RESEARCH

Open Access

Landslide susceptibility assessment of South Korea using stacking ensemble machine learning

Seung-Min Lee¹ and Seung-Jae Lee^{1*}

Abstract

Background Landslide susceptibility assessment (LSA) is a crucial indicator of landslide hazards, and its accuracy is improving with the development of artificial intelligence (AI) technology. However, the AI algorithms are inconsistent across regions and strongly dependent on input variables. Additionally, LSA must include historical data, which often restricts the assessment to the local scale and single landslide events.

Methods In this study, we performed an LSA for the entirety of South Korea. A total of 30 input variables were constructed, consisting of 9 variables from past climate model data MK-PRISM, 12 topographical factors, and 9 environmental factors. Sixteen machine learning algorithms were used as basic classifiers, and a stacking ensemble was used on the four algorithms with the highest area under the curve (AUC). Additionally, a separate assessment model was established for areas with a risk of landslides affecting areas larger than 1 ha.

Results The highest-performing classifier was CatBoost, with an AUC of ~ 0.89 for both assessments. Among the input variables, distance of road, daily maximum precipitation, digital elevation model, and soil depth were the most influential. In all landslide events, CatBoost, lightGBM, XGBoost, and Random Forest had the highest AUC in descending order; in large landslide events, the order was CatBoost, XGBoost, Extra Tree, and lightGBM. The stacking ensemble enabled the construction of two landslide susceptibility maps.

Conclusions Our findings provide a statistical method for constructing a high-resolution (30 m) landslide susceptibility map on a country scale using diverse natural factors, including past climate data.

Keywords Landslide, Susceptibility model, Stacking ensemble, Machine learning

Introduction

Over 63% of South Korea's territory is mountainous and susceptible to landslides. Additionally, the landslide hazard is continuously rising due to the spread of forests and buildings in mountainous areas, increasing the severity of damage caused by landslides. Landslides typically cause more damage in urban areas, as evidenced by casualties

¹ National Center for AgroMeteorology, Seoul National University, 109, Building 36, Gwanak-Ro 1, Gwanak-Gu, Seoul 08826, Korea and property damage from a landslide in Seoul's Woomyeonsan and Chuncheon's Majeoksan caused by heavy rainfall in July 2011 (Pradhan and Kim 2014; Lee and Kim 2016). Therefore, it is important to assess the landslide susceptibility of the country to prevent economic and social losses.

Landslide susceptibility assessment (LSA) can be performed using various methods. Landslide susceptibility began to be analyzed statistically in the 2000s with the development of Geographic Information System (GIS) technology. Researchers have proposed general concepts for landslide mapping, such as maps showing spatiotemporal incidence and landslide prediction (Wang et al.



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

^{*}Correspondence:

Seung-Jae Lee

sjlee@ncam.kr

2005; Chacón et al. 2006), and methods to perform classification using GIS-based zonation into different susceptibilities (Kanungo et al. 2012). Subsequently, methods have been proposed to construct and validate models using an analytical hierarchy process (AHP) based on expert experience or by weighting input variables based on subjective judgment (Yang et al. 2006; Chae et al. 2009; Kayastha et al. 2013). AHP has the advantage of a structured decision approach, which includes a disagreement process for variables. However, since experts subjectively judge the weights between variables, objectivity and consistency of the model may be impacted.

With landslides increasing due to weather anomalies, recent studies have been seeking to improve accuracy by introducing machine learning (ML) technology (Ado et al. 2022). Hong et al. (2019) performed LSA using regression analysis and ML methods; Wang et al. (2021) evaluated the influence of landslide triggering factors using a maximum entropy ML model. Ageenko et al. (2022) used support vector machine (SVM), random forest (RF), and linear regression (LR) algorithms to construct a landslide susceptibility map of the eastern region of Denmark's Jutland Peninsula. Previous studies employing ML algorithms used only a small number of algorithm types, and the best-performing type was different each time. This indicates that the optimal algorithm and performance may vary with the regional characteristics of each LSA application. Since LSA constitutes important data that can be used as an objective indicator to confirm the characteristics of landslides in the target area, objectivity and consistency must be ensured by the methodology.

Recent studies have shown that LSA using deep neural networks such as convolutional (CNN) and recurrent neural networks (RNN) can outperform LSA using ML algorithms. In LSA using CNN algorithms, the core landslide location is collected as an aerial photograph or satellite image, and then a learning model is built together with images of landslide risk factors (Sameen et al. 2020; Zhang et al. 2022). The RNN family has strengths in processing time series data with long temporal dependencies and can analyze dynamic changes over time when continuous information is required, making it useful for LSA (Wang et al. 2020; Ji et al. 2023). However, LSA using deep neural networks is often used in parallel with ML algorithms due to the complexity of the algorithm structure, which makes it difficult to clearly explain the effects of landslide triggers.

ML algorithms in most cases perform relatively poorly compared to deep neural network algorithms. To overcome this limitation, recent studies have proposed stacking ensembles of ML algorithms to improve performance (Wolpert 1992; Wang et al. 2011). Hu et al (2020) used four ML algorithms and a stacking ensemble to confirm the performance improvement gained by combining two or more classifiers. However, since the performance of a stacking ensemble depends on the performance of a single classifier, there are cases where the performance is degraded by the combination of algorithms (Dou et al. 2020). The selection process for the single classifier therefore plays a very important role in stacking ensembles.

LSA represents the spatial probability of landslide occurrence based on a range of input variables. Additionally, the risk of landslide occurrence is mainly influenced by past events; therefore, variables that can represent long-term historical data are required (Ageenko et al. 2022). While the input variables in LSA must be able to sufficiently explain the spatial characteristics of landslide occurrence points, they do not necessarily have to be proportional to the number of characteristics. Rotigliano et al. (2011) used only 3 variables for their assessment, whereas Rossi et al. (2010) used 51. Furthermore, to adequately reflect the spatial characteristics of the landslide occurrence points, the spatial resolution of the input variables must be guaranteed to a certain level. Arnone et al. (2016) proved that LSA can be effectively performed when the spatial resolution of the input variables is between 10 and 50 m; however, they stated that higher resolution does not necessarily lead to better performance.

LSA in South Korea has mainly been localized (Kadavi et al. 2019; Hakim et al. 2022) or focused on individual landslide events (Vasu and Lee 2016; Lee et al. 2017). However, susceptibility assessment results may be biased if large-scale data on landslide occurrence points are collected from only specific cases or areas (Lin et al. 2021; Loche et al. 2022).

Among the factors that trigger landslides, rainfall is the most significant. Landslides occur after prolonged or intense rainfall that causes water to penetrate the shallow stratum within 2 m from the surface and leads to a collapse of soil or rock (Kim and Lee 2023). Data on the rainfall factor are mainly acquired through numerical weather models to generate future data, or from observation stations to obtain past data. Historical climate data that describe the characteristics of rainfall are needed for LSA (Sobie 2020), which covers both past and future domains. However, many recent studies on landslide susceptibility either did not consider rainfall data or used rainfall data from a single point in time (Bruzon et al. 2021; Hodasova and Bednarik 2021; Wang et al. 2021). To avoid spatial or temporal bias, LSA must be performed on landslide cases and triggering factors occurring in various regions over a long period.

This study aims to perform LSA and explain the effects of input variables for the entire territory of South Korea by constructing an optimal model using a stacking ensemble of ML algorithms. This approach was used to explain the effects of input variables while aiming for high performance. We attempted to use unbiased information by collecting landslide cases that occurred nationwide over a long period of time. The input variables of the model were preprocessed terrain, environment, and climate data. For the susceptibility model, we selected the highest-performing algorithm among the ML algorithms and performed an ensemble process, following which the influence of input variables was analyzed by variable importance. The final constructed sensitivity model was applied to input variables across the country to create a map of sensitivity to landslides in South Korea.

Materials and methods

Study area

Figure 1 shows a map of South Korea, the target area of this study. Landslides in South Korea have mostly occurred in thin surface layers (\sim 1-m thickness) due to rainfall (Kim et al. 2000, 2020). The frequency of localized heavy rains is increasing in the country, enhancing the risk of landslides as well as causing life and property

damage in cities and residential areas where thin surface layers are present.

Most of South Korea's rainfall occurs from June to October, and the increase in heavy rainfall events due to the changing weather has caused mountain disasters, such as landslides and debris flows, to occur with an annual frequency.

Landslide occurrence points

In this study, the postal address data of landslide occurrence points from 2011 to 2020 provided by the Korea Forest Service were used. Since the address data are provided without specific coordinates, geocoding had to be performed based on the cadastral map. To this end, we designated the corresponding cadastral polygon on the cadastral map of Korea and used the centroid of the landslides polygon. In South Korea, more than 90% of landslides occur during heavy summer rainfall, when rainfall-induced ground erosion causes debris flows, slides, and land creep on slopes of 10° to 50° (Kim and Chae 2009). We therefore focused on landslides caused by these external factors. We performed geocoding using the centroid of the polygons in the cadastral map, thereby constructing a total of 4,347 points (landslide events). We



Fig. 1 LSA experimental area, covering all of South Korea. A – E represent the administrative districts of South Korea (A: Gyeonggi, B: Gangwon, C: Chungcheong, D: Jeolla, E: Gyeongsang). Locations with landslide occurrence over the past 10 years (2011–2020) are depicted by red points and locations without landslide occurrence are depicted by green points

retained only those landslides that occurred between June and October, the period of intensive summer rainfall, and on slopes of 10° to 50°, and cases with missing addresses were excluded, resulting in a total of 3,174 points. LSA requires standards not only for points where landslides occurred but also for points of non-occurrence. After confirming that each cadastral map polygon had an extent of at most 1 km, we therefore randomly generated a non-occurrence point within a range > 2 km from each occurrence. This ensured that an equal number of occurrence and non-occurrence points was available without any two occupying the same cadastral polygon, and yielded a total of 6,348 points (Fig. 1). The 3,174 constructed landslide data points had varied impact scales. While there were small-scale landslides with an impact area of 0.01 ha, there were also large-scale landslide events with an impact area of up to 20 ha. Since these landslides vary in scale, we used separate scale classes in constructing the models, as their characteristics were expected to differ. As there is no particular standard for the scale of landslide impacts in South Korea, a separate susceptibility assessment model was built targeting 785 landslide cases of at least 1 ha size (large).

Terrain data

LSA is strongly influenced by the surrounding terrain elements of the area when a landslide event occurs. We therefore constructed the input variables using the elevation data to evaluate the terrain elements. For terrain data, we used the 30-m spatial resolution digital elevation model (DEM) data constructed by the United States Geological Survey during the shuttle radar topography mission (SRTM). The slope, aspect, profile, and plan curvature were derived from the DEM. The slope is calculated using dx, the difference in elevation from east to west, and dy, the difference in elevation from north to south. It is then converted to degrees using the arctangent function. The magnitude of the slope is expressed as a value between 0° and 90°, with values closer to 90° indicating a steeper slope. At this point, the aspect is determined by taking the arctangent of the ratio of dy to dx, expressing it as an angle in the plane. Aspect values range from 0° to 360°, depending on their clockwise direction relative to north. Curvature is a variable that indicates whether the target location's terrain is convex or concave. Curvature was denoted by A to I in order on a 3×3 grid, representing DEM values, and the profile and plan curvature were calculated using Eqs. (3) and (4), respectively (Table 1). Terrain ruggedness index (TRI) and topographic position index (TPI) were calculated using N, representing the central elevation, and N_i , representing the surrounding eight elevation values. Topographic wetness index (TWI) and stream power index (SPI) were calculated from A_s , representing the upslope contributing area, and tan β , representing the slope. Valley depth was determined as the vertical distance difference between the target point and the nearest watershed boundary. Soil drainage and effective soil depth were sourced from the Korean Soil Information System.

Environmental data

The environmental data were constructed according to previous research showing that a closer distance of an environmental variable to the landslide occurrence point corresponds to a greater contribution to the risk of landslide occurrence (Reichenbach et al. 2018). Line data for rivers, strata, and roads were derived from the continuous numerical topographic map produced by the Ministry of Land, Infrastructure, and Transport, and the shortest distance to all pixels was calculated to construct the data. Additionally, since landslide-susceptible areas are greatly influenced by surrounding forests, we also used forest data as input variables. These consisted of forest type (FRTP), diameter class (DMCLS), age class (AGCLS), forest density (DNST), and forest height (HEIGHT), and were based on the forest map provided by the Korea Forest Service. Furthermore, the normalized difference vegetation index (NDVI), the most common index for representing forests, was constructed based on Sentinel-2 satellite imagery taken during clear summer days.

