Moon *et al. Geoenvironmental Disasters* (2024) 11:3
https://doi.org/10.1186/s40677-024-00267-8

Geoenvironmental Disasters

RESEARCH

Open Access

Comparison of factors influencing landslide risk near a forest road in Chungju-si, South Korea

Seong-Woo Moon[1], Jeongdu Noh[2], Hyeong-Sin Kim[3], Seong-Seung Kang[4] and Yong-Seok Seo[1*]

Abstract

Background The study aimed to identify the influential factors required to prepare landslide vulnerability maps and establish disaster prevention measures for mountainous areas with forest roads. The target area is Sancheok-myeon, Chungju-si, where several landslides have occurred in a narrow area of approximately 3 km × 4 km. As the area has the same rainfall and vegetation conditions, the influences of the physico-mechanical characteristics of the soil in accordance with compaction and topographic characteristics could be analyzed precisely.

Methods Geological surveying, sampling, and laboratory testing assessed landslide risk in the study area, and data including unit weight, specific gravity, porosity, water content, soil depth, friction angle, cohesion, slope angle, profile/plan curvature, TWI were obtained. Preprocessing and screening such as min-max normalization and multicollinearity were conducted for the data in order to eliminate overestimation of each factor's effectiveness. The influence of each factor was analyzed using logistic regression (LR), structural equation modeling (SEM), extreme gradient boosting (XGBoost), and light gradient boosting machine (LightGBM).

Results All methods showed that soil depth has the greatest impact on landslide occurrence. Friction angle, slope angle, and porosity were also selected as influential factors, although each method ranked them slightly differently. Topographic factors, such as plan curvature, profile curvature, and the topographic wetness index, had minimal influence. This appears to be because landslides near forest roads are more affected by how well compaction was performed during banking than by the concave or convex shape of the slope. This study presents analysis results for an area with the same rainfall and vegetation conditions; therefore, the analysis of the influence of the physico-mechanical characteristics of the soil and topography was more precise than when comparing landslides occurring in different regions. Our results may be helpful in preparing landslide vulnerability maps.

Keywords Landslide influential factor, Logistic regression analysis, Structural equation model, XGBoost, LightGBM

*Correspondence:
Yong-Seok Seo
ysseo@cbu.ac.kr
Full list of author information is available at the end of the article

Introduction

Landslides constitute one of the most dangerous types of natural disaster. They are known to have caused 838 annual deaths globally between 2002 and 2021 (CRED 2023). The precise mapping of landslide susceptibility and methods to assess landslide risk to decrease their potential damage have received substantial research attention. However, predicting landslide occurrence remains difficult despite sustained research efforts, because it is affected by complex interactions among many factors, including geological conditions, geomorphology, climate, earthquakes, and vegetation (Gerrard and Gardner 2002; Wobus et al. 2003; Hasegawa et al. 2009). The main factors influencing landslide occurrence and the relationships among them remain unclear without rainfall factor, thus hindering precise landslide prediction (John and Douglas 2012).

With reference to analysis methods related to landslide, statistical methods including conditional probability, weight of evidence, frequency ratio (FR), and logistic regression (LR) were typically used in the 1990s and 2000s to analyze the influences of factors causing landslides and to predict landslides. Machine learning methods, such as artificial neural networks and deep learning, have been used since the 2010s. For example, EKER and Aydin (2014) prepared a landslide vulnerability map in an analysis of landslide vulnerability for different road types (e.g., forest roads and expressways) by conducting geographic-information-system-based LR analysis of land use, petrology, elevation, slope, side, distance to rivers, distance to roads, and plan curvature. Pham et al. (2016) assessed landslide vulnerability in 930 landslide areas by analyzing Google images using support vector machine, LR, Fisher's linear discriminant analysis, Bayesian network, and naïve Bayes techniques. Wang et al. (2016) proposed a landslide prediction model using LR, FR, decision tree, weight of evidence, and artificial neural networks. Chen et al. (2018) proposed a landslide vulnerability model using a random forest (RF) algorithm based on a digital elevation model and Landsat-8 data. Xiao et al. (2020) proposed a landslide vulnerability model using hybrid models combining RF, FR, CF (certainty factor), and the index of entropy (IOE), namely RF-FR, RF-CF, RF-IOE, IOE-CF, and CF-FR. Further methods—such as big data, machine learning, and deep learning methods, which may overcome existing mathematical and engineering limitations—have been actively used in recent years. Representative prediction models include boosting-based models, such as extreme gradient boosting (XGBoost), light gradient boosting machine (LightGBM), and category boost (CatBoost) (Chen and Guestrin 2016; Ke et al. 2017; Prokhorenkova et al. 2017).

In relation to influential factors of landslide, precipitation or rainfall intensity was pointed out as the most influential factors on landslide in research cases using data-driven analysis because rainfall conditions are different with each other due to greater distance between data collection points as numerous landslide cases need to be analyzed (Chae et al. 2004; Quan et al. 2011; Chen et al. 2013). Also, in physically based analysis, the friction angle and cohesion included in the slope stability analysis equation were evaluated as the most dominant factors on landslide (Regmi et al. 2010; Qu et al. 2021). Besides, it was reported that the stability of slope with complete vegetation cover is higher than that of slope with meagre vegetation (Schmidt et al. 2001; Osman and Barakbah 2006), and there were substantial interests in the effects of forest roads hydrologically and geomorphically on earth surface and landforms (Luce and Wemple 2000; Dutton et al. 2005). Vanacker et al. (2005) reported that changes in the forest landscape or large-scale logging, which change the soil infiltration and ground evapotranspiration rates, thus indirectly affect the water contents in soil and reduce slope stability. It was reported that the soils from landslide prone areas were mainly silty soils with low plasticity (Jotisankasa and Vathananukij 2008). Nugraha et al. (2015) argued that land surface (geomorphometric) characteristics have a significant relationship with the landslide distribution, and even others have emphasized the role of investigating topographic influence (Fernandes et al. 2004; Broothaerts et al. 2012). In addition, Owen (1981) said that the sunny aspects were much more susceptible to landslide than the shady aspects. When the soils are saturated, the liquid limit water content of the sunny aspect subsoil is exceeded, while that of the shady aspect subsoil is not. Meanwhile, Kimaro et al. (2000) suggested that the most important soil characteristics is presence of saprolite or boundary with hard bed rock. As mentioned above, the most influential factors are differently evaluated depending on researcher's perspectives because various factors including vegetation, climate, geology, topography and so on, affect landslides.

Throughout Korea's many mountainous areas, several forest roads have been constructed for forest management. Construction of these roads involves the formation of cutting and banking slopes, which affect slope stability by changing the ground, topography, and water flow (Wempleet al. 1996; Choi et al. 2011). In addition, a recent increase in guerrilla rainstorms caused by climate change has increased landslide risk. According to the Korea Forest Service, the frequency of landslides is increasing annually, and the size of a landslide depends on the region and season, with typhoons and heavy rainfall being concentrated in summer (Korea Forest Service

2021). Therefore, previous regional landslide analyses have focused primarily on rainfall, which is an external factor; therefore, detailed risk plan considering internal factors has not been facilitated when activities such as the selection of areas for reinforcement and the establishment of disaster prevention measures are conducted. The present study analyzes the effects of soil, topographic factors, and rainfall on landslides using statistical and machine learning methods to identify major influencing factors. The results may aid in the preparation of landslide vulnerability maps and establish disaster prevention measures (e.g., prioritizing areas for reinforcement) within budgetary constraints.

Methodology

The theory of logistic regression analysis

Logistic regression (LR) analysis determines correlations between a dependent variable and multiple independent variables influencing it. The probability of an event can be calculated and expressed as a value between 0 and 1. Values can be binarized by those ≥ 0.5 being assigned as 1, and those < 0.5 being assigned as 0. The probability of an event through LR analysis (P_Z) is given by Eqs. (1) and (2):

$$P_Z = \frac{1}{1 + e^{-Z}} \tag{1}$$

$$Z = \alpha + \beta_1 X_1 + \cdots + \beta_n X_n \tag{2}$$

where Z is the LR, α is a constant, and β_n is the regression coefficient of the independent variable (X_n).

The Nagelkerke R-squared, Hosmer and Lemeshow, and confusion matrix verification methods were used to analyze the reliability of the results of LR analysis. Nagelkerke R-squared indicates the degree to which independent variables can explain the dependent variable. A value of ≥ 20% indicates that the independent variables have explanatory power. The Hosmer and Lemeshow test determines the overall goodness-of-fit of a regression model. A significance level of > 0.05 indicates that the model has explanatory power. The confusion matrix estimates prediction accuracy as the area under the curve (AUC) calculated for an ROC curve with an X-axis of "1-specificity" and a Y-axis of "sensitivity". Values of AUC are distributed between 0 and 1, with a value close to 1 indicating an accurate model (Fawcett 2006; Godt et al. 2008; Šimundić 2009). Accuracy, specificity, and sensitivity can be calculated using Eqs. (3), (4), and (5), respectively:

$$Accurancy = \frac{TP + TN}{TP + TN + FP + FN} \tag{3}$$

$$Specificity = \frac{TN}{(TN + FP)} \tag{4}$$

$$Sensiticity = \frac{TP}{(TP + FN)} \tag{5}$$

where a true positive (TP) is the correct prediction of a positive value, a true negative (TN) is a correct prediction of a negative value, a false positive (FP) is a negative value incorrectly predicted as positive, and a false negative (FN) is a positive value incorrectly predicted as negative.

The theory of structural equation model

The SEM used here was first suggested by Wright (1921). It is a path-analysis-based statistical method used to identify causal relationships among multiple variables with complex interrelationships and in cases with many independent and dependent variables. Although it seems similar to multiple regression analysis, it is a more detailed model because it can consider the mutual influences of all variables and can easily identify interrelationships among variables using graphical representation (Hox and Bechger 1999; Yung 2008; Ullman and Bentler 2013). The SEM can be subdivided into a measurement model for confirmatory factor analysis and a theoretical model for multiple regression and path analyses. The former is applied when each variable can explain latent variables perfectly, and the latter is used to group variables into representative latent variables and to find the most descriptive model. When studying landslides, the latter model is more suitable than the former, as many factors related to slope failure or landslide occurrence are difficult to explain fully, and there are limitations related to ground heterogeneity.

Theoretical modeling in an SEM is calculated using partial least squares (PLS), which are subdivided into partial least squares regression (PLS-R) and partial least squares path models (PLS-PMs). PLS-R is used when there are more variables than the number of data, and PLS-PMs are applied to analyze interrelationships. This study applies a PLS-PM, which can deal with a large amount of data and analyze interrelationships and causal analysis among influential factors related to landslide occurrence. PLS-PM analysis is a multivariate analysis technique that analyzes the systems of relationships among many blocks containing variables. This approach follows the component-based estimation procedure, and it is defined by the following two basic concepts: each block of variables acts as a latent variable, and it is assumed that there is a system of linear relationships between blocks. The analysis

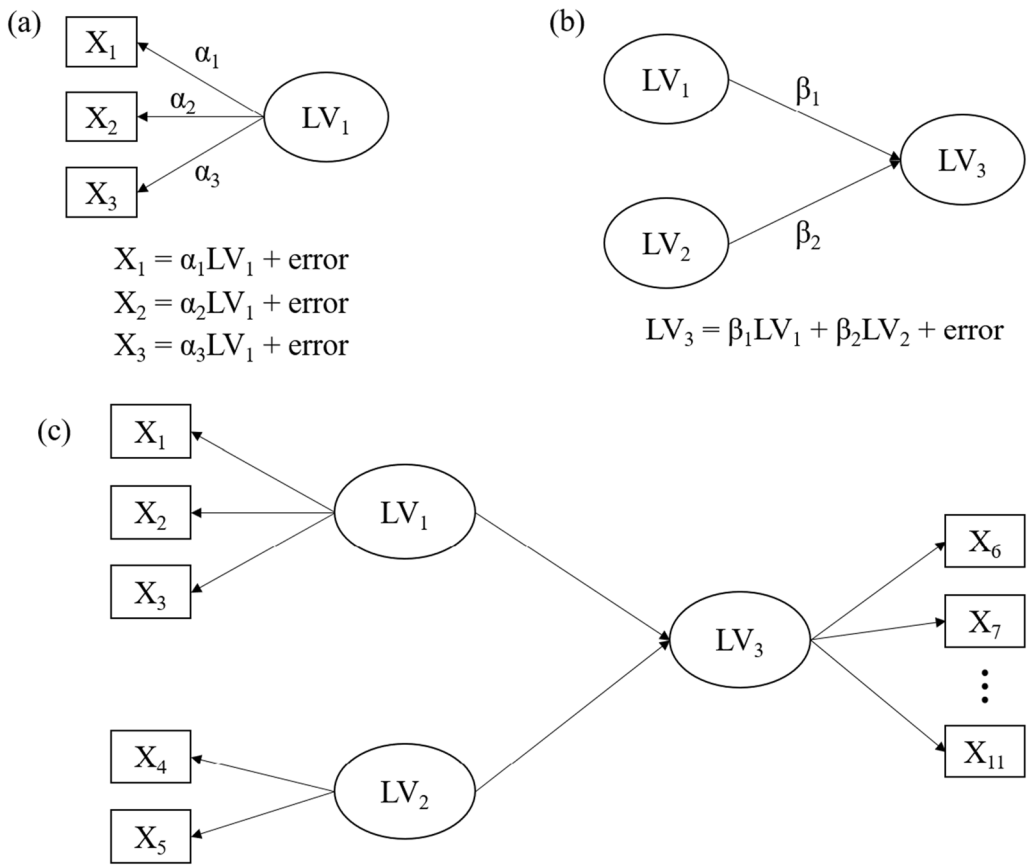

Fig. 1 Schematic diagram of the PLS-PM. **a** The external model comprises manifest variables (dependent and independent variables, X_{ij}) and latent variables. **b** The internal model consists of latent variables (LV_i). **c** The complete PLS-PM includes both the internal and external models

considers multiple relationships between variable blocks, and each variable block is assumed to be represented by a latent variable or theoretical concept. Here, each latent variable is a hypothetical variable created to generate an SEM. The latent variables are grouped with variables that have similar characteristics. The determination and interrelationships of the latent variables are set by the researcher's subjective judgment, and continuous modification and supplementation are required until an optimal model is developed.

Figure 1 outlines the PLS-PM, showing the constituent manifest (dependent and independent variables, X_{ij}) and latent variables. The first step in PLS-PM analysis is to set the latent variables—which comprise manifest variables with similar characteristics (Fig. 1a)—and then the internal model is determined based on the latent variables of the external model (Fig. 1b). The setup of the external and internal models is then modified and updated until statistically significant results are obtained for the arrangement of variables, causal relationships, and error terms. The last step assesses the confidence of the entire model using the external and internal models estimated from the above two steps (Fig. 1c). Here, weight and loading (α and β, respectively; Fig. 1) are essentially correlation coefficients.

The theory of XGBoost

XGBoost uses the classification and regression tree (CART) model for the existing gradient boosting algorithm and enables parallel processing, thereby enabling the resolution of various problems using data mining (Chen and Guestrin 2016; DSBA 2020; Yoon 2020; An 2021). Unlike other tree-based learning methods, the learning of XGBoost uses Eq. (6) based on the CART model. When the data comprise input variable x and output variable y, $\widehat{y}$ is the predicted value of data x, K is the number of CARTs, and f is the CART model (Chen and Guestrin 2016). Equation (7) gives the objective function for training the CART model. Here, $l\left(y_i, \widehat{y}_i\right)$ is the difference between the actual and predicted values, and Ω is the regularization of the model to prevent overfitting. The objective function equation at step t can be expressed using XGBoost's additive method and Taylor expansion, as shown in Eq. (8). According to the definition of the

Taylor expansion, g_i is the first-order derivative of $\widehat{y}_i^{(t-1)}$ and can be defined as $g_i = \delta_{\widehat{y}^{(t-1)}} l\left(y_i, \widehat{y}^{(t-1)}\right)$; h_i is the second-order partial derivative of $\widehat{y}_i^{(t-1)}$ and can be defined as $h_i = \delta^2_{\widehat{y}^{(t-1)}} l\left(y_i, \widehat{y}^{(t-1)}\right)$. The greedy and approximate algorithms are used to optimize the prediction model and identify the optimal split point using above equations (Chen and Guestrin 2016).

$$\widehat{y}_i = \sum\nolimits_{k=1}^{K} f_k(x_i) \tag{6}$$

$$obj(\theta) = \sum\nolimits_{i}^{n} l\left(y_i, \widehat{y}_i\right) + \sum\nolimits_{k=1}^{K} \Omega\left(f_k\right) \tag{7}$$

$$obj(t) = \sum\nolimits_{i=1}^{n} \left[g_i f_t(x_i) + \frac{1}{2} h_i f^2(x_i)\right] + \Omega\left(f_i\right) \tag{8}$$

The theory of LightGBM

LightGBM is a GBM- and tree-based algorithm that performs learning using residuals. However, unlike the symmetric division method of conventional trees, its tree structure is asymmetrical due to its use of a leaf-wise methodology. LightGBM uses a feature histogram that divides continuous variables into discrete sections (bins) during learning. This method learns a function from slope space g to input space X^S using a decision tree. In the presence of the training set of n independent and identically distributed entities $\{x_1, x_2, \cdots x_n\}$, X^S is a vector with a dimension of s. For GBM, the loss function for the model output value generated at each iteration is defined as a negative slope $\left\{g_1, g_2, \cdots g_n\right\}$. This model uses Eq. (9) to divide each node through a variable with the largest information acquisition (Ke et al. 2017):

$$V_{j|O}(d) = \frac{1}{n_O}\left(\frac{\left(\sum_{x_i \in O: x_{ij} \le d} g_j\right)^2}{n^j_{l|O}(d)} + \frac{\left(\sum_{x_i \in O: x_{ij} \le d} g_j\right)^2}{n^j_{r|O}(d)}\right), \tag{9}$$

$$n_O = \sum I[x_i \in O], n^j_{l|O}(d) = \sum I\left[x_i \in O : x_{ij} \le d\right], n^j_{r|O}(d) = \sum I\left[x_i \in O : x_{ij} > d\right]$$

where O is the training data in the tree node, d is the node, and j is the variable performing division at point d. However, this method is inefficient because it searches all divided sections. To prevent this, gradient-based one-side sampling (a method of reducing the number of data) and exclusive feature bundling (a method of reducing the number of variables) are used.

Study area and data collection

Status of landslide occurrence and sampling locations

To analyze the factors influencing landslide occurrence, site investigation and sampling were conducted in Sancheok-myeon, Chungju-si, South Korea, where a number of landslides occurred within a small area (3 km by 4 km) following 259 mm of rainfall on 2 August 2020. The study area is located in the 37° 06′ 02.6″N ~ 37° 08′ 26.5″ N, 127° 57′ 50.3″E ~ 127° 59′ 13.1″ E, and it has a mountainous terrain due to the surrounding mountains, including Ocheong mountain (EL + 656.8 m) of the northeastern part, Cheondeung mountain (EL + 807.1 m) of the eastern part, and Jangbaek mountain (EL + 405.0 m) of the western part. Also, it is reported that Mesozoic granite is mainly distributed, in the study area and this granite contains some of Precambrian gneiss (Kim 2022).

According to precipitation record of the KMA from 2000 to 2023 (KMA 2023), the average annual precipitation of study area is 1,196 mm, and it is similar to that of South Korea (about 1200 mm). So, this region is an area where disaster caused by rainfall are rare. However, annual average precipitation of 1500 mm, a daily precipitation of 316 mm, and a maximum hourly precipitation of 76.5 mm were all record breaking in 2020 when a lot of landslides occurred.

Figure 2 shows sampling locations and photographs of the landslides that occurred in the study area. Sampling locations are colored either red or yellow: the 40 red points are locations at which landslides occurred, and the 45 yellow points indicate sampling locations where no landslides occurred. Locations with or without landslides were sampled in the one point (in case of occurrence, sampling was conducted in the head part) to provide the statistical and machine learning analyses with the necessary data for both landslide and non-landslide locations. Most landslides in the study area occurred near the forest road (Fig. 2) because the slope angle steepens at the cut slope, and the soil thickness increases where the construction of the road involved slope filling. The upper slopes of forest road comprised weathered soil of biotite granite, and the lower slopes of forest road were made up of embanked soil which was excavated when constructing forest road. Therefore, soil type of landslides was all same with weathered soil of granite, and that it was actually composed mostly of sand with SW-SP (well grade sand-poor grade sand). This led to the mineralogical compositions of soil particles being consistent across the sampling locations, as the roads had been constructed

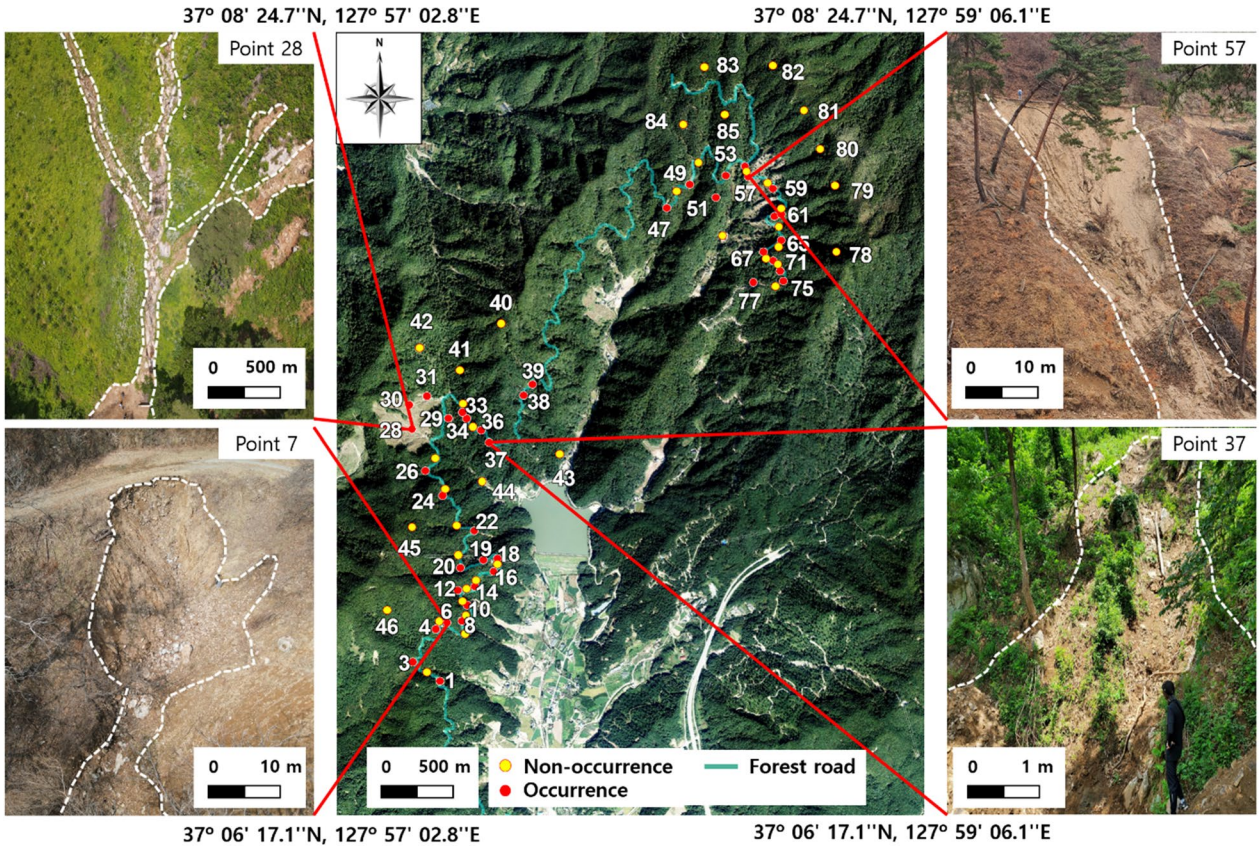

Fig. 2 Locations of sampling points and photographs of landslides along the forest road. Landslides have occurred at 40 of the 85 sampling locations

at the same time. The study area allows specific analysis of the influence of topography and physico-mechanical characteristics associated with soil compaction on landslide occurrence owing to vegetation and rainfall conditions being consistent throughout the area.

The dataset comprises the following information for each site: presence or absence of landslide occurrence (hereafter abbreviated as occurrence or non-occurrence); thickness of the soil layer (hereafter abbreviated as soil depth); slope angle; plan curvature and profile curvature; TWI; dry and saturated unit weights of the soil; and porosity, specific gravity, saturated water content, friction angle, and cohesion of the soil. Sampling was conducted in the head parts of areas where landslides occurred. Elevation was not considered, as the sampling points were at similar altitudes. Data for landslide occurrence and soil depth were obtained from site investigations and dynamic cone penetration testing, and physico-mechanical properties (unit weight, specific gravity, porosity, friction angle, and cohesion) were measured according to the test criteria of the American Society for Testing Materials (ASTM D2216-10; ASTM D2487-17; ASTM D3080-98; ASTM D422-63; ASTM D854-10). Topographic characteristics (slope angle, profile, and plan curvatures) were gained from 1:50,000 digital topographic maps of the National Geographic Information Institute and SAGA GIS software (IBM). The profile and plan curvatures describe whether the slope is concave (negative value) or convex (positive value) longitudinally and in cross-section, respectively (Fig. 3). The TWI is an indicator of the wet content of the soil and is calculated using Eq. (10) (Beven and Kirkby 1979):

$$TWI = ln\frac{SCA}{tan\theta} \tag{10}$$

where SCA denotes the local upslope area draining through a certain point per unit contour length, and θ is the local slope in radians. The SCA is calculated using multiple flow directions, as the flow may vary according to the slope's direction and gradient. Most of the factors are continuous data; the only categorical factor is occurrence (1 for occurrence and 0 for non-occurrence).

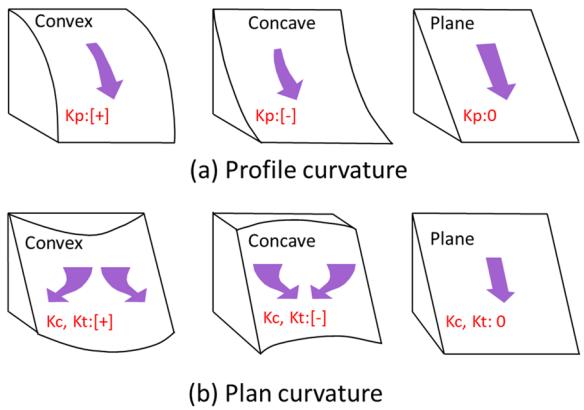

Fig. 3 Schematic diagrams of **a** profile and **b** plan curvature (modified after Dikau 1989)

Distribution of data

The values recorded for the various factors that control landslide occurrence are shown as box-and-whisker plots in Fig. 4. The ends of whiskers indicate the maximum and minimum statistically significant values; any values beyond these ranges were discounted as erroneous outliers. Boxes span the first and third quartiles (the interquartile range); therefore, each box encloses half of the data. The horizontal bar in each box indicates the median, and the plots depict the distribution of data, allowing comparison of the range, interquartile range, and median.

The unit weight shows greater whisker and interquartile ranges for non-occurrence cases than for occurrence cases, regardless of the conditions being dry or saturated (Fig. 4a, b). The ranges of specific gravity are similar for occurrence and non-occurrence cases (Fig. 4c). The interquartile range and median of porosity are higher for occurrence cases than non-occurrence cases; porosity is expected to be proportional to occurrence, as soil can hold much water, increasing its weight and reducing the resistance force (Fig. 4d). The median saturated water content is slightly higher for occurrence cases than for non-occurrence cases (Fig. 4e) and is interpreted similarly to the results of porosity. The interquartile range and median of soil depth are higher for occurrence cases than non-occurrence cases (Fig. 4f). Those of friction angle tend to be lower for occurrence cases than non-occurrence cases, whereas cohesion has the opposite tendency, being positively correlated with occurrence (Fig. 4g, h). In terms of mechanics, cohesion is generally proportional to non-occurrence. However, if the friction angle and cohesion were measured from direct shear tests, they would be inversely proportional to each other according to the Mohr–Coulomb failure criterion (Moon et al. 2020). For this reason, the friction angle is inversely proportional to occurrence, and cohesion is proportional to occurrence. The interquartile range of the slope angle is higher for occurrence than non-occurrence (Fig. 4i). The median profile and plan curvatures are inversely proportional to occurrence, meaning that a number of landslides occurred near valleys with concave topography (Fig. 4j, k). The interquartile range and median of TWI are higher for occurrence than non-occurrence, indicating that soil containing water was prone to landslides (Fig. 4l).

Data preprocessing and screening

Data preprocessing and screening for statistical analysis were performed using min–max normalization and multicollinearity diagnosis. Min–max normalization was performed for the 12 measured independent variables (dry unit weight (kN/m^3), saturated unit weight (kN/m^3), specific gravity, porosity (%), saturated water content (%), friction angle (°), cohesion (kPa), soil depth (m), slope angle (°), profile curvature, plan curvature, and TWI) (Eq. (11)):

$$X_n = \frac{X - X_{min}}{X_{max} - X_{min}} \tag{11}$$

where X_n, X, X_{min}, and X_{max} are the normalized, observed, minimum observed, and maximum observed values, respectively. The normalized results are all between 0 and 1, which facilitates direct comparison of the effects of each dependent variable (despite their initially different distributions and units) on the dependent variable (i.e., landslide occurrence).

Multicollinearity is a phenomenon in which negative effects (such as the overestimation of regression model variables and degraded reliability of regression results) may occur when highly correlated independent variables are used in regression analysis (Ryu 2008). Therefore, collinearity must be assessed before regression analysis. The variation inflation factor (VIF; Eq. (12)) can be used for this. A VIF of ≥ 10 indicates multicollinearity (Kutner et al. 2004).

$$VIF = \frac{1}{1 - R^2} \tag{12}$$

where R^2 is the coefficient of determination. The left side of Table 1 lists the estimated multicollinearity among the 12 independent variables. Those with VIF values of ≥ 10, and thus high correlations corresponding to multicollinearity, are the dry unit weight (γ_d), saturated unit weight (γ_{sat}), specific gravity (G_s), porosity (e), and saturated water content (w). These factors can be related using Eqs. (13) to (15):

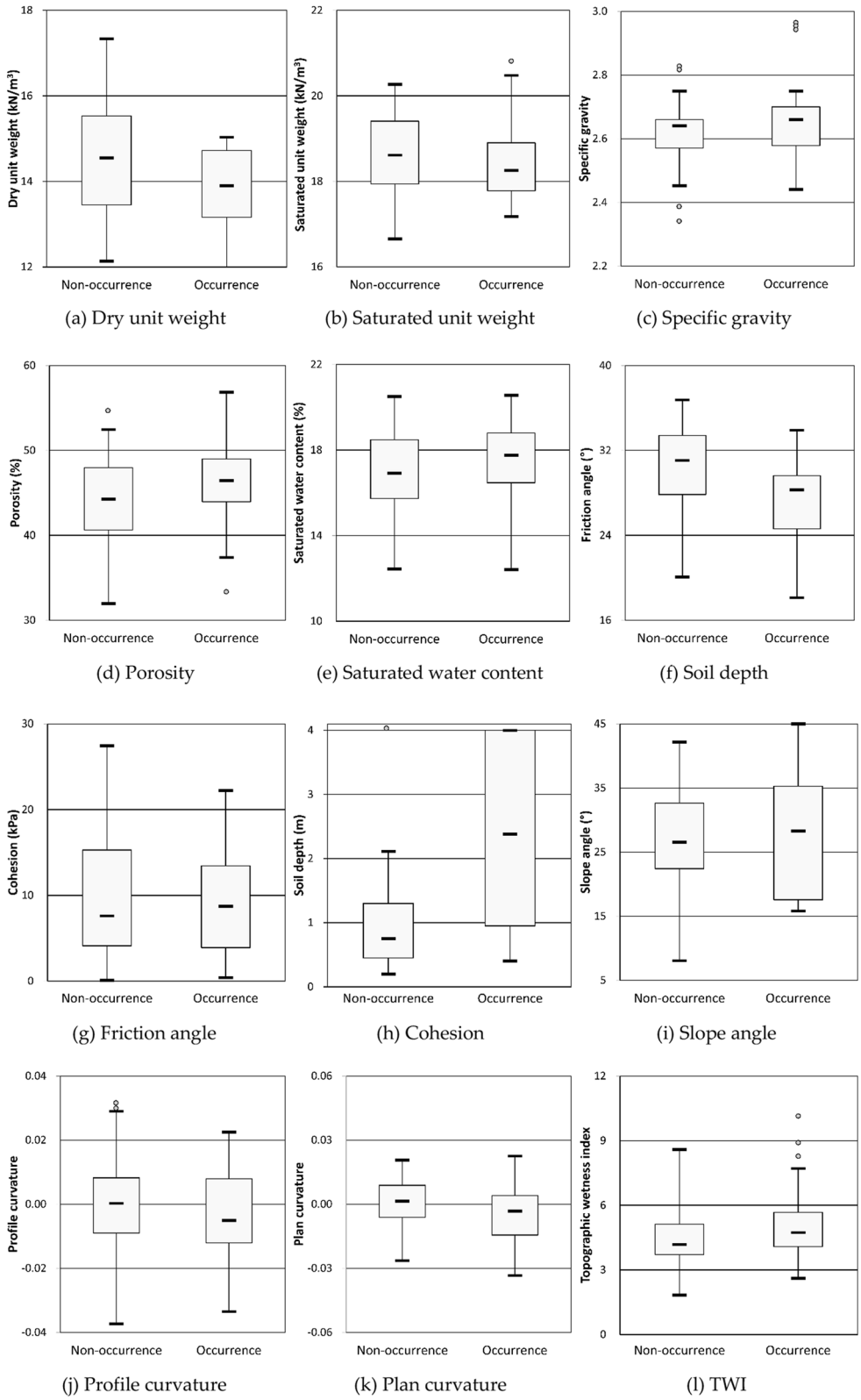

Fig. 4 Box plots showing the values of properties influencing landslides

Table 1 Variation inflation factors (VIFs) for properties influencing landslide susceptibility to assess multicollinearity. Although there are 12 factors in the first step, only nine factors are left after multicollinearity check (three factors are eliminated)

First step		Final step	
Factors	**VIF**	**Factors**	**VIF**
Dry unit weight	7672.21	Saturated unit weight	4.61
Saturated unit weight	4089.78	Porosity	4.66
Specific gravity	473.91	Friction angle	1.36
Porosity	3661.84	Cohesion	1.14
Saturated water content	2050.76	Soil depth	1.25
Friction angle	1.49	Slope angle	1.79
Cohesion	1.16	Profile curvature	1.39
Soil depth	1.29	Plan curvature	1.90
Slope angle	2.05	TWI	2.76
Profile curvature	1.47	–	–
Plan curvature	1.94	–	–
TWI	3.12	–	–

$$\gamma_d = \frac{G_s}{1+e} \tag{13}$$

$$\gamma_{sat} = (1+w)\gamma_d \tag{14}$$

$$G_s \times w = S \times e \tag{15}$$

where S is the degree of saturation.

As collinearity depends on the number of independent variables, factors with high multicollinearity are eliminated step by step so that the VIF of all independent variables could be < 10. Finally, statistical analysis and machine learning are performed using nine independent variables after removing dry unit weight, specific gravity, and saturated water content (the right side of Table 1).

Results of analyses

Logistic regression (LR)

LR analysis uses the nine independent variables filtered through data screening. The AUC of the LR model of 0.776 indicates high prediction ability. In addition, the regression model is determined to be valid, because the Nagelkerke R-squared value, which describes the significance level and reliability of the regression model, is 0.410, and the Hosmer and Lemeshow test significance probability is 0.317. Table 2 shows the regression coefficient of the above regression model and the influence of each independent variable on landside occurrence. Soil depth has the greatest influence (25.76%), followed by porosity and friction angle.

Table 2 Results of logistic regression analysis

Rank	Observed variables	| Coefficient |	Influence (%)
5	Saturated unit weight	1.112	8.37
2	Porosity	2.957	22.27
3	Friction angle	2.099	15.81
8	Cohesion	0.350	2.64
1	Soil depth	3.421	25.76
9	Slope angle	0.161	1.21
7	Profile curvature	0.860	6.48
6	Plan curvature	0.880	6.63
4	TWI	1.440	10.84

Table 3 Results of quantified influence of factors on landslides. The total influence is calculated by multiplying the influence in the external model by that in the internal model

Observed variables	Influence			Rank
	External model	Internal model	Total	
Saturated unit weight	0.870	0.302	0.263	3
Porosity	0.997		0.301	2
Friction angle	8.432	0.019	0.160	5
Cohesion	0.014		0.000	9
Soil depth	0.945	0.579	0.547	1
Slope angle	0.289		0.167	4
Profile curvature	0.018		0.010	8
Plan curvature	0.134		0.078	6
TWI	0.035		0.020	7

Structural equation model (SEM)

Figure 5 and Table 3 show the results of SEM analysis. The entire model system includes the internal and external models. The internal model, comprising physical properties, mechanical properties, topographic properties, and occurrence, is depicted by arrows pointing at occurrence, as each latent variable affects this outcome. The external model is shown by arrows pointing to the independent variables from each latent variable. The label on each arrow gives its statistical weight, which is a measure of how effectively the latent variable can explain the independent or other latent variables. The weight is the same as the effectiveness in Table 3.

According to path model theory, the effectiveness of each factor is the product of the weight in the external model and that in the internal model. For example, the effect of soil depth on occurrence is calculated as 0.945 × 0.579 in the external model and 0.547 in the internal model (Table 3). The most influential factor is soil depth; the next most influential factors in order are porosity, saturated unit weight, and slope angle. The

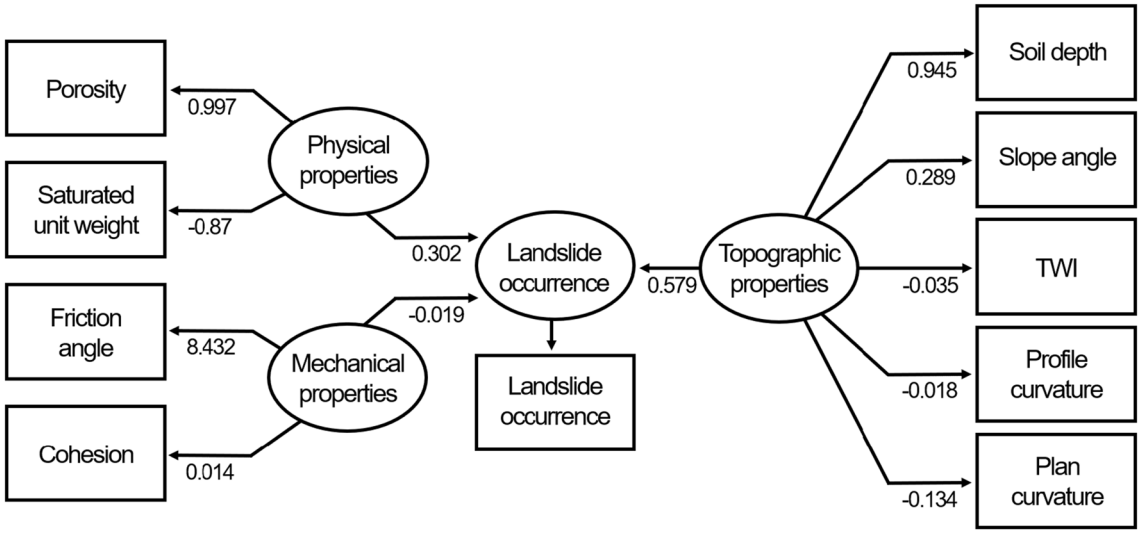

Fig. 5 Results of SEM analysis, in which physical properties, mechanical properties, and topographic characteristics affect landslide occurrence. Numbers near each arrow in the external and internal models indicate absolute values of the statistical weight, which quantifies that factor's effect on occurrence

cohesion, profile curvature, and TWI have little effect on occurrence.

Reliability assessment of the entire model is based on the confidence level using *p*-values and the goodness of fit index (GFI). The statistical criterion evaluating the significance of the results at the 95% confidence level is considered here: $p < 0.05$ indicates satisfying the 95% confidence level. The GFI is calculated from the average communality and/or the geometric mean of the average determination coefficient. The criteria for high and low confidence in the GFI are the same as those used for R^2, as the statistical meaning of GFI is similar to that of the determination coefficient. The criteria are as follows (Zikmund 2000; Moore et al. 2013; Sanchez 2013):

- Low: $R^2 < 0.3$,
- Moderate: $0.3 < R^2 < 0.6$,
- High: $R^2 > 0.6$.

The resulting *p*-value and GFI of the entire model are 0.000 and 0.763, respectively, which means that our results can be considered statistically significant with a "high" confidence grade.

Table 4 Results of hyperparameter optimization using XGBoost

Train and test data ratio	9:1	8:2	7:3	6:4	5:5
learning_rate	0.01	0.01	0.01	0.04	0.02
n_estimators	500	500	500	3	2
max_depth	3	3	3	500	500
gamma	1	2	2	1	1
colsample_bytree	0.5	1	1	0	0
max_delta_step	1	0	0	4	3.0
min_child_weight	1	0.5	0.5	0.5	1.5
reg_alpha	1	0	0	1.5	0
reg_lambda	2.5	2.5	2.5	2.5	0.5
subsample	1	1	1	1	1
scale_pos_weight	1	1	1	2	1

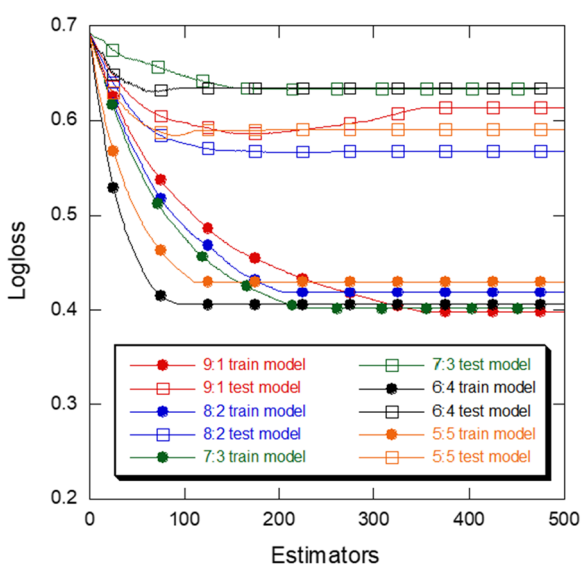

Fig. 6 Results of learning performance for XGBoost prediction models using different ratios of training and test data

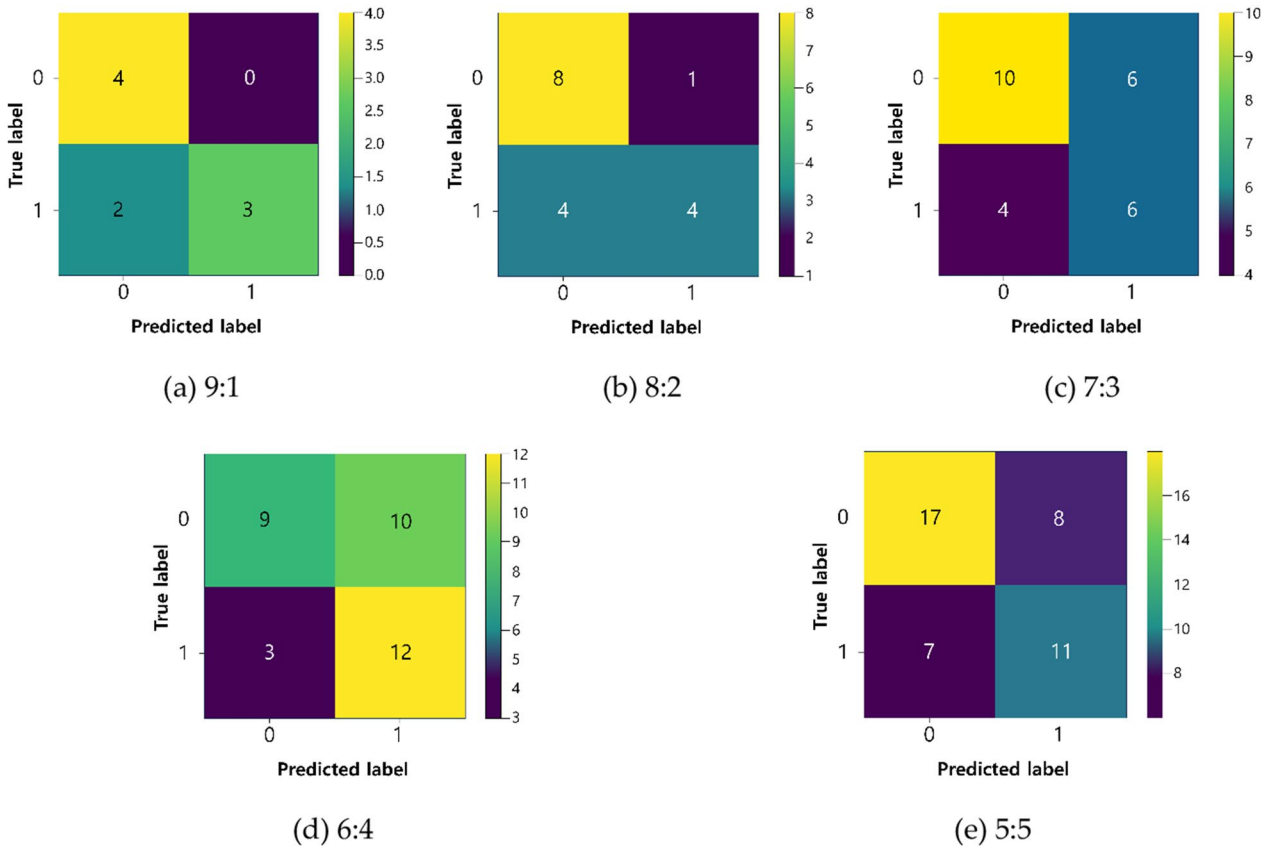

Fig. 7 Confusion matrix results for XGBoost prediction models employing each data ratio

Table 5 Results of XGBoost prediction for different ratios of training and test data

Ratio of train and test data			Precision	Recall	F1-score	Accuracy	AUC
9:1	Train	0	0.92	0.88	0.90	0.89	0.896
		1	0.86	0.91	0.89		
	Test	0	0.67	1.00	0.80	0.78	0.800
		1	1.00	0.60	0.75		
8:2	Train	0	0.89	0.92	0.90	0.90	0.896
		1	0.90	0.88	0.89		
	Test	0	0.60	0.89	0.76	0.71	0.694
		1	0.80	0.50	0.62		
7:3	Train	0	0.93	0.89	0.76	0.90	0.898
		1	0.88	0.50	0.62		
	Test	0	0.71	0.62	0.67	0.62	0.613
		1	0.50	0.60	0.55		
6:4	Train	0	0.95	0.77	0.85	0.86	0.865
		1	0.80	0.96	0.87		
	Test	0	0.75	0.47	0.58	0.62	0.637
		1	0.55	0.80	0.65		
5:5	Train	0	0.85	0.85	0.85	0.86	0.857
		1	0.86	0.86	0.86		
	Test	0	0.71	0.68	0.69	0.65	0.646
		1	0.58	0.61	0.59		

Table 6 Each factor's influence in the XGBoost 8:2 model

Factors	Feature importance	Influence (%)	Rank
Saturated unit weight	0	0	6
Porosity	31	7.1	5
Friction angle	105	24.2	2
Cohesion	0	0	6
Soil depth	164	37.8	1
Slope angle	84	19.4	3
Profile curvature	50	11.5	4
Plan curvature	0	0	6
TWI	0	0	6

Table 7 Results of hyperparameter optimization for LightGBM

Train and test data ratio	9:1	8:2	7:3	6:4	5:5
learning_rate	0.01	0.01	0.01	0.01	0.01
n_estimators	1500	1500	1500	500	500
max_depth	2	2	2	2	2
gamma	–	–	–	–	–
colsample_bytree	–	–	–	–	–
max_delta_step	0	0	0	0	0
min_child_weight	0	0	0	0	0
reg_alpha	0	2	2	3	0
reg_lambda	0	0	0	0	0
subsample	1	1	1	1	1
scale_pos_weight	1	1	1	2	1

XGBoost

Hyperparameter optimization, the most critical analysis process in machine learning, is first performed by grid searching. This involves selecting hyperparameter values that exhibit the highest performance by selecting hyperparameter candidate values at regular intervals. Hyperparameter selection uses verification in five layers (Table 4).

Log-loss is used to evaluate the performance of the training and test models. The learning performance and confusion matrix results are shown in Figs. 6 and 7, respectively, and Table 5 lists the predictive performance results for each model. The learning performance results for the prediction model show similar performance across most of the training models, but a 9:1 ratio of training to test data gives the best performance. For the test models, using an 8:2 ratio gives the best performance. Outstanding predictive performance is obtained, as indicated by the accuracy and AUC ranging from 60 to 90% and the difference in accuracy and AUC between the training and test models being < 20%. For precision and recall, there is a significant trade-off in the prediction model that uses all the test data. The precision and recall of the model using a 9:1 data ratio are both 100%, depending on the label value. Given the excessively small proportion of test data, the probability of predicting an actual "0" value as "0" or an actual "1" value as "1" is considered unreliable. Training–test data ratios of 8:2 and 7:3 minimize the trade-off. As a data ratio of 8:2 leads to a slightly higher performance than 7:3, it is considered optimal for a prediction model using XGBoost.

Table 6 lists the influence of each factor in the XGBoost 8:2 model. Soil depth has the most prominent influence, followed by friction angle, slope angle, plan curvature, and porosity. The saturated unit weight, cohesion, profile curvature, and TWI appear to have no significant influence in this model.

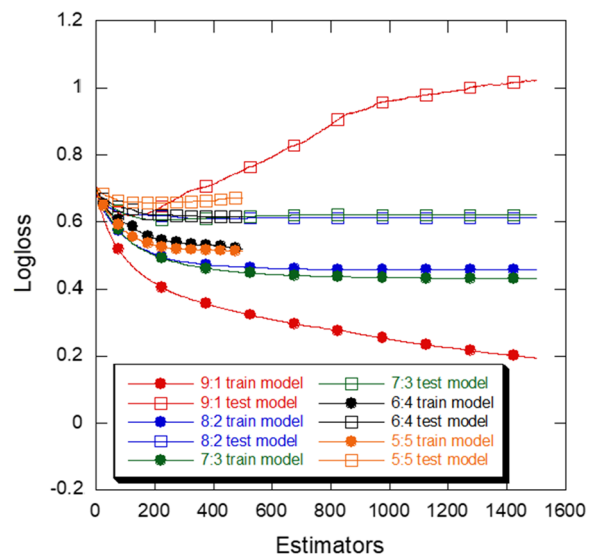

Fig. 8 Results of learning performance for LightGBM prediction models using different data ratios

LightGBM

LightGBM applies grid searching for hyperparameter optimization, which uses verification in five layers (Table 7).

For LightGBM, log-loss is used to evaluate the performance of the training and test models. The learning performance and confusion matrix results are shown in Figs. 8 and 9, respectively, and Table 8 lists the predictive performance results for each model. The learning performance results for the prediction model show that the training model with a 9:1 training-to-test-data ratio performs best, similar to the learning performance results for XGBoost, and the test model with an 8:2 ratio performs best. However, the learning performance of

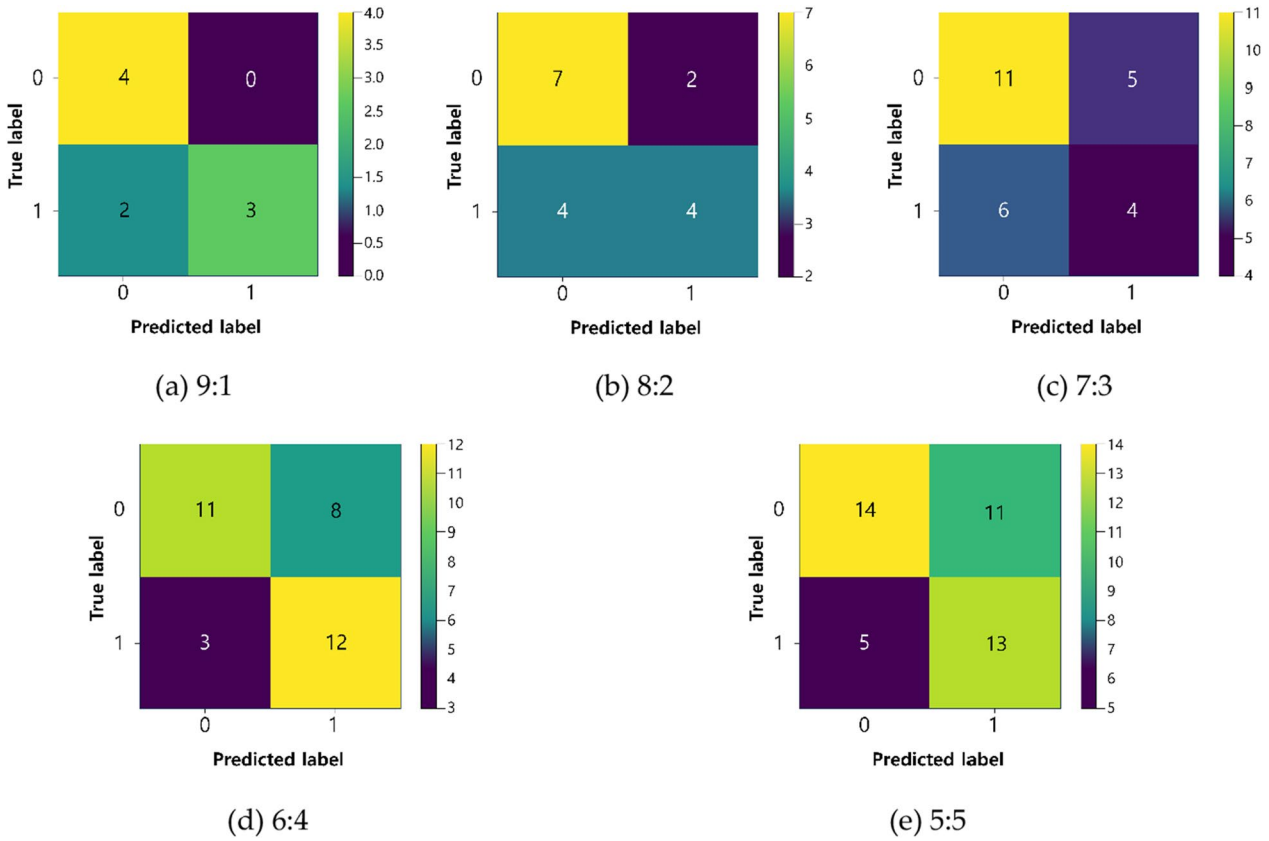

Fig. 9 Confusion matrix results for LightGBM prediction models using different data ratios

Table 8 Results of LightGBM prediction for different training and test data ratios

Ratio of train and test data			Precision	Recall	F1-score	Accuracy	AUC
9:1	Train	0	0.95	0.98	0.96	0.96	0.959
		1	0.97	0.94	0.96		
	Test	0	0.67	1.00	0.80	0.78	0.800
		1	1.00	0.60	0.75		
8:2	Train	0	0.90	0.78	0.84	0.84	0.842
		1	0.78	0.911	0.84		
	Test	0	0.64	0.78	0.70	0.65	0.639
		1	0.67	0.50	0.57		
7:3	Train	0	0.92	0.83	0.87	0.88	0.880
		1	0.85	0.93	0.89		
	Test	0	0.65	0.69	0.67	0.58	0.544
		1	0.44	0.40	0.42		
6:4	Train	0	0.89	0.65	0.76	0.78	0.787
		1	0.72	0.92	0.81		
	Test	0	0.79	0.58	0.67	0.68	0.689
		1	0.60	0.80	0.69		
5:5	Train	0	0.78	0.70	0.74	0.76	0.759
		1	0.75	0.82	0.78		
	Test	0	0.74	0.56	0.64	0.63	0.641
		1	0.54	0.72	0.62		

Table 9 Influence of each factor in the LightGBM 8:2 model

Factors	Feature importance	Influence (%)	Rank
Saturated unit weight	0	0	9
Porosity	60	3.2	5
Friction angle	708	37.3	2
Cohesion	118	6.2	4
Soil depth	853	44.9	1
Slope angle	24	1.3	6
Profile curvature	120	6.3	3
Plan curvature	13	0.7	7
TWI	2	0.1	8

the 9:1 test model decreases as learning progresses. The predictive performance results show that the accuracy and AUC decrease as the proportion of the training data decrease in the training and test models. There is a significant trade-off in most of the test models among the recall, precision, and confusion matrix results. Among them, those with 8:2 and 5:5 ratios perform best. Given the high proportion of test data for the 5:5 model, a 8:2 data ratio is considered optimal for the prediction model.

Table 9 lists the influence of each factor on the LightGBM 8:2 model. Soil depth is the most influential, followed by friction angle, plan curvature, cohesion, porosity, slope, profile curvature, TWI, and saturated unit weight. However, the model depends markedly on the first two factors, which represent 82% of the total influence; the other factors have no significant influence (each < 7%).

Discussion: results and comparison of methods

Table 10 summarizes the influence of the selected factors for each analysis method. Soil depth is consistently the most influential (> 25%). The rankings of friction angle, slope angle, and porosity differ slightly among the analysis methods. Friction angle shows uniform influence (> 10%) across all the methods. Porosity substantially influences LR and SEM analyses but has minimal effect on machine learning. Among the machine learning methods, only XGBoost is substantially influenced by slope angle. Saturated unit weight, profile curvature, plan curvature, TWI, and cohesion generally have small (< 10%) influences.

Soil depth is the most influential factor because it relates directly to conditions that may cause landslides. The data in section "The theory of structural equation model" clearly show its correlation with landslide occurrence: soil depth is generally ≤ 2 m in areas with no landslide and ≥ 1 m in most areas where landslides have occurred. Friction and slope angles are highly influential, as they directly affect the driving and resistance forces of soil (Mehrotra et al. 1992; Budimir et al. 2015; Çellek 2020). Porosity is rarely considered significant when investigating the factors influencing landslides on natural slopes. However, when artificial compaction is performed, as on forest road slopes, porosity is highly influential because it represents the degree of compaction. Unit weight ranks fifth here for artificial slopes because porosity and unit weight are inversely proportional. This study finds topographic factors (profile curvature, plan curvature, and TWI) to be insignificantly influential because landslides around forest roads are more affected by the degree of compaction or resistance force than the concave or convex shape of the slope. Cohesion acts only on resistance force in stability analysis and significantly affects slope activities (Cousins 1978; Ahmadi-Adli et al. 2014; Lin et al. 2016). However, this study attributes insignificant influence to cohesion because the sandy (SP to SW) soil in the study area has low cohesion, and the calculation of cohesion significantly deviated as the Mohr–Coulomb failure criterion was applied within a small range (the median value of

Table 10 Summary of influence and rank for analysis methods

Influential factors	Influence (%)					Rank
	LR	SEM	XGBoost	LightGBM	Arithmetic mean	
Saturated unit weight	8.4	16.8	0	0	6.3	5
Porosity	22.3	19.3	7.1	3.2	13.0	3
Friction angle	15.8	10.3	24.2	37.3	21.9	2
Cohesion	2.6	0	0	6.2	2.2	9
Soil depth	25.8	35.5	37.8	44.9	36.0	1
Slope angle	1.2	0.9	19.4	1.3	8.2	4
Profile curvature	6.5	0.7	11.5	6.3	6.3	5
Plan curvature	6.6	5.2	0	0.7	3.1	7
TWI	10.8	1.3	0	0.1	3.1	7

cohesion is higher for occurrence sites, whereas IQR and whisker are higher for non-occurrence sites; Fig. 4h).

Conclusions

Data for a mountainous area with forest roads were acquired through geological surveying, sampling, and laboratory testing, and the influence on landslide susceptibility of each measured parameter was analyzed using statistical and machine learning methods. The results are summarized as follows.

(1) The target area was Sancheok-myeon, Chungju-si, where rainfall of 259 mm on August 2, 2020 caused several landslides along the forest road in a narrow area of approximately 3 km × 4 km. As the area has the same rainfall and vegetation conditions, the influences of the physico-mechanical characteristics of the soil and topographic characteristics could be analyzed precisely.
(2) Geological surveying and sampling were conducted at 40 survey points where landslides occurred and 45 points where they did not. The soil's physico-mechanical characteristics and topographic factors for each survey point were acquired. Only nine factors were subjected to statistical analysis and machine learning methods.
(3) LR and SEM analysis results showed high accuracy, with values of 0.776 and 0.763, respectively. XGBoost and LightGBM exhibited outstanding performance in predicting landslides, with accuracy and AUC of 60%–90%, and differences of < 20% between the training and test data.
(4) All analysis methods identified soil depth as having the greatest influence on landslide occurrence. Friction angle, slope angle, and porosity were also selected as influential factors, although they differed slightly in the rankings of the different analysis methods.

As the analysis results of this study are for an area across which rainfall and vegetation conditions are largely consistent, the influences of the soil's physico-mechanical characteristics and the topography were analyzed more precisely than in studies comparing landslides across multiple regions. The results of this study are expected to be useful in the preparation of landslide vulnerability maps around forest roads.

Acknowledgements
Authors are grateful to editorial board and anonymous reviewers for the constructive comments that improved the manuscript.

Author contributions
MSW, SYS: Conceptualization, methodology, MSW, KHS: Investigation, data collection, MSW, KHS, NJD, KSS: Data analysis, software, validation, MSW, NJD: Writing-original draft preparation, SYS, KSS: Reviewing and editing of draft, supervision. All authors have read and approved the final manuscript.

Funding
This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(No.2020R1A6A3A03038855).

Availability of data and materials
Data and materials are available upon request.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Author details
[1]Department of Earth and Environmental Sciences, Chungbuk National University, Cheongju, South Korea. [2]Jeollanamdo Carbon Neutral Center, Jeonnam Research Institute, Naju, South Korea. [3]Chungcheongbuk-do Safety Research Institute, Chungbuk Research Institute, Cheongju, South Korea. [4]Department of Energy and Resources Engineering, Chosun University, Gwangju, South Korea.

Received: 20 August 2023 Accepted: 15 January 2024
Published online: 19 January 2024

References

Ahmadi-Adli M, Huvaj N, Toker NK (2014) Effects of the size of particles on rainfall-induced slope instability in granular soils. In: Proceedings of the Geo0Congress 2014, Altanta, GA, USA, 23–26 Feburary 2014
An KM (2021) Developing a prediction model for firm innovation and performance using statistical matching and machine learning ensemble techniques. Dongguk University, Doctoral dissertation 289
ASTM D2216-10 (2010) Standard test methods for laboratory determination of water (moisture) content of soil and rock by mass. ASTM International, West Conshohocken, PA.https://doi.org/10.1520/D2216-10
ASTM D2487-17 (2017) Practice for classification of soils for engineering purposes (Unified Soil Classification System). ASTM International, West Conshohocken, PA, 2017. https://doi.org/10.1520/D2487-17
ASTM D3080-98 (1998) Standard test method for direct shear test of soils under consolidated drained conditions. ASTM International, West Conshohocken, PA.https://doi.org/10.1520/D3080-98
ASTM D422-63 (2007) