Conditioning factor determination for mapping and prediction of landslide susceptibility using machine learning algorithms

Conference Paper · October 2019

DOI: 10.1117/12.2532687

CITATIONS
4

READS
364

4 authors, including:

Bahareh Kalantar
RIKEN
61 PUBLICATIONS **547** CITATIONS
SEE PROFILE

Biswajeet Pradhan
University of Technology Sydney
773 PUBLICATIONS **24,624** CITATIONS
SEE PROFILE

Vahideh Saeidi
Universiti Putra Malaysia
17 PUBLICATIONS **93** CITATIONS
SEE PROFILE

Conditioning factor determination for mapping and prediction of landslide susceptibility using machine learning algorithms

Husam Abdulrasool Hammadeh Al-Najjar, Bahareh Kalantar, Biswajeet Pradhan, Vahideh Saeidi

Husam Abdulrasool Hammadeh Al-Najjar, Bahareh Kalantar, Biswajeet Pradhan, Vahideh Saeidi, "Conditioning factor determination for mapping and prediction of landslide susceptibility using machine learning algorithms," Proc. SPIE 11156, Earth Resources and Environmental Remote Sensing/GIS Applications X, 111560K (3 October 2019); doi: 10.1117/12.2532687

SPIE.

Event: SPIE Remote Sensing, 2019, Strasbourg, France

Conditioning factors determination for mapping and prediction of landslide susceptibility using machine learning algorithms

Husam A. H. Al-Najjar [a], Bahareh Kalantar [b], Biswjaeet Pradhan *[a,c], Vahideh Saeidi[d]
[a]Centre for Advanced Modelling and Geospatial Information Systems (CAMGIS), Faculty of Engineering and IT, University of Technology Sydney, 2007 NSW, Australia; [b]RIKEN Center for Advanced Intelligence Project, Goal-Oriented Technology Research Group, Disaster Resilience Science Team, Tokyo 103-0027, Japan; [c]Department of Energy and Mineral Resources Engineering, Choongmu-gwan, Sejong University, 209 Neungdong-ro Gwangjin-gu, 05006, Seoul, South Korea.
[d]Department of Mapping and Surveying, Darya Tarsim Consulting Engineers Co. Ltd., 1457843993 Tehran, Iran.

*Biswajeet.Pradhan@uts.edu.au, biswajeet24@gmail.com; phone +61 2 95147937; www.uts.edu.au

ABSTRACT

Landslides are type of natural geohazard interfering with many economical and social activities and causing serious damages on human life. It is ranked as a great disaster, threatening life, property and environment. Therefore, early prediction of landslide prone areas is vital. Variety of causative factors such as glaciers melting, excessive raining, mining, volcanic activities, active faults, earthquake, logging, erosion, urbanization, construction, and other human activities can trigger landslide occurrence. Then, identification of factors that directly influences the slide events is highly in demand. Some topographical, geological, and hydrological datasets (e.g., slope, aspect, geology, terrain roughness, vegetation index, distance to stream, distance to road, distance to fault, land use, precipitation, profile curvature, plan curvature) are considered to be effective conditioning factors. However, the importance of each factor differs from one study to another. This study investigates the effectiveness of four sets of landslide conditioning variable(s). Fourteen landslide conditioning variables were considered in this study where they were duly divided into four groups G1, G2, G3, and G4. Three machine learning algorithms namely, Random Forest (RF), Naive Bayes (NB), and Boosted Logistic Regression (LogitBoost) were constructed based on each dataset in order to determine which set would be more suitable for landslide susceptibility prediction. In total, 227 landslide inventory datasets of the study area were used where 70% was used for training and 30% for testing. To this end, in the present research, the two main objectives were: 1) Investigation on effectiveness of 14 landslides conditioning factors (altitude, slope, aspect, total curvature, profile curvature, plan curvature, Stream Power Index (SPI), Topographic Wetness Index (TWI), Terrain Roughness Index (TRI), distance to fault, distance to road, distance to stream, land use, and geology) by analyzing and determining the most important factors using variance-inflated factor (VIF), Pearson's correlation and Chi-square techniques. Consequently, 4 categories of datasets were defined; first dataset included all 14 conditioning factors, second dataset included Digital Elevation Models (DEM) derivatives (morphometrice factors), third dataset was only based on 5 factors namely lithology, land use, distance to stream, distance to road, and distance to fault, and last dataset was included 8 factors selected using factor analysis and optimization. 2) Evaluate the sensitivity of each modeling technique (NB, RF and LogitBoost) to different conditioning factors using the area under curve (AUC). Eventually, RF technique using optimized variables (G4) performed well with AUC of 0.940 followed by LogitBoost (0.898) and NB (0.864).

Keywords: Landslide conditioning factors, machine learning, landslide prediction, Naive Bayes, Random Forest, LogitBoost

Earth Resources and Environmental Remote Sensing/GIS Applications X, edited by Karsten Schulz, Ulrich Michel, Konstantinos G. Nikolakopoulos, Proc. of SPIE Vol. 11156, 111560K

1. INTRODUCTION

Natural disasters such as landslides or mass movements have recently been reflecting important attention around the globe. That is due to their destructive impacts on economy, social and environment particularly in urbanized area [1]. Moreover, a large number of landside events and their consequences have been boosted due to urban sprawls in risky areas [2]. Although it may not be fully achievable to prevent landslide occurrence, prone areas can be mapped to mitigate future losses in human, infrastructures and resources [3].