Climate data

Most landslides that occur in South Korea are debris flows. Therefore, precipitation is the most important factor affecting landslides. Precipitation data exist in various forms such as actual observations and historical climate data. In this study, we employed long-term historical precipitation data to perform LSA. Climate data were sourced from the modified Korea parameter elevation relationships on independent slopes model (MK-PRISM), which uses the past long-term precipitation patterns in the target area as input data. MK-PRISM employs high-resolution (1 km) gridded observational data of the Korean Peninsula. We used data from 537 observation points, consisting of 75 automated surface observing systems and 462 automatic weather stations, over the period 2000-2019 (Kim et al. 2012). Considering the distance, altitude, aspect, and oceanicity of each observation point, the data were interpolated to construct grid data at a 1-km resolution. Monthly total precipitation (RN) represented the aggregate of total precipitation for July, August, and September (RN-july, RN-aug, RN-sep) the months when landslides mainly occurred during 2000-2019. The daily maximum precipitation (RX1day) and 5-day maximum precipitation

Parameter		Description		Formula
Topography	SRTM	Altitude data of 30-m resolution DEM		
	Slope	Degree of steepness of the slope	(1)	$\tan^{-1}\sqrt{\left(\frac{dz}{dx}\right)^2+\left(\frac{dz}{dy}\right)^2}$
	Aspect	Direction of the slope	(2)	$\tan^{-1} \sqrt{\frac{dy}{dx}}$
	Profile curvature	Curvature in the direction perpendicular to the direction of inclination with plan curvature	(3)	$-2\frac{\left(DG^{2}+EH^{2}+FGH\right)}{\left(G^{2}+H^{2}\right)}$
	Plan curvature	Horizontal curvature with respect to the direction of inclination with lateral curvature	(4)	$2\frac{\left(DH^2+EG^2+FGH\right)}{\left(G^2+H^2\right)}$
	TRI	Index of concave or convex conformation of the terrain	(5)	$\sum_{i=1}^{8} N-N_i $
	TPI	Index of degree to which a surface is soft or bumpy	(6)	$N - \sum_{i=1}^{8} \frac{N_i}{8}$
	SPI	Index of the degree of movement and erosion of sediment due to surface runoff	(7)	$A_{ m s}- anoldsymbol{eta}$
	TWI	Index of the effects of terrain on runoff flow	(8)	$\ln \frac{A_s}{\tan \theta}$
	Valley depth	Difference of vertical distance between water channel network and DEM		
	Soil drainage	Duration or frequency of soil being unsaturated by water		
	Soil depth	Vertical depth of soil layer, at a depth at which the roots of the crop can extend sufficiently into the ground		
Environment	Distance to river	Distance from landslide location to water boundary		
	Distance to fault	Distance from landslide location to fault		
	Distance to road	Distance from landslide location to road		
	FRTP	Forest type		
	DMCLS	Forest tree diameter class		
	AGCLS	Forest age class		
	DNST	Forest density		
	HEIGHT	Forest canopy height		
	NDVI	Normalized difference vegetation index-based on Sentinel-2 data	(9)	<u>NIR—Red</u> NIR+Red
Climate	Monthly total precipitation (RN)	Total precipitation in each month, July–September		
	RX1day	Highest precipitation amount in 1 day	(10)	max(RR _{ij})
	RX5day	Highest precipitation amount in 5 days	(11)	$\max(RR_{ki})$
	SDII	Average precipitation on wet days	(12)	$sum\left(\frac{RR_{wj}}{watday}\right)$
	R80	Number of days per year when precipitation \geq 80 mm	(13)	$count(RR_{ii} > 80mm)$
	RD95P	Number of days when daily precipitation is greater than during the top 95% of the reference period	(14)	sum if $\left(\frac{\text{RR}_{wj} > 95\%}{wetday}\right)$
	RD99P	Number of days when daily precipitation is greater than during the top 99% of the reference period	(15)	$sum if\left(\frac{RR_{wj}>99\%}{wetday}\right)$

Та	b	e 1	Descriptions	of	LSA input p	arameters	and	cal	cu	lation	formu	las
----	---	-----	--------------	----	-------------	-----------	-----	-----	----	--------	-------	-----

(RX5day) were sourced from the maximum values recorded between 2000 and 2019. The simple daily intensity index (SDII) is the annual total precipitation divided by the total number of wet days and was constructed based on the total wet days and precipitation for 2000–2019. The number of days when daily precipitation was \geq 80 mm (R80) was computed by simply summing the number of these days in a year. Number of days when daily precipitation is greater than during the top 95% (RD95P) and 99% (RD99P) was constructed based on the number of those days in the base period.

Construction of ML model

This study consisted of four stages (Fig. 2). In the first stage, the raw data for input variables and landslide occurrence points were collected. In the second stage, the input variables were calculated from each raw data source and then preprocessed to be applied to the same



Fig. 2 Flowchart showing the methodology for landslide susceptibility modeling

model. In the third stage, the ML algorithm was executed to tune the model, and in the final stage, landslide susceptibility maps for the entire country area were produced. LSA was based on the landslide events that occurred in South Korea. A diverse set of factors can be considered in the assessment. According to Reichenbach et al. (2018), landslides are mainly influenced by geology, hydrology, land cover, and topography, and depending on the occurrence area, weather and climatic conditions may also be considered. In the present study, the three factors of terrain, surrounding environment, and climate were used (Table 1). Equations (1) to (15) listed in Table 1 represent the formulas for calculating assessment parameters connected to the three fundamental factors. All variables underwent a resampling process for conversion into the same grid form and were constructed as GeoTIFF data with a spatial resolution of 30 m (Fig. 3).

The constructed susceptibility model utilized 16 ML algorithms supported by the Python library Pycaret (Table 2). Pycaret, which applies auto-machine learning, significantly reduces the processing time by automating tasks such as model selection and hyperparameter tuning, which are required for a single ML. Additionally, it allows the simultaneous use of libraries previously used for ML in Python development environments, such as Scikit-learn, XGBoost, and SpaCy.

The parameters for each ML algorithm were determined using K-fold cross-validation without going through a specific pipeline. Each ML algorithm crossed various parameters to automatically select the best performance case. Typically, ML algorithms divide training and validation data in a 50–50 to 90–10 ratio, depending on the conformation of the training data (Joseph 2022). Since we used 3,174 cases of 10-year landslides and a total of 6,348 cases, including non-occurrence cases, in our training model, we decided that sufficient training data were available. We also decided to use a 70–30 ratio to more rigorously validate the training model.

This study used the stacking ensemble approach, based on the ensemble with regard to the results of each ML algorithm. Stacking ensemble, also known as stacked generalization, is an ensemble learning technique that combines multiple classification or regression models via a meta-classifier or meta-regressor. The fundamental principle is to learn from the prediction of various base models and then to use another model to combine these predictions, aiming to achieve better generalization performance than could be achieved by any single base model (Wolpert 1992). An ensemble of ML algorithms has higher performance compared to a single algorithm, as it combines the outputs of different algorithms and compensates for their prediction errors (Rokach 2016; Kardani et al. 2021; Chatterjee and Byun 2022). The stacking ensemble is divided into basic classifiers, constructed through multiple ML algorithms, and the meta classifier, used to combine the basic classifiers. The core process of the stacking ensemble consists of acquiring the training data for the meta-classifier based on the K-fold cross-validation technique. The results of the basic classifiers become individual sets of training data, and the user determines the weights for the K-fold in the stacking ensemble process, ultimately retraining through the meta classifier.

