Predicting susceptibility to landslides under climate change impacts in metropolitan areas of South Korea using machine learning

Sang-Jin Park and Dong-kun Lee

ABSTRACT
Landslides cause considerable damage to life and property worldwide. In order to prevent and respond to landslides, it is necessary to identify vulnerable areas. This study identified areas that are likely to be damaged by landslides and aimed to predict future landslides. We compared and analyzed areas using machine learning (ML) algorithms, and conducted susceptibility mapping and landslide prediction using an algorithm that produced excellent results. For landslide predictions, the probability distribution of precipitation in the representative concentration pathway scenario 8.5 was used. We accounted for future uncertainties by using several regional climate model scenarios. Comparing the performances of different ML algorithms, the overall prediction accuracy of the random forest (0.932) was excellent. Susceptibility to landslides in the future determined using the random forest and five other regional climate models exhibited minor differences, but the average susceptibility increased over time. In addition, many urban areas are distributed around forest areas that have high landslide vulnerabilities, which provide important perspectives for urban and environmental planning.

1. Introduction
Along with flooding, landslides are a major cause of serious damage to life and property worldwide (Malamud et al. 2004; Gomez and Kavzoglu 2005; Lee and Pradhan 2007; Garcia-Rodriguez et al. 2008; Yilmaz 2010; Pham et al. 2020). In the future, when the impacts of climate change become more severe, sudden heavy rains could cause more damage due to landslides and flooding (Yılmaz 2009). Therefore, studies related to landslide-susceptibility assessments and responses are necessary to guide disaster reduction and management measures, including land use planning and
decision-making processes (Akgun 2012; Nsengiyumva et al. 2018; Dou et al. 2019b, 2020b).

Studies of the vulnerability or susceptibility to landslides have been conducted worldwide. Previous research has mostly focused on identifying the factors that cause landslides based on the conditions of the target site, collecting data on those factors, and analyzing vulnerability or sensitivity through statistical models, as shown in Table 1. A review of previous studies indicates that, while the target sites and methodologies differed, the contents of the studies were similar (Tien Bui et al. 2012; Zare et al. 2013; Kumar et al. 2016; Pham et al. 2016; Kornejady et al. 2017; Chen et al. 2017a; Hong et al. 2017; Nsengiyumva et al. 2018; Pham and Prakash 2019; Polykretis et al. 2019; Wang et al. 2020; Dou et al. 2020a).

There have been many great previous studies, but most of them focused on evaluating current susceptibility or sensitivity. Few studies have focused on predicting the future risk of landslides. In the field of disaster research, predictions or future prospects are important in terms of disaster management (UNISDR 2015). The probabilistic statistical techniques used in previous studies (Table 1) to predict and forecast landslides in a specific vulnerable area also have the potential for future studies (Pourghasemi et al. 2020). Therefore, this study aimed to predict future landslides in areas that have not received adequate attention in previous studies by using multiple climate change scenarios. A probabilistic statistical model was constructed to estimate landslide-susceptibility because it can consider the uncertainties in the calculations used for prediction (Barria et al. 2019).

Recently, machine learning (ML) algorithms have become popular and are being used extensively for the spatial prediction of diverse types of hazards (Chen et al. 2017a). Also, data-driven models, such as machine learning models, performed better and were considered more efficient than other approaches, such as expert opinion-based methods (Goetz et al. 2015; Pham et al. 2020). In this study, landslide susceptibility was assessed using five ML algorithms widely used in previous studies: Naive bayes classifier (NB), k-Nearest Neighbor (kNN), Decision Tree (DT), Random Forest (RF), and Support Vector Machine (SVM). Predictions of future landslides were then made by considering the probability distribution of precipitation data obtained from representative concentration pathway (RCP) climate change scenarios provided by the Intergovernmental Panel on Climate Change (IPCC) and regional climate models.

| Author, Year          | Study area                        | Method used in study                                                                 |
|-----------------------|-----------------------------------|---------------------------------------------------------------------------------------|
| Tien Bui et al. (2012)| Hoa Binh province, Vietnam        | Adaptive Neuro-Fuzzy Inference System                                                 |
| Zare et al. (2013)    | Vaz Watershed, Iran               | Multilayer perceptron and radial basic function                                       |
| Pham et al. (2016)    | Uttarakhand state, India.         | Naive Bayes Trees, Support Vector Machines                                           |
| Kumar et al. (2016)   | Indian Himalayas                  | Fuzzy–frequency ratio                                                                |
| Pham et al. (2017)    | Indian Himalayas                  | Multiple Perceptron Neural Networks                                                  |
| Kornejady et al. (2017)| Golestan Province, Iran           | Maximum Entropy                                                                      |
| Chen et al. (2017a)   | Langao County, China.             | Rotation forest ensembles, Naive Bayes Tree                                           |
| Hong et al. (2017)    | Chongren area, China.             | Frequency ratio, Certainty factor, Index of entropy                                   |
| Nsengiyumva et al. (2018)| Eastern Province, Rwanda         | Spatially different criteria evaluation methods                                       |
| Polykretis et al. (2019)| Mediterranean catchment, Greece  | Adaptive neuro-fuzzy modeling                                                        |
| Pham and Prakash (2019)| MuCang Chai, northern Vietnam     | Bagging-based Naive Bayes Trees                                                      |
| Dou et al. (2020a, 2020b)| Northern parts of Kyushu, Japan  | Support vector machine hybrid ensembles                                              |
| Wang et al. (2020)    | Sichuan Province, China           | Deep belief network (DBN)                                                            |
(RCMs) provided by the Korea Meteorological Administration (KMA). In addition, climate models have uncertainties because of uncertainty in the scenario values (Knutti and Sedláček 2013). This is the reason many studies have used the ensemble approach to consider uncertainty by using diverse scenarios (Parker 2013). The findings of this study can serve as a data source for formulating long-term policies for response and disaster management related to landslides.

2. Materials and method

2.1. Study area

Figure 1 shows the study area and locations where landslides have occurred. This study considered the metropolitan area of South Korea (33–38° N, 125–131° E), including the capital city of Seoul (36–38° N, 126–128° E). The metropolitan area of South Korea covers 11,851.26 km², accounting for 11.8% of the total area of Korea. At the end of October 2020, the number of residents registered in the metropolitan area was 26.3 million according to demographic data, accounting for 50.21% of the total population of Korea (Ministry of the Interior and Safety official website: https://www.mois.go.kr). The summer climate in Korea is generally hot and humid, especially from July to August, and typhoons are common from August to September. The annual cumulative precipitation during the summer (June, July, and August) is 892.1 mm, accounting for approximately 61% of the annual precipitation. The highest recorded daily rainfall in 37 years in Seoul was 354.7 mm, which occurred on August 2, 2020 (KMA official website: https://www.weather.go.kr).

Figure 1. Study area (metropolitan area including Seoul, South Korea) and landslide occurrences (n = 326, 2011–2017).
2.2. Data

As shown in Figure 2, nine factors were used in the study. The selected factors were the same as those used for analyzing the impact or sensitivity of landslides by other studies (Tien Bui et al. 2012; Zare et al. 2013; Kumar et al. 2016; Pham et al. 2016; Kornejady et al. 2017; Nsengiyumva et al. 2018; Polykretis et al. 2019). Topographic (elevation, slope, aspect, curvature), geologic (lithology), environmental (road area ratio, forest area ratio), and meteorological (daily maximum precipitation) data pertaining to these variables were collected for analysis from National Geographic Information Institute, Korea Institute of Geoscience and Mineral Resources, Korean Meteorological Administration, and National Disaster Management Research Institute (Table 2). By collecting data on these eight factors, as well as the landslide inventory, four topographic factors based on the digital elevation model (DEM) were created using a 10 m grid (Dou et al. 2015), and others including environmental factors were created using a 250 m grid. All data was resampled to a 250 m grid in consideration of the area of the target site. Because the dataset was created by matching the rainfall data with the landslide occurrence dates from the inventory, the size of the dataset used for analysis was the number of grids multiplied by the number of occurrences. Table 2 shows the source, type, and period of each factor. In addition, Figure 3 shows the mapping of variables used in the study, with July 27, 2011 mapped as an example for Daily Maximum Precipitation (DMP).