In this regard, remote sensing technology including space born, airborne and Unmanned Aerial Vehicle (UAVs) platforms with their broad prosperities can support to implement fast response, recovery, preparation and mitigation strategies. From one side, conditioning factor play an important role in landslide susceptibility analysis. These factors have been investigated using different remote sensing data such as Spaceborne Synthetic aperture radar (SAR), optical LiDAR (Light Detection and Ranging), ground based SAR, terrestrial LiDAR incorporation with in-situ measurements [4]. However, different study area with different type of data has diverse susceptibility result. Accordingly, effective conditioning factors and spatial relationship between them are essential to identify risk zones [5]. From another side, machine learning methods have influential role to generate better landslide susceptibility maps [6 5 7 8 9]. For instance, considerable attempts have been conducted to optimize landslide susceptibility mapping; scholars such as Dou et al. [10] optimized potential of effective factors from fifteen to six factors (distance to faults, distance to geological boundary, drainage density, slope angle, lithology and slope aspect) in Sado Island, Japan. They utilized Statistical Index (SI) and Logistic Regression (LR) to enhance the susceptibility map, and then they concluded that the model using certain factor was more accurate. In another study, Afungang et al. [11] used Informative value Model to enhance the effective conditioning factors from 8 to 6 factors. Their training model gained success rate of 87%, and the validation model achieved a rate of 90%. Another investigation by Mahalingam et al. [12] addressed the performance of six factors derived from digital elevation model (DEM), LiDAR data. They considered six methods for their analysis namely, Support Vector Machine (SVM), Frequency Ratio (FR), Weights of Evidence (WoE), LR, Discriminant Analysis (DA) and Artificial Neural Network (ANN). Among all, SVM ranked the highest compared to other methods and ANN the lowest. However, Hydrology effects and qualitative validation have not been considered in their study. More review studies on landslide susceptibility considering different factors and methods can be found through the following scholars [13 14 15 16 17 4 16].

Through deep investigation from literatures, it appears that different areas with various environmental and geological properties have unique landslide modelling. Hence, the involvement of diverse factor combination with implementation of various machine learning methods is still a great interest to gain more realistic susceptibility modelling. In this study, 4 datasets namely G1, G2, G3 and G4 (derived from initial 14 conditioning factors) were considered to generate landslide susceptibility map. First, each dataset, which contains combination of different factors, were extracted from most important factor analysis methods namely VIF, Pearson's correlation, Chi-square. Then, the four datasets were implemented as feeding input for performance evaluation in order to develop landslide susceptibility models based on 3 different machine learning algorithms such as Random Forest (RF), Naive Bayes (NB), and Boosted Logistic Regression (LogitBoost). Finally, sensitivity of each modelling technique was evaluated.

2. STUDY AREA AND DATA

2.1 Study area

Sajadrood catchment with approximate coverage area of 118 Km² and population of 62,809 (census 2006), is one of Babolrood sub-catchments, which is located in Mazandaran province, Northern Iran with north latitudes 36°9′ and 36°10′ and east longitudes 52°30′ and 52°40′ (Figure 1a). This region experienced several landslides events along recent years. Sajadrood, geologically, categorized as faulted geologic region. In terms of Hydrology, the catchment experiences highest amount of raining in the autumn season. The highest elevation in Sajadrood is 3713 meters, in contrast, the least level is 30 meters. As it is illustrated in the (Figure 1o), the study area has diverse lithology pattern. For the current study, a DEM with 20 meters resolution was generated including a topographic map with scale of 1: 25,000.
In this research, 227 historical landslide events which known as inventory map gathered by means of satellite images as well as field expert surveying in Iran were used. We randomly utilized 70 % of the inventories for training of three

machine learning models namely, RF, NB and LogitBoost. we retrained the remaining 30 % of the inventories data for testing the models. Due to data availability and according to relevant studies by [1] [18] [19] [20] [3] we derived 14 conditioning factors from DEM and topographic directory using Arc Map v. 10.5. The condoning factors including as the following: altitude, slope, aspect, curvature, profile curvature, plan curvature, Stream Power Index (SPI), Topographic Wetness Index (TWI), Terrain Roughness Index (TRI), distance to fault, distance to road, distance to stream, land use and lastly Lithology (Figure 1b-o).

2.2 Landslide conditioning factors preparation

Process of factor arrangements and selection was carried out according to previous studies such as [3] [21]. This process is described briefly in this section.

***Altitude**:* Alternation in the altitude has noticeable effect on landslide susceptibility models. In our study, we divided the altitude to 5 classes via natural break pattern. Accordingly, its classes were defined from minimum value of 74 meters to maximum of 1500 meters (Figure 1b).

***Slope*:** Slope is considered as one of critical causing landslide hazard in areas with sharp slope due to soil weakness and stresses. In this study, slope angle was classified into 5 levels, including: (i) 0°-8.4°, (ii) 8.5°-13°, (iii) 14°-17°, (iv) 18°-23°, and (v) 24°-48° (Figure 1c).

Aspect: This factor, generally, represents the compass orientation that a slope or hillshade confronts. It is also can be used as indication for some measurements of plants group, soil moistures and evaporation. It is categorized into 9 classes including: (i) flat, (ii) north, (iii) northeast, (iv) east, (v) southeast, (vi) south, (vii) southwest, (viii) west, and (ix) northwest (Figure 1d).