Performance evaluation

We used the area under the curve (AUC) of the receiver operating characteristics (ROC) curve to evaluate the performance of the landslide susceptibility model. The ROC curve represents the model's performance at all threshold values for binary classification, and AUC represents the degree or measure of separability. A higher AUC in the binary classification corresponds to a better prediction performance of landslide susceptibility. Table 3 shows the definitions of the four measurement classes (true positive (TP), true negative (TN), false positive (FP), false negative (FN)) used in the confusion matrix.

We also calculated several secondary statistics for evaluation: recall, false positive rate (FPR), accuracy, precision, F1-score, Kappa, and Matthews correlation coefficient (MCC). Recall denotes the ratio of correctly predicted positive observations to actual positives. FPR is the ratio of incorrectly predicted positive observations to total actual negatives, indicating the probability that a false alarm is raised while the actual condition is negative. Accuracy is the ratio of correctly predicted observations to total observations, offering an intuitive measure of overall performance but potentially misleading in imbalanced datasets. Precision is the ratio of correctly predicted positive observations to total predicted positives, emphasizing the accuracy of positive predictions, which is important when the cost of false positives is high. The F1-score is the harmonic mean of precision and recall, balancing the trade-off between the two and providing a single metric for uneven class distributions. Kappa is a statistic that compares an observed accuracy with an expected accuracy (defined as random chance), offering a measure of agreement or consistency in classification tasks beyond chance. In the calculation, P(e) represents the hypothetical probability of chance agreement between the observed ratings and the predicted classifications. MCC is a correlation coefficient between the observed and predicted binary classifications, taking all four confusion matrix categories into account and providing a balanced metric suitable for imbalanced datasets. The calculations for deriving these metrics are shown in Eqs. 16-22.

True Positive Rate (Recall) =
$$\frac{TP}{TP + FN}$$
 (16)



Fig. 3 Pre-processed images for the landslide susceptibility model parameters

Table 2 Description of LSA input parameters

Symbol	Model	Description	References
ADA	Adaptive boosting	Boosting basic algorithm initially generates a weak learner and weights it sequentially	Freund and Schapire (1997)
CatBoost	Categorical boosting	Combines boosting with 'target encoding' to improve categorical data pro- cessing performance that was vulnerable in traditional boosting algorithms	Prokhorenkova et al. (2018)
DT	Decision tree	Classifies variables into nodes based on classification criteria and recursively classifies the process	
Dummy	Dummy classifier	A simple comparison between training and prediction data. Builds a baseline for model performance comparison, but not for actual prediction	
ET	Extremely randomized tree	Uses the entire data without a bagging process in the structure of RF and randomly generates node branches	Geurts et al. (2006)
GBC	Gradient boosting	Friedman et al. (2001)	
KNN	K nearest neighbors	Determined by a majority vote of the nearest k data of the target	
LDA	Linear discriminant analysis	Assumes that all classes share the same covariance matrix, i.e., have a linear structure, by applying a Bayes rule that maximizes the probability that a given data belongs to each class	
lightGBM	Light gradient boosting	While other GBC algorithms apply tree depth minimization through the level- wise method, only certain trees are developed through the leafwise method to minimize loss and shorten the time	Ke et al. (2017)
NB	Naive Bayes	Classified on a Bayes basis with the simple assumption that all characteristics are independent of each other	Lewis (1998)
QDA	Quadratic discriminant analysis	Unlike LDA, assumes that each class is a different covariance matrix. There- fore, the crystal boundary is in the form of a quadratic curve	
RF	Random forest	Samples data through the bootstrap process to perform prediction and aggregation with multiple decision trees; allows to measure the impor- tance of each data	Breiman (2001)
Ridge	Ridge classifier	Performed based on linear regression methods, but adds a normalization process called 'L2 regularization' to avoid overfitting	
SVM	Support vector machine	Distinguished by hyperplane between the two data, and regression is also possible based on hyperplane. Basic 'linear model' and 'RBF kernel model considering multidimensional data' are used	Cortes and Vapnik (1995)
XGBoost	Extreme gradient boosting	Improves performance through normalization, pruning, and missing value processing in traditional GBC	Chen et al. (2016)

 Table 3
 Performance measurement classes used in confusion matrix

	Predicted: No	Predicted: Yes
Training: No	True negative (TN)	False positive (FP)
Training: Yes	False negative (FN)	True positive (TP)

False Positive Rate
$$= \frac{FP}{FP + TN}$$
 (17)

Accuracy =
$$\frac{TP + TN}{TP + TN + FP + FN}$$
 (18)

$$Precision = \frac{TP}{TP + FP}$$
(19)

$$F1score = 2 \cdot \frac{Recall \cdot Precision}{Recall + Precision}$$
(20)

Kappa =
$$\frac{Accuracy - P(e)}{1 - P(e)}$$

$$P(e) = \frac{(TP + FP) \cdot (TP + FN) \cdot (FN + TN) \cdot (FP + TN)}{TP + TN + FP + FN}$$
(21)

$$MCC = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP + FP) \cdot (TP + FN) \cdot (TN + FP) \cdot (TN + FN)}}$$
(22)

The ROC is calculated using Eq. (16) for the true positive rate (TPR) and Eq. (17) for FPR (Bradley 1997). The ROC curve plots all threshold values, using FPR on the x-axis and TPR on the y-axis. The closer the shape of the curve approaches to the upper-left corner, the better the performance of the model.
 Table 4
 Specifications
 of
 implementation
 and
 performance

 evaluation
 environment
 environment

Name	Specification
CPU	AMD Ryzen 7 2700X 8-core, 3.7 GHz
GPU	NVIDIA GeForce GTX 1660 4 GB
RAM	64 GB
OS	Windows 10 Education 64-bit
Language	Python 3.7.11

The details of the computing environment used for implementation and performance evaluation are shown in Table 4. Based on the performance comparison among the 16 ML algorithms, the basic classifiers in this study were constructed using the four algorithms with the highest AUC. Of these four algorithms, the one with the best performance was used in the meta classifier.

We used SHapley Additive exPlanations (SHAP) values to explain the effect of the input variables of LSA. SHAP values are derived from the concept of Shapley values in cooperative game theory, aiming to explain the output of machine learning models by assigning an importance value to each feature for a particular prediction. These values are calculated by considering all possible combinations of features, determining the marginal contribution of each feature to the prediction, and averaging these contributions to provide a fair and consistent attribution. This approach offers insights into model predictions and ensures that the contribution of each feature is accurately and fairly represented, making SHAP a powerful tool for interpreting and understanding complex models in the realm of explainable AI. In machine learning, SHAP values are calculated using the following formula:

$$\varphi_i(\nu) = \sum S \subseteq N \setminus \{i\} \frac{|S|!(N-|S|!-1)!}{N!} [\nu(S \cup \{i\}) - \nu(S)]$$
(23)

where $\varphi_i(\nu)$ is an integrated measure of the SHAP value for the *i*th individual variable, and *S* is a combination of variables other than the *i* th variable, constituting a set to represent the effect of *i*. *N* is the total number of variables included in the model. $\nu(S \cup \{i\}) - \nu(S)$ is a predictive function of the machine learning model. The difference between the inclusion and absence sets of *i* is calculated to represent the marginal contribution of *i*. $\frac{|S|!(N-|S|!-1)!}{N!}$ is a weighting factor that gives weight to each set of variables. In essence, the SHAP value for variables in machine learning represents the average impact of including that feature on the prediction compared to if it were not included, and weighted and summed over all possible combinations of other variables. This quantifies the contribution of the variables while considering all interactions with other features, which renders SHAP values a powerful tool for interpretability in complex models.

Results

Table 5 shows the susceptibility model's performance scores, considering all landslide events regardless of the impact size. The AUC performance was highest for Cat-Boost at 0.891, followed by lightGBM (0.886), XGBoost (0.885), and RF (0.885). CatBoost also showed the highest scores in performance metrics other than AUC, including accuracy, F1-score, Kappa, and MCC. However, Catboost's Recall was about 0.009 lower than that of lightGBM and its Precision was about 0.002 lower than that of ET. Overall, gradient boosting algorithms showed high AUC and excellent performance, with AUC values of at least 0.9. AUC was 0.857 for GBC and 0.8278 for ADA, showing that specifically those boosting algorithms incorporating gradients had superior performance. The AUC of SVM and Ridge was 0 due to a limitation of using Pycaret, which returns a zero value if the algorithm does not support the tool 'predict proba' for calculating prediction probabilities. However, SVM and Ridge yielded accuracies of 0.629 and 0.713, respectively, suggesting a relatively poorer performance.

Table 6 shows the susceptibility model's performance scores, considering only landslides of relatively large impact scale (impact size ≥ 1 ha). CatBoost showed the highest AUC at 0.89, followed by XGBoost (0.889), Extra Tree (0.887), and lightGBM (0.883); this constituted a slight decline compared to the model performance when all landslides were included. The AUC of CatBoost, which showed the highest performance, also was lower by approximately 0.019 compared to that scenario. Overall, algorithms based on gradient boosting, as well as the tree ensemble-based Extra Tree, demonstrated high performance.

Landslide susceptibility maps were created based on the top four models by performance (Fig. 4). The landslide-susceptible areas, based on all landslide events, primarily appeared around Gyeonggi (A in Fig. 1), northern Chungcheong (C in Fig. 1), Jeolla (D in Fig. 1), and the southeastern coast of Gyeongsang (E in Fig. 1).

Landslide-susceptible areas primarily appeared in forested areas, centered around slopes of 10° to 50°. Mountain peaks were classified as non-susceptible areas; however, as the slopes became gentler, susceptibility gradually developed. The central regions of Gangwon (B in Fig. 1) and Gyeongsang mostly appeared non-susceptible. The distribution pattern of landslide-susceptible areas showed similar characteristics to the spatial

Model	Accuracy	AUC	Recall	Precision	F1	Kappa	MCC	LogLoss	TT(sec)
CatBoost	0.806	0.891	<u>0.813</u>	<u>0.802</u>	0.807	0.612	0.612	0.420	2.906
lightGBM	0.801	<u>0.886</u>	0.807	0.798	<u>0.802</u>	0.602	0.602	0.427	1.506
XGBoost	0.805	0.885	0.814	0.800	0.807	0.610	0.610	0.459	1.554
RF	0.799	<u>0.885</u>	0.802	0.797	<u>0.799</u>	0.598	0.598	0.452	1.360
ET	0.798	0.879	0.788	0.804	0.796	0.595	0.596	0.461	1.310
GBC	0.773	0.858	0.793	0.764	0.778	0.546	0.547	0.474	1.682
ADA	0.752	0.828	0.770	0.745	0.757	0.505	0.506	0.678	1.344
QDA	0.735	0.807	0.791	0.712	0.749	0.470	0.473	0.957	1.160
LDA	0.713	0.776	0.748	0.700	0.723	0.427	0.428	0.570	1.182
LR	0.699	0.763	0.743	0.683	0.711	0.397	0.399	0.584	2.542
NB	0.664	0.729	0.762	0.637	0.694	0.328	0.334	1.120	1.756
DT	0.715	0.715	0.729	0.709	0.719	0.429	0.429	<u>9.857</u>	1.200
KNN	0.639	0.683	0.681	0.628	0.653	0.277	0.278	1.854	1.832
Dummy	0.500	0.500	0.600	0.300	0.400	0.000	0.000	0.693	1.196
SVM	0.629	0.000	0.656	0.635	0.587	0.258	0.283	0.000	1.158
Ridge	0.713	0.000	0.748	0.699	0.723	0.426	0.427	0.000	1.178

Table 5 Performance scores for all landslide events, regardless of the impact scale

0.000: best performance, 0.000: 2nd best performance, 0.000: 3rd best performance

distribution of landslides that occurred in the country over the past 10 years.

Landslide-susceptible areas based on large landslide events were mainly concentrated in Gyeonggi and Jeolla. Due to their smaller impact areas, those areas present in the susceptibility maps for all landslide events on the east coast and southeastern coast of Gyeongsang were mostly classified as non-susceptible in the map of large landslide events, with the exception of a small proportion.

A common feature of all maps was that the boostingbased algorithms clearly distinguished between landslidesusceptible and non-susceptible areas without creating any ambiguous regions. This is believed to be due to the classification of all ambiguous parts that could degrade performance by dichotomy during the learning process of boosting algorithms. In contrast, tree-based algorithms, such as RF or Extra Trees (Fig. 4), also generated ambiguous areas, possibly due to the inherent randomness of the tree structures. The SHAP value and ROC curves for the classifiers that showed the highest performance among the basic classifiers for constructing the landslide susceptibility model are shown in Figs. 5 and 6, respectively. Magnitude of SHAP values is shown by color (blue to red for low to high) and input values increase from left to right.

For both all and large landslide events types, distance from road and RX1day were the most influential variables. Blue features for distance from the road are clustered around SHAP values of 0.5 and above, while red features are clustered around SHAP values of -1.0 and below. This indicates a correlation between increasing distance from roads and lower landslide susceptibility. Blue features for RX1day are clustered around SHAP values of -0.5 or less, while red features are clustered around SHAP values of 0.5 or more. This indicates a positive correlation with landslide susceptibility.

Soil depth, SRTM, River, and FRTP were most influential for all landslide events, while SRTM, soil depth, FRTP, and RN-aug were most influential for large events.

Model	Accuracy	AUC	Recall	Precision	F1	Kappa	MCC	LogLoss	TT (sec)
CatBoost	0.821	0.890	0.804	0.834	<u>0.818</u>	0.642	0.643	0.420	2.052
XGBoost	0.828	$\underline{0.889}^{\dot{0}\dot{0}\dot{0}\dot{0}\dot{0}\dot{0}\dot{0}\dot{0}\dot{0}0$	<u>0.812</u>	0.840	0.825	0.656	0.657	0.514	0.710
ET	0.823	<u>0.887</u>	0.819	0.827	0.822	<u>0.646</u>	0.647	<u>0.448</u>	0.760
lightGBM	0.809	0.883	0.800	0.816	0.807	0.618	0.619	0.508	0.844
RF	0.816	0.882	0.790	0.836	0.811	0.632	0.634	0.450	0.780
GBC	0.818	0.879	<u>0.816</u>	0.821	0.818	0.637	0.637	<u>0.441</u>	0.752
ADA	0.757	0.838	0.761	0.757	0.759	0.515	0.516	0.662	0.676
QDA	0.767	0.836	0.809	0.747	0.777	0.534	0.537	1.277	0.638
LDA	0.753	0.823	0.766	0.749	0.756	0.506	0.508	0.527	0.640
LR	0.730	0.805	0.718	0.738	0.727	0.460	0.462	0.548	2.012
NB	0.744	0.796	0.768	0.735	0.751	0.489	0.490	1.216	1.212
DT	0.718	0.718	0.723	0.721	0.720	0.436	0.439	<u>9.735</u>	0.648
KNN	0.640	0.678	0.735	0.620	0.672	0.281	0.287	1.931	1.242
Dummy	0.498	0.500	0.600	0.299	0.400	0.000	0.000	0.693	0.634
SVM	0.571	0.000	0.412	0.631	0.413	0.140	0.180	0.000	0.630
Ridge	0.756	0.000	0.762	0.756	0.757	0.513	0.515	0.000	0.642

Table 6 Performance score for landslides with impact scale greater than 1 ha

0.000: best performance, <u>0.000</u>: 2nd best performance, <u>0.000</u>: 3rd best performance

Deeper soil depth corresponded to greater landslide susceptibility. Distance from rivers functioned similarly to distance from roads, with landslide susceptibility decreasing at increasing distance. Greater FRTP corresponded to greater landslide susceptibility, but the majority of values are concentrated in the center of the graph, indicating no substantial impact in either direction. The blue features for RN-aug were mainly clustered around SHAP values below 0, while the red features were mainly clustered around SHAP values above 0. This indicates a positive correlation with landslide susceptibility. However, RN-aug had mixture of SHAP values around 0, indicating that the relationship may be ambiguous. This phenomenon of ambiguous effect on landslide susceptibility was also observed for input variables such as RN-sep, RN-july, RX5day, profile curvature, RD95P, R80, and NDVI. For SDII, fault, HEIGHT, and DNST, greater values corresponded to lower landslide susceptibility, while the opposite applied to slope. Figure 7 shows the final landslide susceptibility map produced through the meta-classifier's ensemble. In both assessments, CatBoost showed the highest AUC and was chosen as the algorithm for the meta-classifier. The ensemble results of the susceptibility model for all landslide events showed a tendency to mitigate the assessment of areas previously determined as risky.

In some western and northern regions of Gyeonggi, all areas, including mountain peaks and foothills, appeared to be susceptible. However, in the ensemble susceptibility map, areas corresponding to mountain peaks and ridges were mostly excluded from susceptible areas. Moreover, the areas in Gyeonggi, Jeolla, the east coast, and the southeastern coast of Gyeongsang that are prone to soil collapse due to precipitation, such as foothills, mainly appeared susceptible. The ensemble results for large landslide cases also showed a similar pattern. In the susceptibility map for large landslide events, not only the foothills but also high mountain areas, including mountain peaks and ridges, appeared susceptible, thus failing to fully reflect the characteristics of landslide occurrence points. However, in the ensemble results, the assessment of these parts



Fig. 4 Landslide susceptibility map for the last 10 years (2011–2020) generated with a basic classifier by the top four machine learning algorithms based on AUC. a All landslide events, b large landslide events



Fig. 5 SHAP value of landslide susceptibility model with basic classifier. a All landslide events, b large landslide events



Fig. 6 ROC curve of landslide susceptibility model with a basic classifier. a All landslide events, b large landslides events



Fig. 7 Landslide susceptibility map during the last 10 years (2011–2020) generated with meta classifier. a All landslide events, b large landslide events

was mitigated, and some mountain peaks and ridges appeared as non-susceptible areas.

Discussion

For all landslide events and as well as only large events, gradient boosting showed relatively high performance, and the CatBoost algorithm yielded the highest performance. This may be because CatBoost is not only a gradient-boosting algorithm but is also particularly suited for categorical variables. LSA analyzes which areas are susceptible, targeting landslides that occurred in the past rather than predicting where landslides will occur in the future; therefore, input variables reflecting long-term characteristics are required. Because such data inevitably include many categorical variables, CatBoost appears to have excellent performance in this respect.

A number of recent studies have attempted to improve the performance of landslide susceptibility models using AI techniques. Hu et al. (2020) and Huan et al. (2023) used stacking ensembles but only compared combinations of ML algorithms and only provided limited descriptions of input variables. Li et al. (2022) attempted to maximize landslide susceptibility model performance by using a deep neural network and a stacking ensemble at the same time. However, a limitation of deep neural networks has always been their requirement for a strict variable selection process in the preliminary stage, which suffers from the difficulties of clearly determining the effects of variables when building the model. Nevertheless, in this study, we were able to quantitatively determine the impact of numerous variables on landslide susceptibility using SHAP values.

The locations of landslides that occurred in South Korea over the past 10 years show a clear occurrence pattern. However, SHAP value distribution suggests that different variables were both positively and negatively correlated with susceptibility (Fig. 5). For this reason, while input variables used in the LSA apparently had an effect on landslide susceptibility, it was difficult to determine critical input variables. This applied to all aspects of the assessment and was not limited to specific variables or models. The reasons for this may be twofold. First, the SHAP values alone are not sufficient to explain the impact of the variables. Variables such as distance from roads and RX1day clearly had directional effects (Fig. 5) that sufficiently explained their impact. However, variables such as FRTP showed ambivalent directionality, although they were overall less impactful. Since such categorical variables do not lend themselves to a linear form of outcome, additional analysis techniques should be integrated. This is left as a task for our future research. Second, the limited accuracy of the landslide locations and the spatial resolution of the input variables likely were an issue. Since the landslide occurrence locations were determined from postal addresses, errors of several tens of meters must be expected even if converted to coordinates using a cadastral map. The spatial resolution of the input data used in this study is 30 m; hence, such errors cannot be sufficiently reflected. Overcoming this limitation is a challenge that must be solved to create a high-precision landslide susceptibility map. In future work, we intend to prepare an improved training data construction method to identify the effect of input variables on susceptibility.

Conclusion

This study constructed a model based on ML algorithms to perform a nationwide LSA of South Korea. A total of 30 input variables were constructed from 12 topographical, 9 environmental, and 9 climatic factors in three ranges to construct an effective model. Sixteen ML algorithms were used to quantitatively evaluate performance based on AUC, and an ensemble analysis was performed using four algorithms that showed superior performance. Additionally, a separate model was constructed for landslide events with a large impact area.

The most important element in an LSA is the information about the location of landslide occurrences. The relevant information covering the past 10 years allowed us to acquire sufficient data, however, there were shortcomings in terms of location accuracy. Ultra-high-resolution satellite images are as of recently being produced by various countries, and their numbers are rising. By supporting higher spatial resolution than that of satellite image sources like Landsat-8 (30 m) and Sentinel-2 (10 m), this imagery is expected to enhance the accuracy of landslide occurrence location data. Our future studies will focus on using these high-resolution satellite images in our susceptibility assessment model and to derive information on geoengineering works undertaken to build roads, apartments, parking lots, etc.

Abbreviations

AGCLS	Age class
AHP	Analytical hierarchy process
AUC	Area under the curve
CNN	Convolutional neural network
DCMLS	Diameter class
DEM	Digital elevation model
DNST	Forest density
FPR	False positive rate
FRTP	Forest type
GIS	Geographic information system
HEIGHT	Forest height
LR	Linear regression
LSA	Landslide susceptibility assessment
MCC	Matthews correlation coefficient
MK-PRISM	Modified Korea parameter elevation relationships on independ-
	ent slopes model
ML	Machine learning
NDVI	Normalized difference vegetation index
RBF	Radial basis function
RD95P	Number of days in a year when the daily precipitation was more
	than the upper 95%
RD99P	Number of days in a year when the daily precipitation was more
	than the upper 99%
RF	Random forest
RN	Total precipitation in each month
RN-aug	Total precipitation in August
RN-july	Total precipitation in July
RN-sep	Total precipitation in September
RNN	Recurrent neural networks
ROC	Receiver operating characteristics
RX1day	Daily maximum precipitation
RX5day	5-Day maximum precipitation
R80	Daily rainfall exceeding 80 mm
SHAP	SHapley Additive exPlanations
SII	Simple daily intensity index
SPI	Stream power index
SRTM	Shuttle radar topography mission data
SVM	Support vector machine

IPI Iopographic position in	de>
-----------------------------	-----

- TPR True positive rate
- TRI Terrain ruggedness index
- TWI Topographic wetness index

Author contributions

Conceptualization: SML; SJL; data curation: SML; formal analysis: SML; funding acquisition: SJL; investigation: SML, SLJ; methodology: SLM, SLJ; writing, original draft preparation: SLM, SJL; writing, review, and editing: SML, SJL; supervision: SJL. All authors have read and agreed to the publishable version of the manuscript.

Funding

This study was carried out with the support of the R&D Program for Forest Science Technology (Project No. 2021341A00-2323-CD01) provided by the Korea Forest Service (Korea Forestry Promotion Institute).

Data availability

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Competing interests

The authors declare no competing interests.

Received: 3 October 2023 Accepted: 2 February 2024 Published online: 12 February 2024

References

- Ado M, Amitab K, Maji AK, Jasińska E, Gono R, Leonowicz Z, Jasiński M (2022) Landslide susceptibility mapping using machine learning: a literature survey. Remote Sens 14:3029
- Ageenko A, Hansen LC, Lyng KL, Bodum L, Arsanjani JJ (2022) Landslide susceptibility mapping using machine learning: a Danish case study. ISPRS Int J Geoinf 11(6):324
- Arnone E, Francipane A, Scarbaci A, Puglisi C, Noto LV (2016) Effect of raster resolution and polygon-conversion algorithm on landslide susceptibility mapping. Environ Model Softw 84:467–481
- Bradley AP (1997) The use of the area under the roc curve in the evaluation of machine learning algorithms. Pattern Recognit 30:1145–1159

Breiman L (2001) Random forests. Mach Learn 45:5-32

- Bruzon AG, Arrogante-Funes P, Arrogante-Funes F, Martin-Gonzalez F, Novillo CJ, Fernandez RR, Vazquez-Jimenez R, Alarcon-Paredes A, Alonso-Silverio GA, Cantu-Ramirez CA, Ramos-Bernal RN (2021) Landslide susceptibility assessment using an AutoML framework. Int J Environ Res Public Health 18:10971
- Chacón J, Irigaray C, Fernández T, El Hamdouni R (2006) Engineering geology maps: landslides and geographical information systems. B Eng Geol Environ 65:341–411
- Chae BG, Cho YC, Song YS, Seo YS (2009) Development of an evaluation chart for landslide susceptibility using the AHP analysis method. J Eng Geol 19:99–108
- Chatterjee S, Byun YC (2022) EEG-Based emotion classification using stacking ensemble approach. Sensors 22:8550
- Chen TQ, Guestrin C, Machinery Assoc Comp (2016) XGBoost: a scalable tree boosting system. In: KDD'16: proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining, 785–94 Cortes C, Vapnik V (1995) Support-vector networks. Mach Learn 20:273–297
- Dou J, Yunus AP, Bui DT, Merghadi A, Sahana M, Zhu ZF, Chen CW, Han Z, Pham BT (2020) Improved landslide assessment using support vector machine with bagging, boosting, and stacking ensemble machine learning framework in a mountainous watershed, Japan. Landslides 17:641–658
- Freund Y, Schapire RE (1997) A decision-theoretic generalization of on-line learning and an application to boosting. J Comp Syst Sci 55:119–139
- Friedman JH (2001) Greedy function approximation: a gradient boosting machine. Ann Stat 29:1189–1232

- Geurts P, Ernst D, Wehenkel L (2006) Extremely randomized trees. Mach Learn 63:3–42
- Hakim WL, Rezaie F, Nur AS, Panahi M, Khosravi K, Lee CW, Lee S (2022) Convolutional neural network (CNN) with metaheuristic optimization algorithms for landslide susceptibility mapping in Incheon. South Korea J Environ Manage 305:114367
- Hodasova K, Bednarik M (2021) Effect of using various weighting methods in a process of landslide susceptibility assessment. Nat Hazards 105:481–499
- Hong HY, Shahabi H, Shirzadi A, Chen W, Chapi K, Bin Ahmad B, Roodposhti MS, Hesar AY, Tian YY, Bui DT (2019) Landslide susceptibility assessment at the Wuning area, China: a comparison between multi-criteria decision making, bivariate statistical and machine learning methods. Nat Hazards 96:173–212
- Hu XD, H. Zhang H. B. Mei, D. H. Xiao Y. Y. Li, M. D. Li, (2020) Landslide susceptibility mapping using the stacking ensemble machine learning method in Lushui, Southwest China. Appl Sci 10:21
- Huan YK, Song L, Khan U, Zhang BY (2023) Stacking ensemble of machine learning methods for landslide susceptibility mapping in Zhangjiajie City Hunan Province, China. Environ Earth Sci 82(1):35
- Ji JJ, Y. Z. Zhou Q. M. Cheng, S. J. Jiang S. T. Liu, (2023) Landslide susceptibility mapping based on deep learning algorithms using information value analysis optimization. Land 12:22
- Joseph VR (2022) Optimal ratio for data splitting. Stat Anal Data Min 15:531–538
- Kadavi PR, Lee CW, Lee S (2019) Landslide-susceptibility mapping in Gangwon-do, South Korea, using logistic regression and decision tree models. Environ Earth Sci 78:1–17
- Kanungo DP, Arora Shantanu Sarkar MK, Gupta RP (2012) Landslide susceptibility zonation (LSZ) mapping–a review. J South Asian Disaster Stud 2(1):81–105
- Kardani N, Zhou AZ, Nazem M, Shen SL (2021) Improved prediction of slope stability using a hybrid stacking ensemble method based on finite element analysis and field data. J Rock Mech Geotech Eng 13:188–201
- Kayastha PM, Dhital R, De Smedt F (2013) Application of the analytical hierarchy process (AHP) for landslide susceptibility mapping: a case study from the Tinau watershed, west Nepal. Comput Geosci 52:398–408
- Ke GL, Meng Q, Finley T, Wang TF, Chen W, Ma WD, Ye QW, Liu TY (2017) LightGBM: a highly efficient gradient boosting decision tree. Adv Neural Inf Process Syst 30.
- Kim WY, Chae BG (2009) Characteristics of rainfall, geology and failure geometry of the landslide areas on natural terrains, Korea. J Eng Geol 19:331–344
- Kim MI, Lee SJ (2023) An analysis of landslide risk using the change in the volumetric water content gradient in the soil layer per unit time of effective cumulative rainfall. Water 15:1699
- Kim KS, Kim WY, Chae BG, Cho YC (2000) Engineering geologic characteristics of landslide induced by rainfall - Boeun. Chungcheong Buk-Do J Eng Geol 10(2):163–174
- Kim MK, Han MS, Jang DH, Baek SG, Lee WS, Kim YH, Kim S (2012) Production technique of observation grid data of 1 km resolution. J Clim Res 7:55–68
- Kim M-I, Lee S, Kim BS (2017) Slope stability assessment induced by variation in mountain topography and rainfall infiltration. J Eng Geol 27(2):125–132
- Kim YT, Park MH, Kwon HH (2020) Spatio-temporal summer rainfall pattern in 2020 from a rainfall frequency perspective. J Korean Soc Disaster Secur 13:93–104
- Lee MS, Kim MI (2016) Investigation and management of landslide areas. J Disaster Prev 18:34–42
- Lee S, Hong SM, Jung HS (2017) A support vector machine for landslide susceptibility mapping in Gangwon Province. Korea Sustainability 9:48
- Lewis DD (1998) Naive (Bayes) at forty: the independence assumption in information retrieval. European conference on machine learning. Springer, Berlin Heidelberg, pp 4–15
- Li WJ, Fang ZC, Wang Y (2022) Stacking ensemble of deep learning methods for landslide susceptibility mapping in the Three Gorges Reservoir area, China. Stoch Env Res Risk A 36:2207–2228
- Lin QG, Lima P, Steger S, Glade T, Jiang T, Zhang JH, Liu TX, Wang Y (2021) National-scale data-driven rainfall induced landslide susceptibility mapping for China by accounting for incomplete landslide data. Geosci Front 12:101248
- Loche M, Alvioli M, Marchesini I, Bakka H, Lombardo L (2022) Landslide susceptibility maps of Italy: lesson learnt from dealing with multiple landslide

types and the uneven spatial distribution of the national inventory. Earth Sci Rev 232:104125

- Pradhan AMS, Kim YT (2014) Relative effect method of landslide susceptibility zonation in weathered granite soil: a case study in Deokjeok-ri Creek, South Korea. Nat Hazards 72:1189–1217
- Prokhorenkova L, Gusev G, Vorobev A, Dorogush AV, Gulin A (2018) CatBoost: unbiased boosting with categorical features. Adv Neural Inf Process Syst 31 (NIPS 2018)
- Reichenbach P, Rossi M, Malamud BD, Mihir M, Guzzetti F (2018) A review of statistically-based landslide susceptibility models. Earth Sci Rev 180:60–91
- Rokach L (2016) Decision forest: twenty years of research. Inf Fusion 27:111–125
- Rossi M, Guzzetti F, Reichenbach P, Mondini AC, Peruccacci S (2010) Optimal landslide susceptibility zonation based on multiple forecasts. Geomorphology 114:129–142
- Rotigliano E, Agnesi V, Cappadonia C, Conoscenti C (2011) The role of the diagnostic areas in the assessment of landslide susceptibility models: a test in the sicilian chain. Nat Hazards 58:981–999
- Sameen MI, Pradhan B, Lee S (2020) Application of convolutional neural networks featuring Bayesian optimization for landslide susceptibility assessment. CATENA 186:13
- Sobie SR (2020) Future changes in precipitation-caused landslide frequency in British Columbia. Clim Change 162:465–484
- Vasu NN, Lee SR (2016) A hybrid feature selection algorithm integrating an extreme learning machine for landslide susceptibility modeling of Mt. Woomyeon. South Korea Geomorphol 263:50–70
- Wang HB, Liu GJ, Xu WY, Wang GH (2005) GIS-based landslide hazard assessment: an overview. Prog Phys Geogr Earth Enviro 29:548–567
- Wang G, Hao JX, Ma JA, Jiang HB (2011) A comparative assessment of ensemble learning for credit scoring. Expert Syst Appl 38:223–230
- Wang Y, Fang ZČ, Wang M, Peng L, Hong HY (2020) Comparative study of landslide susceptibility mapping with different recurrent neural networks. Comput Geosci 138:18
- Wang D, Hao MM, Chen S, Meng Z, Jiang D, Ding FY (2021) Assessment of landslide susceptibility and risk factors in China. Nat Hazards 108:3045–3059
- Wolpert DH (1992) Stacked generalization. Neural Netw 5:241-259
- Yang IT, Chun KS, Park JH (2006) The effect of landslide factor and determination of landslide vulnerable area using GIS and AHP. J Korean Soc Geospat Inf Sci 14:3–12
- Zhang T, Li Y, Wang T, Wang H, Chen T, Sun Z, Luo D, Li C, Han L (2022) Evaluation of different machine learning models and novel deep learningbased algorithm for landslide susceptibility mapping. Geosci Lett 9:16

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.