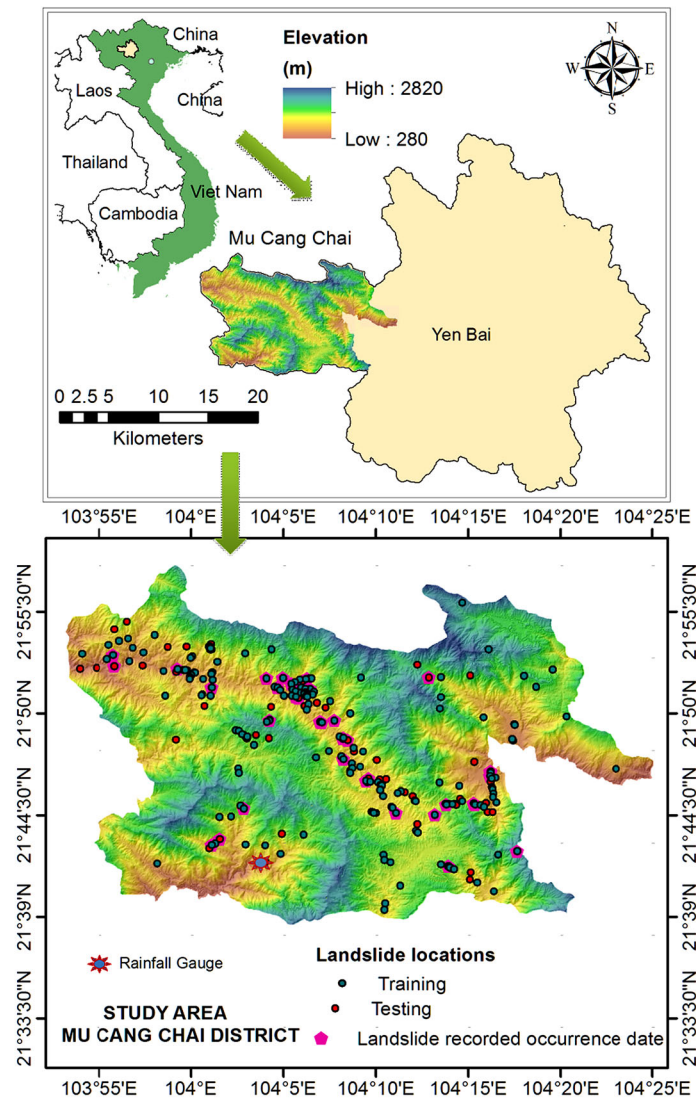


Fig. 1 Location map and landslide photos of the study area (courtesy: Yen Bai and Dan Tri newspapers)

Spatial Prediction of Landslides Hazards

Geo-environmental Factors in Relation with Landslide Occurrences

In spatial prediction of landslides, spatial relationship between geo-environmental factors and landslide occurrences is often analyzed based on the assumption that future landslides will occur under identical conditions of past landslides (Pham et al. 2015). Thus determination of geo-environmental factors that affected past landslide occurrences is very important. Based on the analysis of mechanism of landslide occurrences and geo-environmental characteristics of the study region, a total of fifteen geo-environmental factors (slope, aspect, curvature, plan curvature, profile curvature, elevation, land use, lithology, rainfall, distance to faults, fault density, distance to roads, road density, distance to rivers, and river density) have been considered as landslide affecting factors in the present study. Maps of these affecting factors have been constructed as raster data (pixel size of 20×20 m) with different classes (Table 1; Fig. 2) based on the degree of susceptibility of each class to landslide occurrences. These classes are based on the experience of analysis of adjacent area carried out by authors Pham

et al. (2016a) and other workers Cevik and Topal (2003) and Dai and Lee (2002).

Evaluation of Predictive Capability of the RSSCE Model

In literature, Receiver Operating Characteristic (ROC) curve method has been utilized as a standard quantitative method to evaluate the predictive capability of landslide models (Pham et al. 2016a). Therefore, in the present study, the ROC curve has been selected to validate the performance of the RSSCE model. Basically, the ROC curve is generated by plotting pairs of two statistical indexes such as “sensitivity” and “100-specificity” (Tien Bui et al. 2016). The AUC value is area under the ROC curve that is employed to evaluate quantitatively the performance of landslide models (Pham et al. 2015). Additionally, statistical indexes namely accuracy (ACC), kappa (k), and root mean squared error (RMSE) have also been used to evaluate the performance of landslide models (Bennett et al. 2013).

Furthermore, other benchmark landslide models such as Support Vector Machines (SVM) (Vapnik 1995), Multiple Perceptron Neural Network (MLPN Nets) (Zare et al. 2013), and Logistic Regression (LR) (Akgun 2012) have

Table 1 Geo-environmental factors and their classes

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

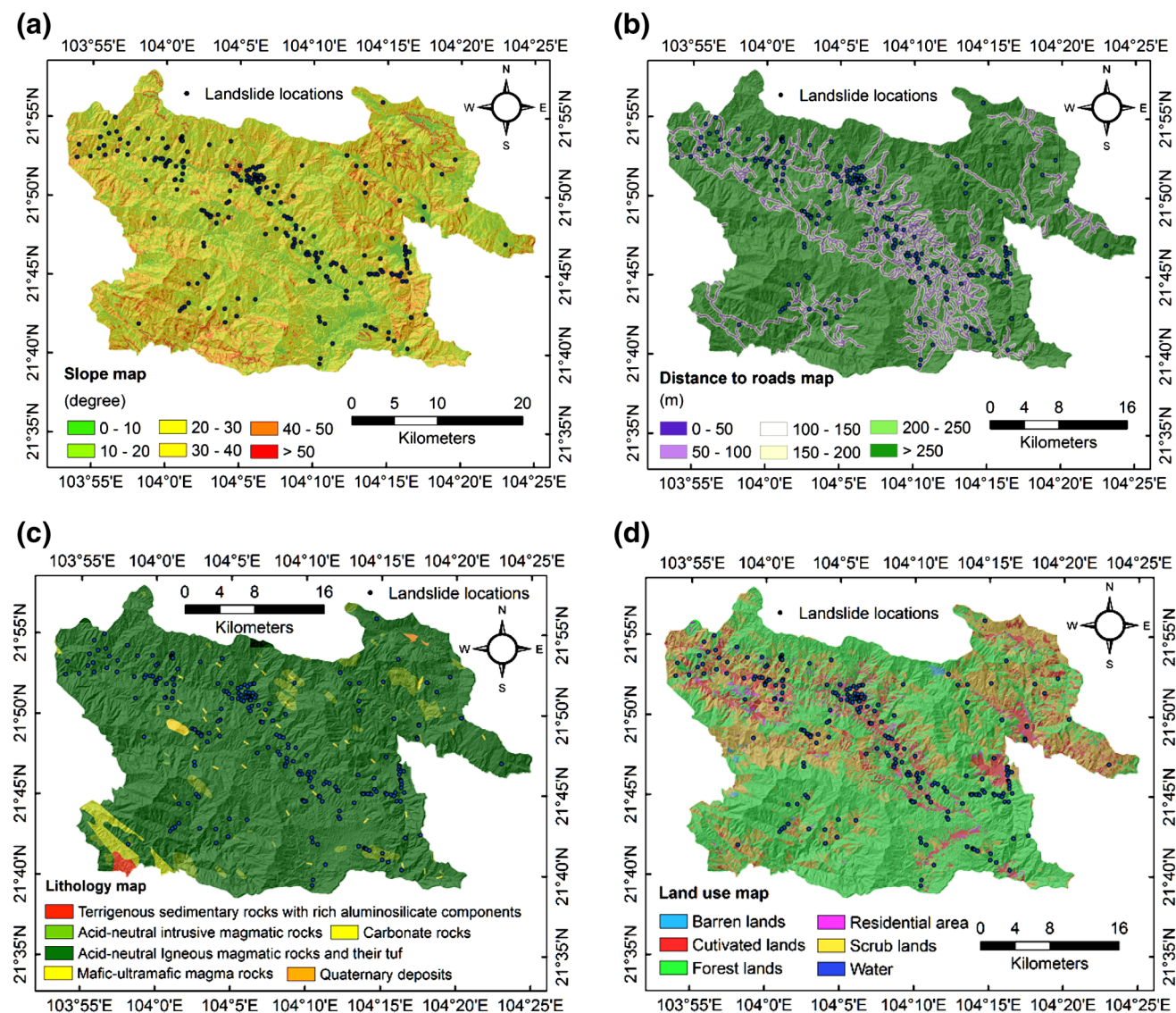


Fig. 2 Thematic maps in the study area: **a** slope map, **b** distance to roads map, **c** lithology map, and **d** land use map

been utilized for comparison with the RSSCE model. More specifically, SVM is known as one of the most efficient machine learning methods for landslide prediction, it is based on the statistical approach of finding an optimal hyper-plane for separating two classes (landslide and non-landslide) (Pourghasemi et al. 2013). Meanwhile, MLPN Nets is one of artificial neural networks which are known as a branch of artificial intelligence has been applied efficiently in landslide problems (Pham et al. 2015). LR is known as a more accurate machine learning method compared to conventional methods (Akgun 2012; Choi et al. 2012).

Results of the performance of the RSSCE model and other benchmark landslide models are shown in Fig. 3 and Table 2. It can be observed that the RSSCE model has the highest predictive capability of spatial prediction of

landslides compared to other benchmark landslide models including the LR model, the SVM model, and the MLPN Nets model. It proves that the RSSCE model is the best choice for spatial prediction of landslides in the present study. Therefore, the results of spatial prediction of landslides from the RSSCE model have been used for landslide hazard assessment.

Temporal Prediction of Landslide Hazards

Rainfall Data Analysis

In the present study, rainfall data have taken into account as a time related factor to analyze temporally landslide occurrences. Rainfall data have been collected from the

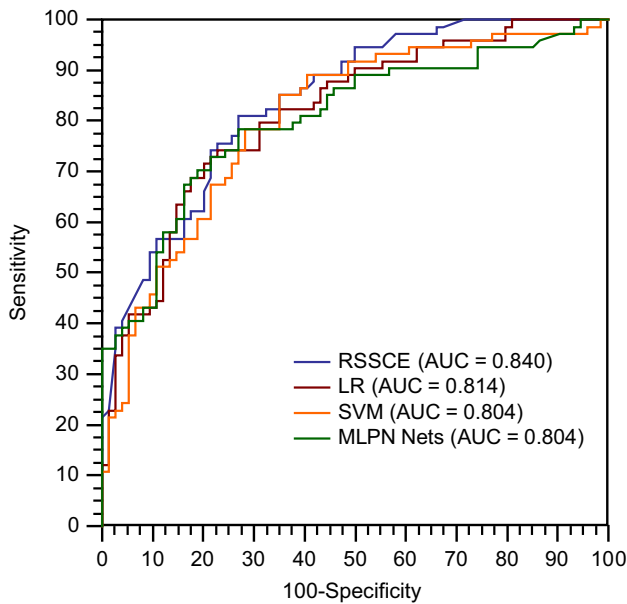


Fig. 3 Analysis of the ROC curve of different landslide models

Table 2 Statistical index values of different landslide models

| No. | Statistical indexes | RSSCE | LR | SVM | MLPN nets |
|-----|---------------------|--------|--------|--------|-----------|
| 1 | ACC (%) | 0.758 | 0.756 | 0.723 | 0.750 |
| 2 | k | 0.5135 | 0.5130 | 0.4459 | 0.500 |
| 3 | RMSE | 0.4159 | 0.4192 | 0.4261 | 0.4644 |

rainfall gauge located in the Mu Cang Chai district, Yen Bai province (Viet Nam). Available Rainfall data for the period 2008–2013 obtained from Global Weather data for SWAT (NCEP 2014) has been analysed. The daily rainfall, in the study area, is shown in Fig. 4a. It can be observed that most intense rainfall usually occurs for short period that is within few days (Fig. 4a). Analysis also shows that the highest annual rainfall occurred in the year 2008 (4362 mm), followed by 2009 (3522 mm), 2010 (3493 mm), 2012 (2160 mm), and 2013 (1950 mm) 2011 (1748 mm), respectively. Out of 248 landslide locations, at 42 locations intense rainfall (more than 100 mm) occurred in a single day (Table 3).

Determination of Rainfall Threshold

In general, determination of rainfall threshold is required for the temporal prediction of landslide (Tien Bui et al. 2013). Rainfall threshold is the minimum rainfall at which a landslide might happen in a certain region (Guzzetti et al. 2007). In recent decades, many methods have been developed to calculate the rainfall threshold for landslide study, these methods can be grouped into five approaches namely (1) physical-based approach, (2) empirical based approach, (3) intensity–duration based approach, (4) normalized

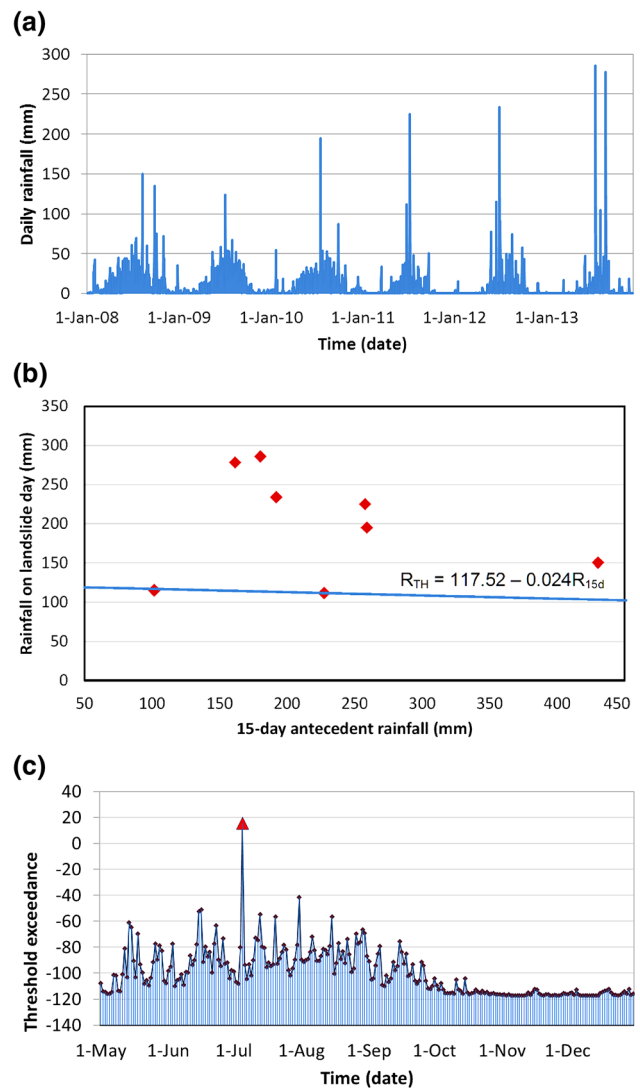


Fig. 4 Rainfall analysis at Mu Cang Chai district: **a** daily rainfall for the period of 2008–2013, **b** the rainfall threshold, **c** validation of the rainfall threshold (the red mark indicates threshold exceeded rainfall) (color figure online)

intensity–duration based approach, and (5) antecedent rainfall based approach (Guzzetti et al. 2008). Out of these approaches, the intensity–duration based approach is known as the most widely used method (Larsen and Simon 1993; Aleotti 2004). This approach requires the data of intensity of rainfall during the day on which landslide occurred (Larsen and Simon 1993). This data is generally not available in most of the cases. The antecedent rainfall plays important role in the initiation of landslides as it increases the pore-water pressure in the slope formed materials (Tien Bui et al. 2013). Therefore, the antecedent rainfall based approach has been utilized to determine the rainfall threshold for temporal prediction of landslides in the present study.

Rainfall threshold in the study area has been determined based on the experience of other adjacent areas even

Table 3 Temporal occurrence of rainfall triggered landslides in the Mu Cang Chai district from 2008 to 2013

| No. | Date of landslides | Number of landslides | Daily rainfall (mm) | 3 days rainfall (mm) | 5 days rainfall (mm) | 7 days rainfall (mm) | 10 days rainfall (mm) | 15 days rainfall (mm) | 30 days rainfall (mm) |
|-----|--------------------|----------------------|---------------------|----------------------|----------------------|----------------------|-----------------------|-----------------------|-----------------------|
| 1 | 9/8/2008 | 2 | 150.5985 | 157.364 | 207.085 | 251.075 | 286.034 | 430.219 | 769.220 |
| 2 | 5/7/2009 | 3 | 124.235 | 37.55092 | 60.82649 | 80.2637 | 141.0215 | 268.2655 | 611.5101 |
| 3 | 01/7/2010 | 4 | 195.168 | 80.313 | 112.174 | 152.779 | 205.144 | 258.936 | 509.060 |
| 4 | 25/06/2011 | 2 | 112.098 | 47.904 | 96.105 | 135.415 | 147.881 | 227.402 | 317.381 |
| 5 | 08/07/2011 | 3 | 225.117 | 13.843 | 23.785 | 43.698 | 56.154 | 257.768 | 482.405 |
| 6 | 15/06/2012 | 7 | 115.098 | 30.981 | 32.695 | 53.668 | 90.871 | 101.658 | 331.705 |
| 7 | 28/06/2012 | 5 | 234.054 | 0.915 | 13.364 | 21.267 | 49.241 | 192.038 | 282.914 |
| 8 | 15/07/2013 | 8 | 286.035 | 114.310 | 117.788 | 126.568 | 129.192 | 180.137 | 277.123 |
| 9 | 24/08/2013 | 8 | 278.235 | 46.96312 | 63.26751 | 115.8869 | 125.8879 | 161.487 | 497.6292 |