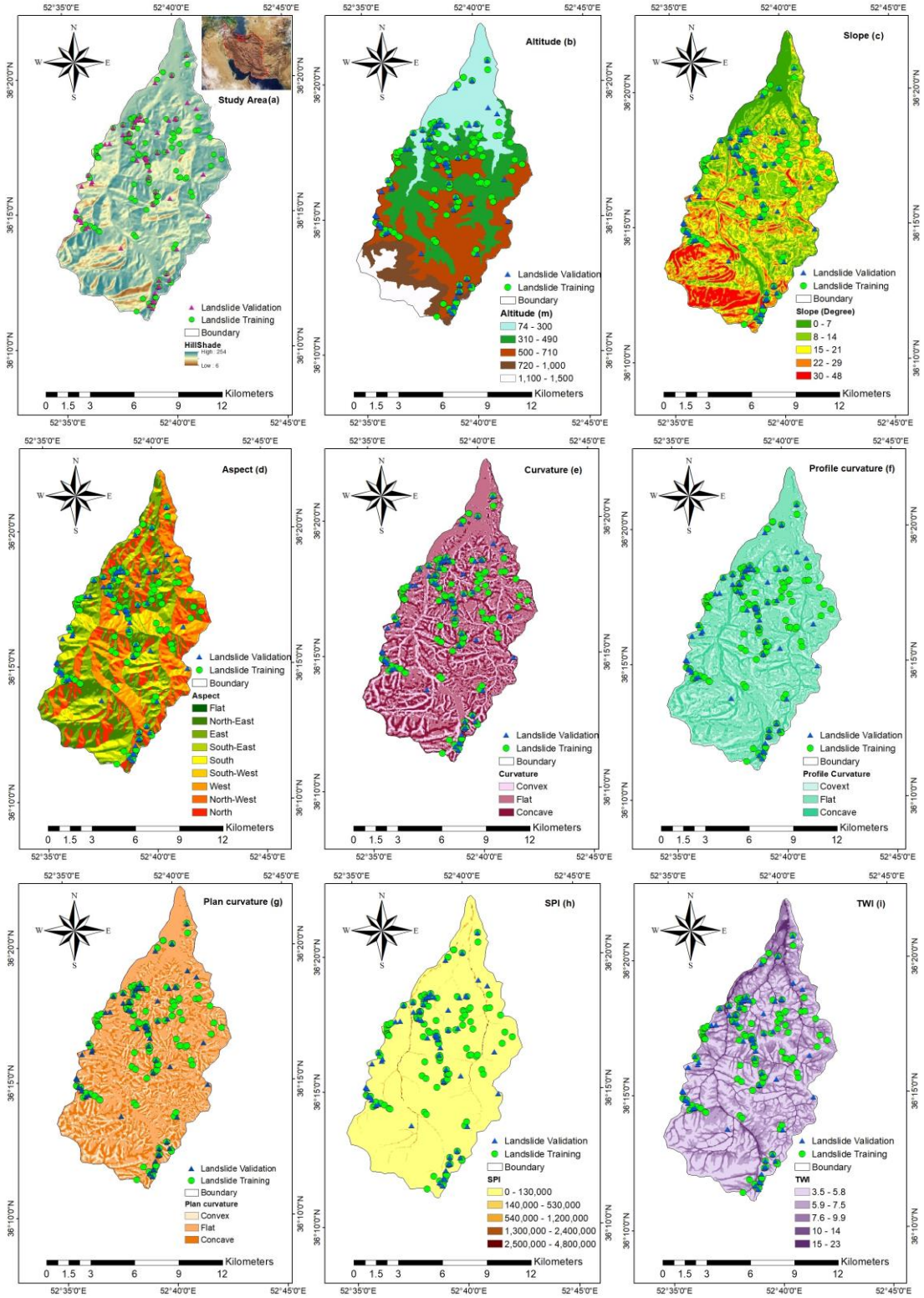
Curvature, Profile curvature, Plan curvature: Surface curvature reveals the earth's ground shape including soil run off. Profile curvature is in parallel direction with maximum slope. It has influence on speeding up or slowing down of run off on the surface. The other factor namely plan curvature, however, is at the 90 degrees to the orientation of maximum slope which indicates the encounter or separation of flow on a surface. In principal, curvature is summation of plan and profile curvature. Further detail is addressed in the reference [22]. In this research, total, profile and plan curvature were categorized based on three clasess namely concave, flat and covex [23] (Figure 1e,f,g).

Stream Power Index (SPI), Topographic Wetness Index (TWI), Terrain Roughness Index (TRI): SPI illustrates the discharge erosion of a given point in a surface. The SPI and risk of erosion increase if the amount of supply up-stream water increases. TWI, on the other hand, estimates amount of wetness in the soil. Whereas, TRI indicates the variation in the elevation of neighboring cells in a digital elevation cells network. We classified the SPI, TWI, TRI (Figure 1h,i,j) based on 5 classes as per reference [1], which further definitions are available there.

Distance to fault, Distance to road, Distance to stream: Potential of landslide occurrence generally increases by adjacent to fault, road, and stream. This may be related to human activities and erosion in those areas. Considering references [24] [25] [23] we classified these factors into 5 classes via ARCGIS 10.5 software by implementing Euclidean distance function (Figure 1k, l, m).

***Land use*:** Land use map illustrates how a land resources are used for. As an example, in rural area a land can be used as agriculture, forestry or water bodies. While, in urban area, a land can be used as industrial purpose, housing, green spaces or parks. In this research, we categorized the land use based on 6 classes using supervised classification with accuracy of 90 %, namely agricultural lands, harvested forest, massive forest, residential land, rural lands and finally water and rain gardens (Figure 1n). For this scope, we utilized satellite image of Landsat Thematic Mapper which was taken in 2017.

***Lithology**:* This factor considers the physical characteristics of outcrop rocks in the case study. The study area has 13 types of lithology classes (Figure 1o).