2.3. Landslide factors analysis

To control unnecessary factors used in the analysis prior to a landslide susceptibility assessment (LSA), multi-collinearity analysis was used to find the relationship among the factors (Bui et al. 2019, Dou et al. 2019a), and the use of information gain ratio (IGR) to determine the degree of factor influence on the results was analyzed. This is because multi-collinearity and influencing factor analysis using IGR affect the results...
and accuracy of the model (Zhou et al. 2018). Variance Inflation Factor (VIF) and tolerances were used to calculate the accuracy of multi-collinearity. Then, the influence of each factor was determined using the IGR technique (Zhou et al. 2018).

### 2.4. Machine learning algorithms and validation

Each of the five ML algorithms used in this study had their own characteristics. NB is a stochastic-statistical method based on Bayes’ rule, where prior probability is used to estimate the posterior probability. Bayes’ theorem is stated mathematically as follows: \( P(A|B) = \frac{P(A)P(B|A)}{P(B)} \). In that equation, \( P(A) \) and \( P(B) \) is the prior probability, \( P(B|A) \) is the likelihood, and \( P(A|B) \) is the posterior probability. kNN was developed by Cover and Hart (1967), is easy to run, and is as simple as NB (Jadhav and Channe 2016). The designation of the number \( k \), which is the proximity of the data points to one another, is important because it affects the result of the algorithm (Bhavsar and Ganatra 2012; Kim et al. 2012). DT is a popular ML algorithm resembling a tree and based on decision tree theory. It is useful for decision-making because it provides a simple representation of the results (DeFries and Chan 2000). RF is an algorithm that is often used in studies that use ML, combined with algorithms such as SVM and neural networks (e.g., Breiman 2001). SVM, devised by Cortes and Vapnik (1995), is a multi-purpose algorithm that can classify unlabeled datasets. It identifies and analyzes the characteristics of data clusters.

Using the receiver operating characteristic (ROC) curve score, the results of the LSA derived from the five ML algorithms were compared. ROC analysis was mainly used to assess model performance (Pham et al. 2016), the relationship between the false positive rate (1 – Specificity), and the true positive rate determining the model’s performance. The closer the average AUC (area under the ROC curve) is to 1, the higher the accuracy of the model (Chen et al. 2018).

### Table 2. Data acquisition for the landslide-susceptibility analysis.

| Category       | Factor (Abbreviation) | Source                  | Type     | Period     |
|----------------|-----------------------|-------------------------|----------|------------|
| Topographical  | Elevation (E)         | *NGII                   | Grid     | 2014       |
|                | Slope (S)             |                         |          |            |
|                | Aspect (A)            |                         |          |            |
|                | Curvature (C)         |                         |          |            |
| Geological     | Lithology (L)         | **KIGAM                 | Polygon  | –          |
| Environmental  | Roads area ratio (RA) | *NGII                   | Polyline | 2018       |
|                | Forest Area Ratio (FA)|                        | Polygon  | 2017       |
| Meteorological | Daily Maximum Precipitation (DMP) | ***KMA | Points | 2011–2013 |
|                | Precipitation from Regional Climate Models (*used for prediction) | Grid | 2026.1.1–2085.12.31 |
| Target         | Landslides (LS)       | ****NDMI                | Points   | 2011–2013  |

*NGII (National Geographic Information Institute)
**KIGAM (Korea Institute of Geoscience and Mineral Resources)
***KMA (Korean Meteorological Administration)
****NDMI (National Disaster Management Research Institute)
2.5. LSA using different algorithms

The landslide inventory was divided into inventories for landslide occurrence and non-occurrence. Because there was a considerable difference in frequency between the occurrence and non-occurrence areas, under-sampling was performed based on the occurrence area (He and Garcia 2009). The analysis was then conducted by dividing the data into a training set and a test set with a ratio of 70:30 (Dou et al. 2015, 2019b, 2020a, 2020b; Zhou et al. 2018). Tens of thousands of iterations were required to analyze all of the grids because under-sampling was performed prior to grid creation. Landslide susceptibility was assessed using five ML algorithms that have been used widely in recent years: NB, kNN, DT, RF, and SVM. Additionally, these five algorithms were used to account for the uncertainty of the model (Hao et al. 2019). Results were obtained through an analysis involving approximately 50,000 iterations.

Figure 3. Maps of variables used in this study: a) Elevation, b) Slope, c) Aspect, d) Curvature, e) Lithology, f) Roads area ratio, g) Forest area ratio, h) Daily maximum rainfall (ex. July 27 2011), i) Time series graph in summer season (Jun to Aug) of daily maximum rainfall in RCP 8.5 scenario according to the five different Regional Climate Models.
2.6. Predicting landslide susceptibility

Susceptibility was predicted using the highest performing algorithm. Precipitation data from RCP climate change scenario 8.5, obtained from five different RCMs (GRI Ms, HadGEM3-RA, RegCM4, SNURCM, and WRF), were used as the main variables for prediction. RCP climate change scenarios are scenarios using radiative forcing to measure the amount of carbon emissions that are the main cause of climate change, and there were four scenarios (2.6, 4.5, 6.0, 8.5). The higher the number, the higher the carbon emission scenario, with 8.5 indicating the scenario with carbon emission continuously carried out at the current level (IPCC 2014). Precipitation data for these five RCMs were obtained from the Regional Climate Detailing Project in East Asia (CORDEX-EA: Coordinated Regional Downscaling Experiment-East Asia, http://cordexea.climate.go.kr/cordex). The climate information portal of the KMA provided the RCM data (Table 2). We used the daily maximum rainfall amounts for the different RCP scenarios of the RCMs, and their probability distributions were used as inputs to determine the future susceptibility to landslides. The temporal targets for forecasting are the 2030s, 2050s, and 2080s, and daily maximum rainfall amounts for 10 years before and after each year were used (Figure 2).

3. Results

3.1. Multi-collinearity and influencing factor analysis

Table 3 shows the result of the multi-collinearity analysis. If the tolerances were less than 0.2 or the VIF was greater than 5, it was interpreted that there was multi-collinearity (O’Brien 2007). The VIF and tolerances of the factor ‘slope’ were 5.013 and 0.199, respectively. Since the VIF was greater than 5, and the tolerance was less than 0.2, the ‘slope’ was removed for LSA. As a result of performing the multi-collinearity analysis again without the ‘slope’, it was found that there was no multi-collinearity among the factors.

Figure 4 shows the result of influencing factor analysis and the average AUC value of the result of LSA by removing the influencing factor analysis result for low influence in order. As can be seen from the results (Figure 4a), the most influential variable was 0.45, which was shown as daily maximum precipitation (DMP). The variable with the least influence was aspect (0.05), which was much higher than the altitude (0.023) that most affected the colluvial landslide calculated by Zhou et al. (2018). The
factor’s influence is difficult to compare accurately because the number of variables used is different, but even after taking this into account, it was a high value. As a result of performing LSA ignoring factors having low influence, it can be seen that the AUC value is not significantly different (Figure 4b). Then, LSA was implemented by using all factors without factor ‘slope,’ which was removed at the multi-collinearity analysis. In addition, DMP appeared to be the most influential factor, similar to those studies that analyzed flooding in coastal areas (Park and Lee 2020). These results reiterate the importance of responding to heavy rain as part of disaster management.

3.2. Comparison of machine learning algorithms

As shown in Figure 5a, the accuracies of the average ROC curves produced from the five algorithms were: 0.822 (NB), 0.896 (kNN), 0.869 (DT), 0.932 (RF), and 0.866 (SVM). The accuracy of the model using RF was the highest of the 5 models; however, the results of the other models were also high. In addition, the graph depicting the density of RF versus ROC accuracies shows that the kurtosis of the graph is slightly higher than those of other algorithms. This indicates that the results obtained with RF were precise. Therefore, we used RF for the landslide susceptibility mapping and prediction. Figure 5b shows the ROC curves obtained using RF.