though correlation between the numbers of days for the antecedent rainfall and the triggering of a landslide is relatively complex (Guzzetti et al. 2007). Aleotti (2004) considered rainfall of 10 and 15 days. In general, no agreement has been reached to select the exact number of days for the antecedent rainfall in determining the rainfall threshold. Therefore, the selection of the number of days is often based on the analysis of the rainfall data at the time landslide occurred for different number of days (Tien Bui et al. 2013). In the present study, using the results from the study carried out in the Hoa Binh province which is an adjacent region of the study area (the Mu Cang Chai district), the number of days has been utilized as 15 for analyzing the rainfall threshold (Tien Bui et al. 2013). Data of the rainfall-induced landslides of 5 years (2008, 2010, 2011, 2012, and 2013) has been utilized to determine the rainfall threshold whereas the rainfall-induced landslides of 2009 have been utilized for the validation of the rainfall threshold. Finally, the envelope line for landslide occurrences (Fig. 4b) has been determined using two lowest points in the scattered graph (Tien Bui et al. 2013) which is expressed as following mathematical equation:

$$R_{TH} = 117.52 - 0.024R_{15d} \quad (2)$$

where R_{TH} is defined as the rainfall threshold; R_{15d} is inferred as the antecedent rainfall of 15 days.

Evaluation of the Rainfall Threshold

Evaluation of the performance of landslide models is to be done by dividing them into two subsets for training and validation of the models (Chung and Fabbri 2003). Therefore, the recorded rainfall induced landslide events, in the present study, have been divided into two parts. One part includes induced landslide events of the years of 2008, 2010, 2011, 2012, and 2013 which have been utilized to determine the rainfall threshold. Another part includes

landslide events occurred in the year 2009 which have been used for of the validation of rainfall threshold.

Analysis of results shows that there is only one day rainfall exceeded the threshold value that is on July 5, 2015 (Fig. 4c). This result can be correlated with the landslide events recorded during the year 2009 (Table 3). Thus rainfall threshold obtained from the present study can be used to analyze the temporal prediction of landslides for landslide hazard assessment in this area.

Temporal Probability of Landslide Occurrences for Landslide Hazard Assessment

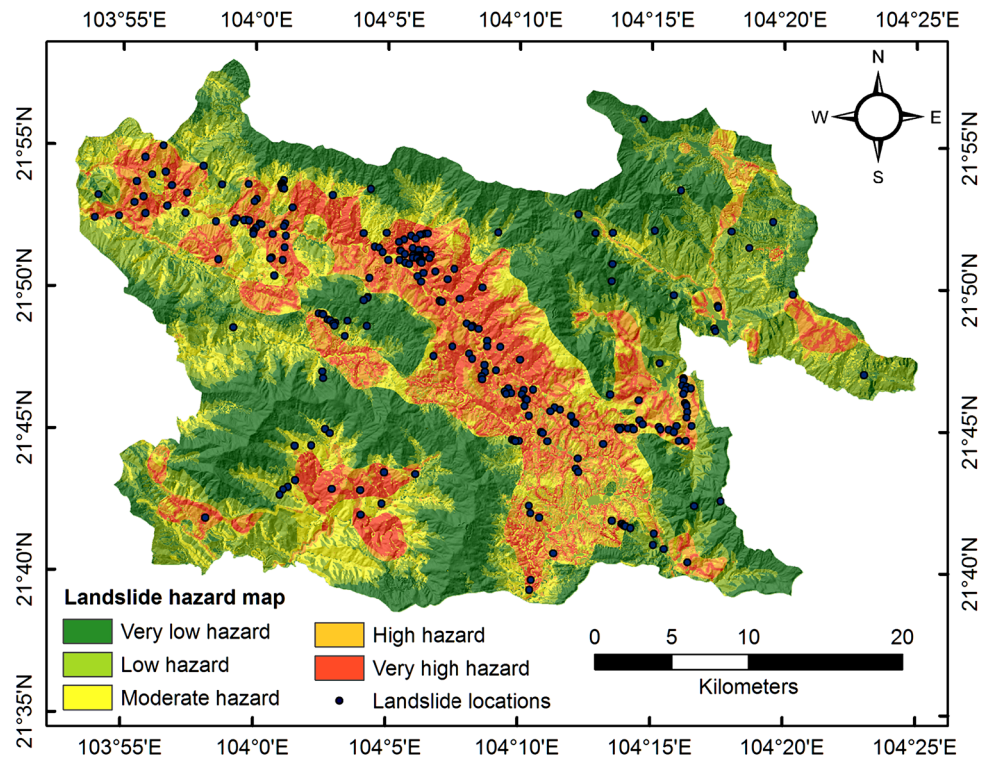
The temporal probability of landslide occurrences is determined based on the assumption that the past landslide events could be considered as independent random point-events in time (Guzzetti et al. 2005). Therefore, the exceeded probability of landslide occurrences during time “t” is expressed as below (Guzzetti et al. 2005):

$$P_L = P[L(t) \geq 1] \quad (3)$$

where $L(t)$ is the number of landslide events that occur during time “t” in the study region.

In order to determine the exceeded probability of landslide occurrences during time “t”, there are two common methods utilized namely poisson and binomial methods (Crovelli 2000). Out of these methods, poisson is a continuous time method that is based on the independent relationship between the occurrences of random-point events and time (Coe et al. 2000) whereas binomial is a discrete-time method that consists of the occurrence of random-point events in certain time (Coe et al. 2000). In comparison of these two methods, Crovelli (2000) stated that these are quite different in the case of short periods (t is short), but quite coincided in case of long periods (t is long). Moreover, Poisson method is more commonly used for landslide hazard assessment (Guzzetti et al. 2005; Coe

Fig. 5 Landslide hazard map of the study area for the return period of 5 years



et al. 2000). Thus in the present study the poisson method has been adopted to analyze the temporal probability of landslide occurrences.

In the Poisson method (Guzzetti et al. 2005), the probability of “m” landslides during time “t” can be estimated as below:

$$P[L(t) = m] = \exp(-\beta t) \frac{(\beta t)^m}{m!}; \quad m = 1, 2, 3, \dots \quad (4)$$

where β is the estimated rate of occurrence of landslides, it could be obtained from the catalogue of historical landslide events (Guzzetti et al. 2005).

Based on the Eq. (4), the exceeded probability (the probability of one or more landslides happened during time “t”) can be estimated as following equation:

$$P[L(t) \geq 1] = 1 - \exp\left(-\frac{t}{\eta}\right) \quad (5)$$

where η is the future average recurrence interval, t is a period of time in the future at which the exceeded probability is calculated.

Using the poisson method, the temporal probability of landslide hazards for the Mu Cang Chai district have been calculated for the return period of 1, 3, and 5 years. The results show that the number of times at which rainfall exceeded the rainfall threshold is 10, and the probability of landslide hazards increases with the return period that is 0.865 in 1 year, 0.998 in 3 years, and 1.000 in 5 years.

Landslide Hazard Assessment

Landslide hazard assessment has been carried out in consideration of both the spatial prediction and temporal prediction of landslides in the study area. Based on this, landslide hazard maps have been constructed in three main steps such as (1) generating landslide susceptibility indexes, (2) calculating landslide hazard indexes by multiplying landslide susceptibility indexes with temporal probability of landslides for different periods, and (3) reclassifying landslide hazard indexes. In the first step, using the spatial prediction results by applying the RSSCE model, landslide susceptibility indexes have been generated for all pixels of whole study area. In the next step, the temporal probability in different return periods (1, 3, 5 years) obtained from the temporal prediction have been multiplied individually with landslide susceptibility indexes for obtaining landslide hazard indexes for all pixels of entire study area. In the final step, landslide hazard indexes have been reclassified into five intervals using natural breaks method which is widely applied in landslide studies (Pham et al. 2015). Based on the hazard index intervals, hazard has been classified from very high, high, low and very low hazards. In the development of landslide hazard map highest temporal probability observed in the 5 years period has been used for the study area (Fig. 5).

Discussions

Landslide hazard assessment has been carried out in the present study using Random SubSpace fuzzy rules based Classifier Ensemble (RSSCE) method and probability analysis of rainfall data. Out of these, RSSCE is novel hybrid approach of Fuzzy Unordered Rules Induction Algorithm (FURIA) classifier and Random SubSpace (RSS) ensemble which has been proposed to predict spatially landslide occurrences. Probability analysis of rainfall data has been utilized to predict temporally landslide occurrences for the Mu Cang Chai district, Yen Bai province (Viet Nam) which is highly landslide prone area.

It has been observed that for the spatial prediction of landslides occurrences, the RSSCE model performed well in the present study (AUC = 0.840) in comparison to other models such as LR model (AUC = 0.810), SVM model (AUC = 0.804) and the MLPN Nets model (AUC = 0.804). This result is expected as the classifier ensemble method of RSSCE uses the RSS ensemble which has capability in improvement of the performance of individual classifiers (Onan 2015). Moreover, RSSCE uses learning dataset which are optimized during training process for classification thus significantly improving its predictive capability in comparison to other landslide models.

For temporal probability of landslide occurrences, time factors such as rainfall and landslide frequency usually take into account (Tien Bui et al. 2013). In the present study, landslide frequency data is not available in landslide inventory. Therefore, rainfall has been utilized to analyze the temporal prediction of landslides for the study area. Based on the analysis of temporal relationship between the rainfall data and the historical landslide events during the period 2008–2013, temporal probability of landslide occurrences has been analyzed for different periods (1, 3, 5 years). The temporal prediction of landslide occurrences has been done in four main steps (1) determining the rainfall threshold by analysis the temporal relationship between the antecedent rainfall of 15 days and past landslide events, (2) evaluating the rainfall threshold and (3) calculating the temporal probability of landslide occurrences using poison method. This methodology can be applied in other areas also where multi-temporal landslide inventory is not properly available.

Landslide hazard assessment has been accomplished in the study area by the consideration of both spatial and temporal prediction of landslide occurrences. Past and present landslide events and geoenvironmental conditions have been considered in the study. In general, geoinfluencing factors for short period of 5 years may not be of significance. However, it is desirable to consider some dynamic factors such as changes in the slope due to road

cutting, and change in the land use pattern for development even for short period analysis.

Conclusions

Landslide is a common geo-environmental hazard in hilly and mountainous areas in Viet Nam especially during rainy season. Therefore, rainfall is considered as main triggering factors to landslide occurrences in the study area. Landslide hazard assessment for the Mu Cang Chai district, Vietnam has been carried out in consideration of both spatial and temporal prediction of landslide occurrences. In the present study, a novel classifier ensemble method called Random SubSpace fuzzy rules based Classifier Ensemble (RSSCE) has been applied to predict spatially landslide occurrences, and probability analysis of rainfall data for the period 2008–2013 has been considered to predict temporally occurrences of landslides in the area.

Result of present study show that a hybrid approach of RSSCE method and the probability analysis of rainfall data is a promising approach for landslide hazard assessment which can be applied in other landslide prone areas where multi-temporal landslide inventory is not adequately available. Landslide hazard map developed for the study area would be of use for land use planning and proper landslide hazard management.

Acknowledgments Authors would like to sincerely thank to the Vietnam Institute of Geosciences and Mineral Resources for sharing the data for the present study. Authors are also thankful to the Director, Bhaskarcharya Institute for Space Applications and Geo-Informatics (BISAG), Department of Science and Technology, Government of Gujarat, Gandhinagar, Gujarat, India for the encouragement and for providing facilities to carry out this research work.

References

- 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.
- Aleotti, P. (2004). A warning system for rainfall-induced shallow failures. *Engineering Geology*, 73(3), 247–265.
- Alkhasawneh, M., Ngah, U. K., Tay, L. T., Isa, N. A. M., & Al-Batah, M. S. (2014). Modeling and testing landslide hazard using decision tree. *Journal of Applied Mathematics*,. doi:10.1155/2014/929768.
- Bennett, N. D., Croke, B. F. W., Guariso, G., Guillaume, J. H. A., Hamilton, S. H., Jakeman, A. J., Marsili-Libelli, S., Newham, L. T. H., Norton, J. P., Perrin, C., Pierce, S. A., Robson, B., Seppelt, R., Voinov, A. A., Fath, B. D., Andreassian, V. (2013). Characterising performance of environmental models. *Environmental Modelling and Software*, 40, 1–20.
- Cevik, E., & Topal, T. (2003). GIS-based landslide susceptibility mapping for a problematic segment of the natural gas pipeline,

- Hendek (Turkey). *Environmental Geology*, 44(8), 949–962. doi:10.1007/s00254-003-0838-6.
- Choi, J., Oh, H.-J., Lee, H.-J., Lee, C., & Lee, S. (2012). Combining landslide susceptibility maps obtained from frequency ratio, logistic regression, and artificial neural network models using ASTER images and GIS. *Engineering Geology*, 124, 12–23. doi:10.1016/j.enggeo.2011.09.011.
- Chung, C.-J. F., & Fabbri, A. G. (2003). Validation of spatial prediction models for landslide hazard mapping. *Natural Hazards*, 30(3), 451–472.
- Coe, J. A., Michael, J., Crovelli, R., & Savage, W. (2000). *Preliminary map showing landslide densities, mean recurrence intervals, and exceedance probabilities as determined from historic records*. Washington: Seattle.
- Cohen, W. W. (1995). Fast effective rule induction. In *Proceedings of the twelfth international conference on machine learning* (pp. 115–123).
- Corominas, J., & Moya, J. (2008). A review of assessing landslide frequency for hazard zoning purposes. *Engineering Geology*, 102(3), 193–213.
- Crovelli, R. A. (2000). *Probability models for estimation of number and costs of landslides*. U.S. Geological Survey Open File Report 00-249. (p. 23).
- Dai, F., & Lee, C. (2002). Landslide characteristics and slope instability modeling using GIS, Lantau Island, Hong Kong. *Geomorphology*, 42(3), 213–228. doi:10.1016/S0169-555X(01)00087-3.
- Gao, Y., & Wang, Y. (2006). Boosting in random subspace for face recognition. In Huang, D.-S., & Irwin, G. W. (Eds.), *Intelligent computing in signal processing and pattern recognition* (pp. 172–181). Berlin: Springer.
- Guzzetti, F., Peruccacci, S., Rossi, M., & Stark, C. P. (2007). Rainfall thresholds for the initiation of landslides in central and southern Europe. *Meteorology and Atmospheric Physics*, 98(3–4), 239–267.
- Guzzetti, F., Peruccacci, S., Rossi, M., & Stark, C. P. (2008). The rainfall intensity–duration control of shallow landslides and debris flows: An update. *Landslides*, 5(1), 3–17.
- Guzzetti, F., Reichenbach, P., Cardinali, M., Galli, M., & Ardizzone, F. (2005). Probabilistic landslide hazard assessment at the basin scale. *Geomorphology*, 72(1), 272–299.
- Ho, T. K. (1998). The random subspace method for constructing decision forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(8), 832–844.
- Hühn, J., & Hüllermeier, A. (2009). FURIA: An algorithm for unordered fuzzy rule induction. *Data Mining and Knowledge Discovery*, 19(3), 293–319. doi:10.1007/s10618-009-0131-8.
- 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. doi:10.1007/s10346-013-0391-7.
- Kjekstad, O., & Highland, L. (2009). Economic and social impacts of landslides. In Sassa, K., & Canuti, P. (Eds.), *Landslides—Disaster risk reduction* (pp. 573–587). Berlin: Springer.
- Kleinberg, E. (1990). Stochastic discrimination. *Annals of Mathematics and Artificial Intelligence*, 1(1), 207–239.
- Larsen, M. C., & Simon, A. (1993). A rainfall intensity–duration threshold for landslides in a humid-tropical environment, Puerto Rico. *Geografiska Annaler. Series A. Physical Geography*, 75, 13–23.
- NCEP. (2014). Global weather data for SWAT. <http://globalweather.tamu.edu/home>
- Onan, A. (2015). Classifier and feature set ensembles for web page classification. *Journal of Information Science*,. doi:10.1177/0165551515591724.
- Pham, B. T., Tien Bui, D., Dholakia, M. B., Prakash, I., & Pham, H. V. (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), 1–18. doi:10.1007/s10706-016-9990-0.
- Pham, B. T., Tien Bui, D., Pourghasemi, H. R., Indra, P., & Dholakia, M. B. (2015). 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. *Theoretical and Applied Climatology*, 122(3–4), 1–19. doi:10.1007/s00704-015-1702-9.
- Pham, B. T., Tien Bui, D., Prakash, I., & Dholakia, M. B. (2016b). Rotation forest fuzzy rule-based classifier ensemble for spatial prediction of landslides using GIS. *Natural Hazards*,. doi:10.1007/s11069-016-2304-2.
- 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. *Journal of Earth System Science*, 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. *Computers & Geosciences*, 51, 350–365. doi:10.1016/j.cageo.2012.08.023.
- Saez, J. L., Corona, C., Stoffel, M., Schoeneich, P., & Berger, F. (2012). Probability maps of landslide reactivation derived from tree-ring records: Pra Bellon landslide, southern French Alps. *Geomorphology*, 138(1), 189–202.
- Terlien, M. T., Van Westen, C. J., & van Asch, T. W. (1995). Deterministic modelling in GIS-based landslide hazard assessment. In Carrara, A., & Guzzetti, F. (Eds.), *Geographical information systems in assessing natural hazards* (pp. 57–77). Berlin: Springer.
- 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. *International Journal of Digital Earth*, 9(11), 1–21.
- Tien Bui, D., Pradhan, B., Lofman, O., Revhaug, I., & Dick, Ø. B. (2013). Regional prediction of landslide hazard using probability analysis of intense rainfall in the Hoa Binh province, Vietnam. *Natural Hazards*, 66(2), 707–730.
- Vapnik, V. N. (1995). *The nature of statistical learning theory*. New York: Springer.
- Varnes, D. J. (1984). *Landslide hazard zonation: A review of principles and practice*. Paper presented at the UNESCO Press, Paris.
- 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. *Arabian Journal of Geosciences*, 6(8), 2873–2888.