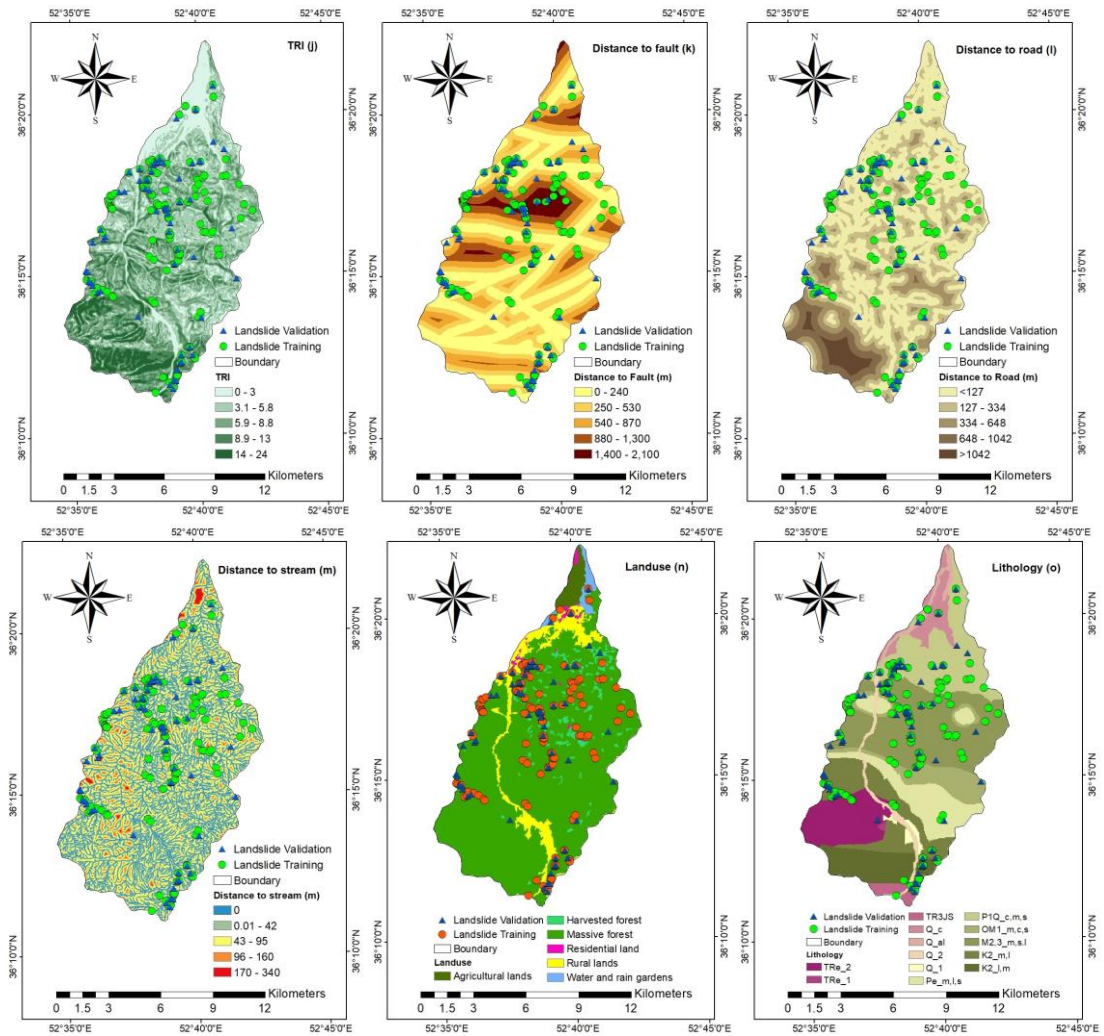


Figure 1. a) study area, b) Altitude, c) Slope, d) Aspect, e) Curvature, f) Profile curvature, g) plan curvature, h) SPI, i) TWI, j) TRI, k) distance to fault, l) distance to road, m) distance to stream, n) landuse, o) lithology.

3. METHODOLOGY

Figure 2 shows the overall methodology flowchart of this study. First, the dataset that we use consisted of 227 landslide inventory points collected by the Geological Survey of Iran from the Sajadrood District of the Sari County. 70% of the dataset was used for training and 30% was used for testing. Next, the dataset was divided into four groups, G1, G2, G3, and G4. Then three machine learning algorithms namely, NB, RF, and LogitBoost were constructed on each group of factors. Lastly, Area Under Curve (AUC) was analyzed to evaluate the performance of three groups of conditioning factors in each model.

3.1 Conditioning factor analysis and optimization

We implemented the highly related features discard approach using an estimation of variance-inflated factor (VIF) as the following equation:

$$VIF = \frac{1}{1-R'^2} \tag{1}$$

where R' represent the multi correlation coefficient among single factor and other factors in the model. In the this study, factors with a VIF value greater than 5 or 10 were identified as the high correlation and should be removed [26]. The correlation coefficient of two conditioning factors was calculated using Pearson's correlation coefficients method (Eq. 2):

$$r_{xy} = \sum_{i=1}^{n} \frac{X_i - \bar{X}}{\sqrt{\sum_{k=1}^{n}(X_i - \bar{X})^2}} \frac{Y_i - \bar{Y}}{\sqrt{\sum_{k=1}^{n}(Y_i - \bar{Y})^2}} \quad (2)$$

In landslide susceptibility analysis, both training samples and computational cost are generally raised by increasing the conditioning factors. This issue leads to misguiding the regression coefficients. Accordingly, in order to decrease the obstacles, these factors require an optimization process which is known as factor optimization. In this investigation, the Chi-square was implemented as factor optimization method to drop out the senseless components at confidence grade of 0.05 (95%).