3.3. Predicting landslide susceptibility

The predictions were conducted using monthly average rainfall amounts. In this process, rainfall predicted to occur in the future was used as a density function to consider the uncertainty of future rainfall. The rainfall was estimated by substituting the kernel density in one month, and the model was used to calculate the monthly predicted susceptibility. By comparing the probability of monthly susceptibilities, we observed that the probability increased during the months of June, July, and August for most of the RCP 8.5 scenarios. Based on these data, the average probabilities of

Figure 4. a) Influencing factor analysis, and b) comparison of the average AUC value of the result of LSM by removing the influencing factor analysis result with low influence in order.
susceptibility for each scenario (8.5, 2030s, 2050s, and 2080s) during June, July, and August were calculated to create a probability map of the maximum susceptibility for the given scenario.

The results of the predictions are shown in Figure 6. These results were obtained for the RCP climate change scenario 8.5 using five different RCMs with the 2030s, 2050s, and 2080s as the target periods. The susceptibility was high, on average, when HadGEM3-RA was used. Moreover, when GRIMs were used, the susceptibility increased over time, while the remaining three RCMs had tendencies for reduced susceptibility. This is because the peak precipitation value appeared differently depending on the scenario considered by the RCMs (Figure 3i).

4. Discussion

4.1. Analysis of results from different ML algorithms

This study accounted for the model uncertainty by performing LSA using five different ML algorithms. The resulting performance was good for most models, with RF having the best performance, similar to the findings of previous studies such as those using supervised learning classification with spatially geographical data (Cracknell and Reading 2014; Naghibi et al. 2017; Chen et al. 2017b; Pourghasemi et al. 2020).

4.2. Difference in susceptibilities based on land cover type

We determined how susceptibility changed according to the type of land cover when the same land cover is maintained in the future using five different RCMs under climate change scenario 8.5 (Figure 7a). The risk in forest areas is largely due to the occurrence of landslides. The uncertainty of the results increased over time. This can be attributed to the uncertainty of climate models and the distribution characteristics of the precipitation values in each RCM.

Figure 5. (a) Comparison of the accuracy scores of each ML algorithm and (b) plot of ROC curves of the RF algorithm.
Figure 6. Landslide susceptibility changes for different RCMs under RCP climate change scenario 8.5 from the 2030s to the 2080s.
Moreover, it is important to know which areas among other land cover types near the forest are more sensitive in the future. Figure 7b shows how three land-cover types near the forest are relatively susceptible during the target periods under the RCP 8.5 scenario. The uncertainty increased with time; however, the urban area remained more susceptible than other areas. This is because many urban areas are distributed around forest areas that have high susceptibilities to landslides. More indirectly, it could be explained that due to the high urbanization around the forest area, economic damage is high when landslides occur due to heavy rains. These results highlight the need for future efforts in land-use planning to reduce landslide susceptibility in urban areas located near forest areas.

5. Conclusions

This study evaluated landslide susceptibility in the metropolitan area that includes Seoul, South Korea. Prior to LSA, multi-collinearity was analyzed among eight factors. As a result of the multi-collinearity analysis, the factor ‘slope’ had high multi-collinearity, and it was removed for LSA. Then, to improve the ML model’s performance, the influencing factors were addressed by using the IGR technique. The most influential factor was the daily maximum precipitation (0.45). In addition, the results of performing LSA while ignoring factors having low influence in turn showed that the AUC value is not significantly different from each result. Based on evaluations using five different ML algorithms with seven factors, the average AUC using RF exhibited the best performance (0.932). To predict future landslide susceptibility, projected future precipitation values from the RCP climate change scenario were used, and the target years were set to the 2030s, 2050s, and 2080s. Additionally, various future possibilities were predicted based on five different RCMs. This allowed us to consider future uncertainties. The landslide susceptibility generally increased over time, with the exception of some results where RegCM4, SNURCM, and WRF were used.

From this study, we arrived at the following additional inferences. First, among the five ML algorithms used to assess landslide susceptibility, RF had the best performance. Similar studies using supervised learning classification have shown comparable
results. Second, we found that it is necessary to reduce the landslide susceptibility in forest areas, based on an analysis of changes in landslide susceptibility over the target period (2030s, 2050s, and 2080s) and land cover types (urban, agricultural, forest, and grassland). We found that urban areas were more susceptible than other land-cover types because they were distributed around the forest areas that were estimated as more susceptible to landslides. It might indicate that many areas were urbanized around the forest area, and it also highlighted the need for future efforts in land-use planning to reduce landslide susceptibility.