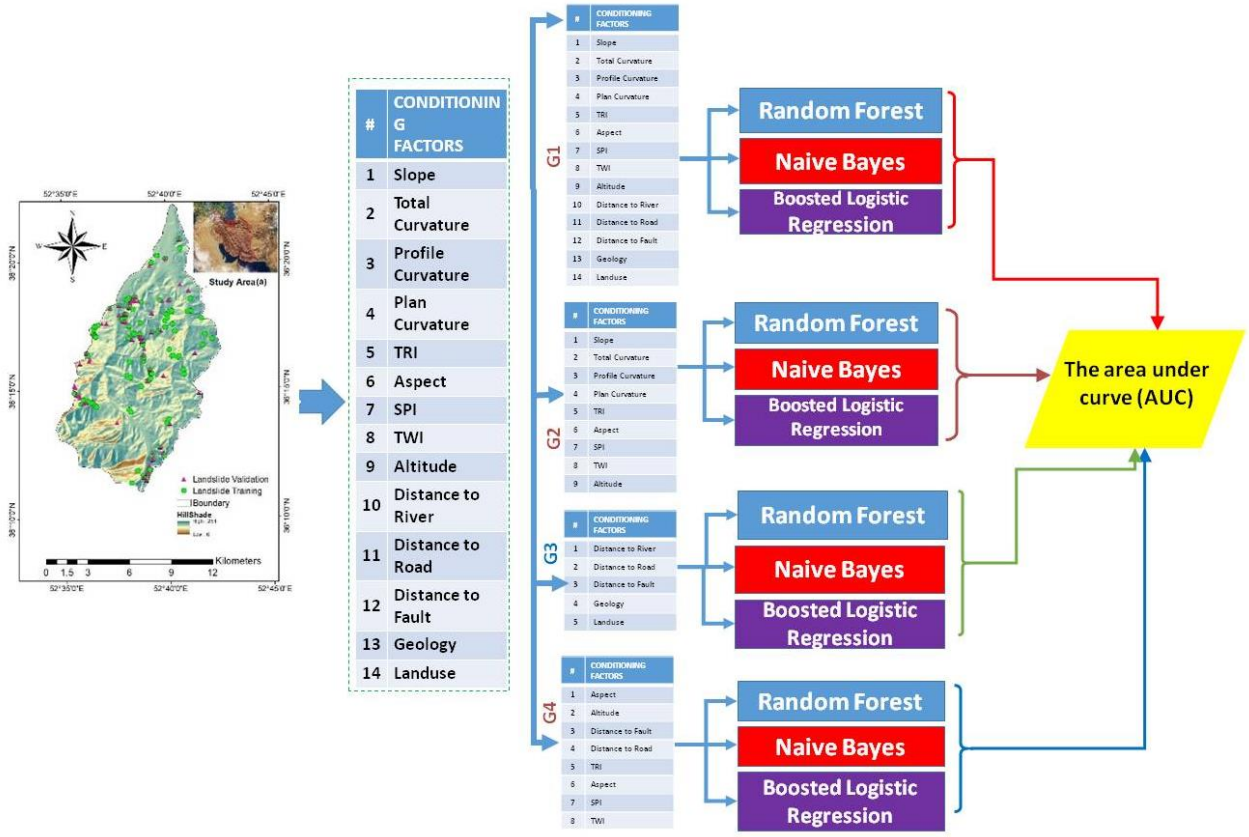


Figure 2. Methodological flowchart.

3.2 Principle of machine learning models

3.2.1 Random forest (RF)

Random forest (RF) is one of the powerful non-parametric learning approaches which is extensively used in variety of geospatial applications such as image classification and analysis [27 28]. It relies on multiple decision trees from set of training data. In complex dataset, compared with other decision trees, RF is less sensitive to over-fitting obstacles. The output of every individual random forest is anticipated by a decision tree, and that output receives a weightage by votes. The final classification generated from the greater part of voting for an output accompanying a convergency level [2930].

3.2.2 Naive Bayes (NB)

Naive Bayes (NB) is one of machine learning classifiers, which is well-employed in remote sensing applications such as landslide susceptibility models [313233]. It is considered as a part of simple Probabilistic classifiers family, Bayes's law based and its performance relies on self-reliant variables estimation [34]. It is noticeably convenient to use because there is no complex repeated parameter assessment in its structure [35].

3.2.3 Boosted Logistic Regression (LogiBoost)

LogitBoost is boosting algorithm which is mainly used for best fitting the linier logistic regression [36]. It determines the over-fitting obstacle using LogitBoost design [37]. Basically, a singular class is well-fitted with least square fitting using additional logistic regressions, taking an example of landslide and non-landslide events [38].

3.3 Validation

In this research, the area under the receiver operating characteristic curve (AUC) by evaluation the prediction and success rates were looked at to evaluate the performance of three machine learning algorithms namely, NB, RF and LogitBoost. The AUC value is classified into scales related to qualitative classes. Values from 0.5-0.6 indicates poor, 0.6-0.7 average, 0.7-0.8 as good, 0.8-0.9 means very good and 0.9-1 is exceptional (or excellent) [26]

4. RESULT

Results are presented in two sections.

(i) First, factor analysis and optimization using VIF, Pearson's correlation, and Chi-Square is presented.

(ii) Second, we review the performance of three machine learning algorithms with four different datasets (G1, G2, G3, and G4). Then the AUC accuracies are used to compare the outcomes of the models.

4.1 Factor analysis and optimization results

Both VIF and Pearson's coefficients were used to detect the multicollinearity through the conditioning factors. A value of VIF greater than 5 or 10 implies a high correlation, which means higher multicollinearity[39]. According to Table 1, the highest VIF value was 3.02255E+13 (curvature). Moreover, the slope, plan curvature, profile curvature and TRI with VIF values of 23.29, 1.11475E+13, 1.18828E+13, 23.51 were detected as the highly correlated factors. In addition, the linear correlations between two conditioning factors were calculated using Pearson's correlation coefficient (Table 2). The coefficient more than 0.7 indicate high collinearity[18]. As it can be seen in Table 2, the highest correlation was between TRI and slope with a value of 0.9. Moreover, there were strong correlation between profile curvature and curvature, and curvature and plan curvature with a value of 0.82 and 0.80, respectively. In this case, the simple method to overcome with high Pearson correlation between conditioning factor is to remove one of the factors from the dataset and rebuild analtysis [1]. On the other hand, factor optimization using Chi-Square indicated that higher Chi-square values with p-value less than 0.05 ranks the significance of each factor for landslide prediction. Therefore, the factor optimization results determined that distance to road, altitude, lithology, TWI and distance to fault were found the most important landslide conditioning factors; however, the land use and slope and distance to stream were identified as less significant factors (Figure 3).

Table 1. The estimated variance information factor (VIF) for landslide conditioning factors.

	Summary statistics and multicollinearity			
Variable	**Means**	**Std.Devs**	**Multiple**	**VIF**
Altitude	507.60	236.5	0.67	1.814552713
Slope	16.07	8.2	0.98	**23.29487996**
Curvature	0.05	0.4	1.00	**3.02255E+13**
Plan Curvature	0.03	0.3	1.00	**1.11475E+13**
Profile Curvature	-0.02	0.3	1.00	**1.18828E+13**

TWI	6.37	2.1	0.63	1.658710203
TRI	6.50	3.4	0.98	**23.51014998**
SPI	17070.64	152054.2	0.30	1.100591715
Distance to Stream	44.90	32.1	0.24	1.062830076
Distance to Road	144.73	248.6	0.39	1.183368206
Distance to Fault	501.08	436.9	0.18	1.032186586
Landuse	3.13	0.7	0.07	1.005608413
Lithology	7.20	3.0	0.57	1.481543944
Aspect	4.61	2.8	0.03	1.0008435

Table 2. Pearson correlations between landslide conditioning factors.

	Correlation matrix between landslide conditioning factors													
Variable	**Altitude**	**Slope**	**Total Curvature**	**Plan Curvature**	**Profile Curvature**	**TWI**	**TRI**	**SPI**	**Distance to Stream**	**Distance to Road**	**Distance to Fault**	**Landuse**	**Geology**	**Aspect**
Altitude	1.00	0.53	0.11	0.02	-0.15	-0.30	0.55	-0.07	0.02	0.55	0.05	-0.21	-0.69	-0.12
Slope	0.53	1.00	0.09	0.04	-0.10	-0.47	**0.99**	-0.12	0.05	0.32	0.08	-0.08	-0.45	-0.07
Total Curvature	0.11	0.09	1.00	**0.80**	**-0.82**	-0.55	0.07	-0.18	0.41	-0.12	0.09	0.00	0.03	-0.03
Plan Curvature	0.02	0.04	**0.80**	1.00	-0.31	-0.49	0.03	-0.09	0.44	-0.17	0.06	0.00	0.03	-0.03
Profile Curvature	-0.15	-0.10	**-0.82**	-0.31	1.00	0.39	-0.08	0.20	-0.22	0.02	-0.09	-0.01	-0.02	0.03
TWI	-0.30	-0.47	-0.55	-0.49	0.39	1.00	-0.44	0.46	-0.31	0.04	-0.09	0.04	0.18	0.02
TRI	0.55	**0.99**	0.07	0.03	-0.08	-0.44	1.00	-0.09	0.04	0.35	0.08	-0.08	-0.47	-0.07
SPI	-0.07	-0.12	-0.18	-0.09	0.20	0.46	-0.09	1.00	-0.10	0.00	0.00	-0.11	0.02	-0.06
Dis. to Stream	0.02	0.05	0.41	0.44	-0.22	-0.31	0.04	-0.10	1.00	-0.13	0.19	-0.02	0.08	-0.04
Dis. to Road	0.55	0.32	-0.12	-0.17	0.02	0.04	0.35	0.00	-0.13	1.00	-0.09	-0.06	-0.39	-0.09
Dis. to Fault	0.05	0.08	0.09	0.06	-0.09	-0.09	0.08	0.00	0.19	-0.09	1.00	0.03	0.21	-0.07
Landuse	-0.21	-0.08	0.00	0.00	-0.01	0.04	-0.08	-0.11	-0.02	-0.06	0.03	1.00	0.14	0.04
Lithology	-0.69	-0.45	0.03	0.03	-0.02	0.18	-0.47	0.02	0.08	-0.39	0.21	0.14	1.00	0.08
Aspect	-0.12	-0.07	-0.03	-0.03	0.03	0.02	-0.07	-0.06	-0.04	-0.09	-0.07	0.04	0.08	1.00

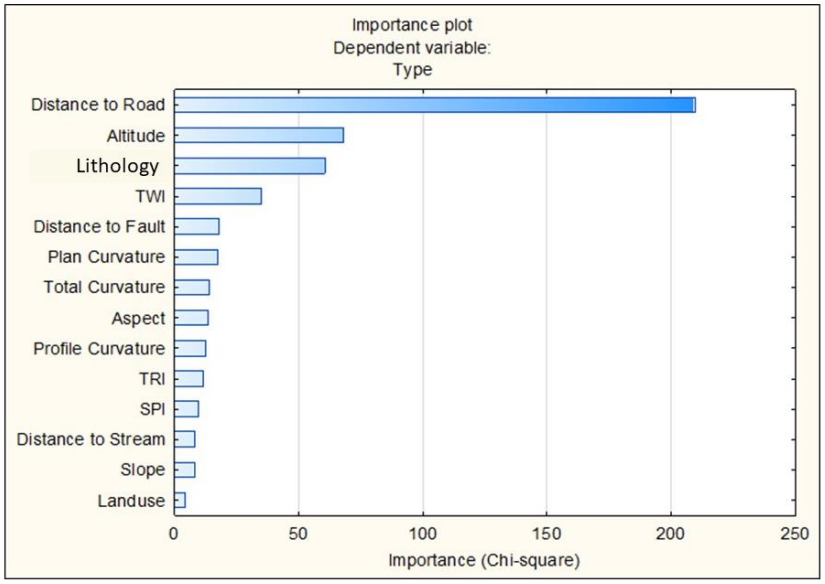


Figure 3. The important plot of conditioning factors using Chi-Square.

4.2 Models validation

Finally, for the evaluation of the prediction models and the best choice of datasets, the results of AUC for 4 groups of datasets using 3 models are shown in Table 3. RF model's prediction performances were reported by the value of 0.939 in G1, 0.819 in G2, 0.903 in G3, and 0.940 in G4 dataset. The second model, NB, obtained the maximum AUC of 0.866 using G3, and minimum AUC of 0.658 using G2 dataset. Lastly, LogitBoost model's maximum and minimum performance were recorded by 0.911and 0.688 using G3 and G2, respectively. All models had satisfactory results except NB and LogitBoost using DEM-derived variables (G2).

Table 3. The ROC area value for RF, NB, and LogitBoost.

Models	ROC AREA			
	G1	G2	G3	G4
RF	0.939	0.819	0.903	0.940
NB	0.857	0.658	0.866	0.864
LogitBoost	0.893	0.688	0.911	0.898

5. DISCUSSION

Based on the results of factor analysis and optimizations, between slope and TRI, we removed slope factor, because it was labeled redundant data by VIF and Pearson's correlation, also it was ranked as one of the least important factors by Chi-square analysis, as well. Similarly, distance to stream, land use, all curvatures variables were excluded from G1 to create the last group of data (G4) including the most significant factors namely, distance to road, altitude, lithology, TWI, distance to fault, aspect, SPI, and TRI. Consequently, three aforementioned models were applied on G1, G2, G3, and G4.

According to Table 3, RF model successfully was evaluated by very good and excellent level of prediction performance in all 4 groups of dataset and it reached almost the highest rate of success among 3 applied models, as well. Applying RF, G2 dataset (DEM derivatives only) represented the lowest reliability to compare with G1, G3, and G4, while the optimization procedures lead to the best AUC result in G4. However, the AUC's values enhancement from using all 14 conditioning factors (G1) to optimized factors (G4) were not significant by exploiting RF. This highlights the power of RF model to handle all types of variables and causative factors even redundant or incomplete datasets. The similar pattern was followed by NB model whereas, NB was not successful in using G2 and it obtained an average level of AUC rate. Meanwhile, LogitBoost model obtained an average accuracy when it was dealing with G2 dataset, as well. In contrast, LogitBoost performed excellent with G3 dataset. It showed LogitBoost result was insignificantly better than RF using only 5 factors (G3). Generally speaking, RF was the best prediction model for landslide susceptibility mapping (AUC=0.940) and factor optimization could slightly improve the accuracy of the results. In addition, NB and LogitBoost models were not reliable in a certain place where we have incomplete and redundant data such as G2.

6. CONCLUSION

In this study, the precision of four groups of conditioning factors were compared to analysis and determinate the most important factors for landslide susceptibility mapping. The first dataset named as G1 included 14 landslide conditioning factors (altitude, slope, aspect, total curvature, profile curvature, plan curvature, SPI, TWI, TRI, distance to fault, distance to road, distance to stream, land use, and geology). The second dataset G2 included only DEM derived factors. However, the third dataset (G3) was only based on 5 factors namely lithology, land use, distance to stream, distance to road, and distance to fault. In other words, G3 included all factors excluding DEM-derived factors. The last dataset G4 included 8 factors which selected using factor analysis (VIF, Pearson's correlation) and factor optimization (Chi-square technique). The study analyzed landslide susceptibility on a catchment scale in Sajadrood, Iran, using well-known machine learning techniques, NB, RF and LogitBoost. This research emphasized on the significance of distance to road,

altitude, and lithology as landslide causal factors. The results showed the RF has higher accuracy in almost every group of conditioning factors. The multiple decision trees in RF were successfully trained by all forms of datasets and it was proved by our research. However, NB and LogitBoost algorithms had an average performance in G2 which was only DEM-derived factors. The level of accuracy improvement using factor optimization methods was not considerably highlighted while NB and LogitBoost models were significantly sensitive to the use of morphometric datasets only. To great extent, the results of RF technique showed this algorithm was more suitable for landslide prediction with all type of variables.

REFERENCE

[1] Kalantar, B., Ueda, N., Al-Najjar, H. A. H., Gibril, M. B. A., Lay, U. S. and Motevalli, A., "An evaluation of landslide susceptibility mapping using remote sensing data and machine learning algorithms in Iran," ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci. **4**(2/W5), 503–511 (2019).

[2] Novellino, A., Jordan, C., Ager, G., Bateson, L., Fleming, C. and Confuorto, P., [Geological disaster monitoring based on sensor networks], Springer Singapore (2019).

[3] Kalantar, B., Ueda, N., A. Alnajjar, H., Idrees, M. O., Motevalli, A. and Pradhan, B., "Landslide susceptibility mapping at Dodangeh watershed, Iran using LR and ANN models in GIS," Proc. SPIE 10790, Earth Resour. Environ. Remote Sensing/ GIS Appl. IX, 107901D, 41 (2018).

[4] Zhao, C. and Zhong, L., "Remote Sensing of Landslides — A Review," Remote Sens. **10**(279), 8–13 (2018).

[5] Arabameri, A., Pradhan, B., Rezaei, K. and Lee, C. W., "Assessment of landslide susceptibility using statistical- and artificial intelligence-based FR-RF integrated model and multiresolution DEMs," Remote Sens. **11**(9) (2019).

[6] Bui, D. T., Shahabi, H., Shirzadi, A., Chapi, K., Pradhan, B., Chen, W., Khosravi, K., Panahi, M., Ahmad, B. Bin and Saro, L., "Land subsidence susceptibility mapping in South Korea using machine learning algorithms," Sensors (Switzerland) **18**(8) (2018).

[7] Zhou, C., Yin, K., Cao, Y., Ahmed, B., Li, Y., Catani, F. and Pourghasemi, H. R., "Landslide susceptibility modeling applying machine learning methods: A case study from Longju in the Three Gorges Reservoir area, China," Comput. Geosci. **112**(April 2017), 23–37 (2018).

[8] Pham, B. T., Pradhan, B., Tien Bui, D., Prakash, I. and 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).

[9] Huang, Y. and Zhao, L., "Review on landslide susceptibility mapping using support vector machines," Catena **165**(January), 520–529 (2018).

[10] Dou, J., Bui, D. T., Yunus, A. P., Jia, K., Song, X., Revhaug, I., Xia, H. and Zhu, Z., "Optimization of causative factors for landslide susceptibility evaluation using remote sensing and GIS data in parts of Niigata, Japan," PLoS One **10**(7) (2015).

[11] Afungang, R. N., de Meneses Bateira, C. V. and Nkwemoh, C. A., "Assessing the spatial probability of landslides using GIS and informative value model in the Bamenda highlands," Arab. J. Geosci. **10**(17), 1–15 (2017).

[12] Mahalingam, R., Olsen, M. J. and O'Banion, M. S., "Evaluation of landslide susceptibility mapping techniques using lidar-derived conditioning factors (Oregon case study)," Geomatics, Nat. Hazards Risk **7**(6), 1884–1907 (2016).

[13] Ma, Z., Qin, S., Cao, C., Lv, J., Li, G., Qiao, S. and Hu, X., "The influence of different knowledge-driven methods on landslide susceptibility mapping: A case study in the Changbai Mountain Area, Northeast China," Entropy **21**(4) (2019).

[14] Chen, W., Xie, X., Peng, J., Shahabi, H., Hong, H., Bui, D. T., Duan, Z., Li, S. and Zhu, A. X., "GIS-based landslide susceptibility evaluation using a novel hybrid integration approach of bivariate statistical based random forest method," Catena **164**(April 2017), 135–149 (2018).

[15] Chen, W., Xie, X., Wang, J., Pradhan, B., Hong, H., Bui, D. T., Duan, Z. and Ma, J., "A comparative study of logistic model tree, random forest, and classification and regression tree models for spatial prediction of landslide susceptibility," Catena **151**, 147–160 (2017).

[16] Reichenbach, P., Rossi, M., Malamud, B. D., Mihir, M. and Guzzetti, F., "A review of statistically-based landslide susceptibility models," Earth-Science Rev. **180**(March), 60–91 (2018).

[17] Pourghasemi, H. R., "Analysis and evaluation of landslide susceptibility : a review on articles published during 2005 – 2016 (periods of 2005 – 2012 and 2013 – 2016)" (2018).
[18] Pradhan, B., Seeni, M. I. and Kalantar, B., [Performance Evaluation and Sensitivity Analysis of Expert-Based, Statistical, Machine Learning, and Hybrid Models for Producing Landslide Susceptibility Maps] (2017).
[19] Nguyen, H., Mehrabi, M., Kalantar, B., Moayedi, H. and Abdullahi, M. M., "Potential of hybrid evolutionary approaches for assessment of geo-hazard landslide susceptibility mapping," Geomatics, Nat. Hazards Risk **10**(1), 1667–1693 (2019).
[20] Bui, D. T., Tuan, T. A., Hoang, N., Thanh, N. Q., Nguyen, D. B., Liem, N. and Pradhan, B., "Spatial prediction of rainfall-induced landslides for the Lao Cai area (Vietnam) using a hybrid intelligent approach of least squares support vector machines inference model and artificial bee colony optimization," 447–458 (2017).
[21] Shirzadi, A., Soliamani, K., Habibnejhad, M., Kavian, A., Chapi, K., Shahabi, H., Chen, W., Khosravi, K., Pham, B. T., Pradhan, B., Ahmad, A. and Ahmad, B. Bin., "Shallow Landslide Susceptibility Mapping," Sensors (Switzerland) **18**(11) (2018).
[22] Kavzoglu, T., Sahin, E. K. and Colkesen, I., "Landslide susceptibility mapping using GIS-based multi-criteria decision analysis, support vector machines, and logistic regression," Landslides **11**(3), 425–439 (2014).
[23] Kadavi, P. R. and Lee, C., "Application of Ensemble-Based Machine Learning Models to Landslide Susceptibility Mapping," Remote Sens. **10**(8), 1–18 (2018).
[24] Hong, H., Pradhan, B., Sameen, M. I., Kalantar, B., Zhu, A. and Chen, W., "Improving the accuracy of landslide susceptibility model using a novel region-partitioning approach," Landslides **15**(4), 753–772 (2018).
[25] Golkarian, A., Naghibi, S. A., Kalantar, B. and Pradhan, B., "Groundwater potential mapping using C5.0, random forest, and multivariate adaptive regression spline models in GIS," Environ. Monit. Assess. **190**(3) (2018).
[26] Kalantar, B., Ueda, N., Lay, U. S., Al-Najjar, H. A. H. and Halin, A. A., "Conditioning Factors Determination for Landslide Susceptibility," IEEE Int. Conf. Geosci. Remote Sens. (2019).
[27] Breiman, L., "Random forests," Mach. Learn. **45**(1), 5–32 (2001).
[28] Melville, B., Lucieer, A. and Aryal, J., "Object-based random forest classification of Landsat ETM+ and WorldView-2 satellite imagery for mapping lowland native grassland communities in Tasmania, Australia," Int. J. Appl. Earth Obs. Geoinf. **66**(May 2017), 46–55 (2018).
[29] Xu, R., Lin, H., Lü, Y., Luo, Y., Ren, Y. and Comber, A., "A modified change vector approach for quantifying land cover change," Remote Sens. **10**(10), 1–20 (2018).
[30] Chen, T., Trinder, J. C. and Niu, R., "Object-oriented landslide mapping using ZY-3 satellite imagery, random forest and mathematical morphology, for the Three-Gorges Reservoir, China," Remote Sens. **9**(4) (2017).
[31] Pourghasemi, H. R. and Rahmati, O., "Prediction of the landslide susceptibility: Which algorithm, which precision?," Catena **162**(May 2017), 177–192 (2018).
[32] Roodposhti, M. S., Aryal, J., Shahabi, H. and Safarrad, T., "Fuzzy Shannon entropy: A hybrid GIS-based landslide susceptibility mapping method," Entropy **18**(10) (2016).
[33] Tsangaratos, P. and Ilia, I., "Comparison of a logistic regression and Naïve Bayes classifier in landslide susceptibility assessments: The influence of models complexity and training dataset size," Catena **145**, 164–179 (2016).
[34] Soria, D., Garibaldi, J. M., Ambrogi, F., Biganzoli, E. M. and Ellis, I. O., "A 'non-parametric' version of the naive Bayes classifier," Knowledge-Based Syst. **24**(6), 775–784 (2011).
[35] Wu, X., Kumar, V., Ross, Q. J., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G. J., Ng, A., Liu, B., Yu, P. S., Zhou, Z. H., Steinbach, M., Hand, D. J. and Steinberg, D., [Top 10 algorithms in data mining] (2008).
[36] Landwehr, N., Hall, M. and Frank, E., "Logistic model trees," Mach. Learn. **59**(1–2), 161–205 (2005).
[37] L. Breiman, J. Friedman, C. J. Stone, R. A. O., "Classification Algorithms and Regression Trees," Classif. Regres. Trees, 246–280 (1984).
[38] Doetsch, P., Buck, C., Golik, P., Hoppe, N., Kramp, M., Laudenberg, J., Oberdörfer, C., Steingrube, P., Forster, J., Mauser, A., Dror, G., Boullé, M., Guyon, I., Lemaire, V., Vogel, D., Doetsch, P., Buck, C., Golik, P., Hoppe, N., et al., "Logistic Model Trees with AUC Split Criterion for the KDD Cup 2009 Small Challenge Human Language Technology and Pattern Recognition," JMLR Workshop Conf. Proc. **7**, 77–88 (2009).
[39] O'Brien, R. M., "A caution regarding rules of thumb for variance inflation factors," Qual. Quant. **41**(5), 673–690 (2007).