Additionally, the result and process of this study reveal some limitations. First, since the purpose of this study was to predict future landslide susceptibility, the effect of data resolution was not considered, although the resolution of the area of the target site was larger than the resolution of the two other studies (Dou et al. 2015, Chang et al. 2019); thus, the resolution of data should be considered in future studies because it is important for determining landslide susceptibility. Second, we predicted future landslides by assuming that the socioeconomic factors (as indicated by land cover type) did not change. In the future, socioeconomic factors may become more important for determining susceptibility; therefore, these factors should be considered in future studies.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the Korea Environment Industry & Technology Institute (KEITI) through the Decision Support System Development Project for Environmental Impact Assessment, funded by the Korea Ministry of Environment (MOE) (No. 2020002990009).

Data availability statement

The data that support the findings of this study are available upon reasonable request from the authors.

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 Izmir, Turkey. Landslides. 9(1):93–106.
Barría P, Cruzat ML, Cienfuegos R, Gironás J, Escauriaza C, Bonilla C, Moris R, Ledezma C, Guerra M, Rodríguez R, et al. 2019. From multi-risk evaluation to resilience planning: The case of central Chilean Coastal Cities. Water. 11(3):572.
Bhavsar H, Ganatra A. 2012. A comparative study of training algorithms for supervised machine learning. Int J Soft Comput Eng. 2(4):74–81. https://www.ijsce.org/wp-content/uploads/papers/v2i4/D0887072412.pdf.
Breiman L. 2001. Random forests. Mach Learn. 45(1):5–32.
Bui DT, Shirzadi A, Shahabi H, Geertsema M, Omidvar E, Clague JJ, Pham BT, Dou J, Asl DT, Ahmad BB, et al. 2019. New ensemble models for shallow landslide susceptibility modeling in a semi-arid watershed. Forests. 10(9):743.

Chang KT, Merghadi A, Yunus AP, Pham BT, Dou J. 2019. Evaluating scale effects of topographic variables in landslide susceptibility models using GIS-based machine learning techniques. Sci Rep. 9(1):12221–12296.

Chen W, Pourghasemi HR, Naghibi SA. 2018. Prioritization of landslide conditioning factors and its spatial modeling in Shangnan County, China using GIS-based data mining algorithms. Bull Eng Geol Environ. 77(2):611–629.

Chen W, Shirzadi A, Shahabi H, Ahmad B, Bin Zhang S, Hong H, Zhang N. 2017a. A novel hybrid artificial intelligence approach based on the rotation forest ensemble and naïve Bayes tree classifiers for a landslide susceptibility assessment in Langao County, China. Geomat Nat Hazards Risk. 8(2):1955–1977.

Chen W, Xie X, Wang J, Pradhan B, Hong H, Bui DT, Duan Z, Ma J. 2017b. 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.

Cortes C, Vapnik V. 1995. Support-vector networks. Mach Learn. 20(3):273–297.

Cover T, Hart P. 1967. Nearest neighbor pattern classification IEEE Trans Inf Theory. 13:21–27.

Cracknell MJ, Reading AM. 2014. Geological mapping using remote sensing data: A comparison of five machine learning algorithms, their response to variations in the spatial distribution of training data and the use of explicit spatial information. Comput Geosci. 63:22–33.

DeFries RS, Chan JCW. 2000. Multiple criteria for evaluating machine learning algorithms for land cover classification from satellite data. Remote Sens Environ. 74(3):503–515.

Dou J, Yamagishi H, Pourghasemi HR, Yunus AP, Song X, Xu Y, Zhu Z. 2015. An integrated artificial neural network model for the landslide susceptibility assessment of Osado Island, Japan. Nat Hazards. 78(3):1749–1776.

Dou J, Yunus AP, Bui DT, Merghadi A, Sahana M, Zhu Z, Chen CW, Han Z, Pham BT. 2020a. Improved landslide assessment using support vector machine with bagging, boosting, and stacking ensemble machine learning framework in a mountainous watershed, Japan. Landslides. 17(3):641–658.

Dou J, Yunus AP, Merghadi A, Shirzadi A, Nguyen H, Hussain Y, Avtar R, Chen Y, Pham BT, Yamagishi H. 2020b. Different sampling strategies for predicting landslide susceptibilities are deemed less consequential with deep learning. Sci Total Environ. 720:137320.

Dou J, Yunus AP, Tien Bui D, Merghadi A, Sahana M, Zhu Z, Chen CW, Khosravi K, Yang Y, Pham BT. 2019a. Assessment of advanced random forest and decision tree algorithms for modeling rainfall-induced landslide susceptibility in the Izu-Oshima Volcanic Island, Japan. Sci Total Environ. 662:332–346.

Dou J, Yunus AP, Tien Bui D, Sahana M, Chen C-W, Zhu Z, Wang W, Pham BT. 2019b. Evaluating GIS-based multiple statistical models and data mining for earthquake and rainfall-induced landslide susceptibility using the LiDAR DEM. Remote Sens. 11(6):638.

Garcia-Rodriguez MJ, Malpica JA, Benito B, Diaz M. 2008. Susceptibility assessment of earthquake-triggered landslides in El Salvador using logistic regression. Geomorphology. 95(3–4):172–191.

Goetz JN, Brenning A, Petschko H, Leopold P. 2015. Evaluating machine learning and statistical prediction techniques for landslide susceptibility modeling. Comput Geosci. 81:1–11.

Gomez H, Kavzoglu T. 2005. Assessment of shallow landslide susceptibility using artificial neural networks in Jabonosa River Basin, Venezuela. Eng Geol. 78(1–2):11–27.

Hao T, Elith J, Guillera-Arroita G, Lahoz-Monfort JJ. 2019. A review of evidence about use and performance of species distribution modelling ensembles like BIOMOD Divers. Divers Distrib. 25(5):839–852.

He H, Garcia EA. 2009. Learning from imbalanced data IEEETrans. Knowl Data Eng. 21:1263–84.
Hong H, Chen W, Xu C, Youssef AM, Pradhan B, Tien Bui D. 2017. Rainfall-induced landslide susceptibility assessment at the Chongren area (China) using frequency ratio, certainty factor, and index of entropy. Geocarto International. 32(2):1–154.

IPCC. 2014. Climate change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, R.K. Pachauri and L.A. Meyer, editors]. IPCC, Geneva, Switzerland, p. 151.

Jadhav SD, Channe HP. 2016. Comparative Study of K-NN, Naive Bayes and Decision Tree Classification Techniques. Int J Sci Res (IJSR). 5(1):1842–1845.

Kim J, Kim B-S, Savarese S. 2012. Comparing image classification methods: K-nearest-neighbor and support-vector-machines. Appl Mathematics Electr Comput Eng :133–138.

Knutti R, Sedláček J. 2013. Robustness and uncertainties in the new CMIP5 climate model projections. Nature Clim Change. 3(4):369–373.

Korea Meteorological Administration official website: https://www.weather.go.kr.

Kornejady A, Ownegh M, Bahremand A. 2017. Landslide susceptibility assessment using maximum entropy model with two different data sampling methods. Catena. 152:144–162.

Kumar S, Srivastava PK, Gore A, Singh MK. 2016. Fuzzy–frequency ratio model for avalanche susceptibility mapping. Int J Digital Earth. 9(12):1168–1184.

Lee S, Pradhan B. 2007. Landslide hazard mapping at Selangor, Malaysia using frequency ratio and logistic regression models. Landslides. 4(1):33–41.

Malamud BD, Turcotte DL, Guzzetti F, Reichenbach P. 2004. Landslide inventories and their statistical properties. Earth Surf Process Landforms. 29(6):687–711.

Ministry of the Interior and Safety official website: https://www.mois.go.kr.

Naghibi SA, Ahmadi K, Daneshi A. 2017. Application of support vector machine, random forest, and genetic algorithm optimized random forest models in groundwater potential mapping. Water Resour Manage. 31(9):2761–2775.

Nsengiyumva JB, Luo G, Nahayo L, Huang X, Cai P. 2018. Landslide susceptibility assessment using spatial multi-criteria evaluation model in Rwanda. Int J Environ Res Public Health. 15(2):243.

O’Brien RM. 2007. A caution regarding rules of thumb for variance inflation factors. Qual. Quant. 41:673–690.

Park SJ, Lee DK. 2020. Prediction of coastal flooding risk under climate change impacts in South Korea using machine learning algorithms. Environ Res Lett. 15(9):094052. https://doi.org/10.1088/1748-9326/ab5b3.

Parker WS. 2013. Ensemble modeling, uncertainty and robust predictions. WIREs Clim Change. 4(3):213–223.

Pham BT, Bui D, Prakash I, Dholakia MB. 2016. Evaluation of predictive ability of support vector machines and naive Bayes trees methods for spatial prediction of landslides in Uttarakhand state (India) using GIS. J Geomat. 10(1):71–79. https://isgindia.org/JOG/abstracts/APR-2016/pap012.pdf.

Pham BT, Prakash I. 2019. A novel hybrid model of Bagging-based Naïve Bayes Trees for landslide susceptibility assessment. Bull Eng Geol Environ. 78(3):1911–1925.

Pham BT, Prakash I, Dou J, Singh SK, Trinh PT, Tran HT, Le TM, Van Phong T, Khoi DK, Shirzadi A, et al. 2020. A novel hybrid approach of landslide susceptibility modelling using rotation forest ensemble and different base classifiers. Geocarto Int. 35(12):1267–1292.

Pham BT, Tien Bui D, Prakash I, Dholakia MB. 2017. Hybrid integration of Multilayer Perceptron Neural Networks and machine learning ensembles for landslide susceptibility assessment at Himalayan area (India) using GIS. Catena. 149:52–63.

Polykretis C, Chalkias C, Ferentinou M. 2019. Adaptive neuro-fuzzy inference system (ANFIS) modeling for landslide susceptibility assessment in a Mediterranean hilly area. Bull Eng Geol Environ. 78(2):1173–1187.

Pourghasemi HR, Kariminejad N, Amiri M, Edalat M, Zarafshar M, Blaschke T, Cerda A. 2020. Assessing and mapping multi-hazard risk susceptibility using a machine learning technique. Sci Rep. 10(1):3203. https://doi.org/10.1038/s41598-020-60191-3.
Tien Bui D, Pradhan B, Lofman O, Revhaug I, Dick OB. 2012. Landslide susceptibility mapping at Hoa Binh province (Vietnam) using an adaptive neuro-fuzzy inference system and GIS. Comput Geosci. 45:199–211.

UNISDR. 2015. Making Development Sustainable: The Future of Disaster Risk Management. Global Assessment Report on Disaster Risk Reduction. Geneva, Switzerland: United Nations Office for Disaster Risk Reduction (UNISDR). https://www.preventionweb.net/english/hyogo/gar/2015/en/gar-pdf/GAR2015_EN.pdf.

Wang W, He Z, Han Z, Li Y, Dou J, Huang J. 2020. Mapping the susceptibility to landslides based on the deep belief network: a case study in Sichuan Province, China. Nat Hazards. 103(3):3239–3261.

Yılmaz I. 2009. Landslide susceptibility mapping using frequency ratio, logistic regression, artificial neural networks and their comparison: a case study from Kat landslides (Tokat-Turkey). Comput Geosci. 35(6):1125–1138.

Yılmaz I. 2010. The effect of the sampling strategies on the landslide susceptibility mapping by conditional probability and artificial neural networks. Environ Earth Sci. 60(3):505–519.

Zare M, Pourghasemi HR, Vafakhah M, Pradhan B. 2013. Landslide susceptibility mapping at Vaz Watershed (Iran) using an artificial neural network model: a comparison between multilayer perceptron (MLP) and radial basic function (RBF) algorithms. Arab J Geosci. 6(8):2873–2888.

Zhou C, Yin K, Cao Y, Ahmed B, Li Y, Catani F, Pourghasemi HR. 2018. Landslide susceptibility modeling applying machine learning methods: a case study from Longju in the Three Gorges Reservoir area, China. Comput Geosci. 112:23–37.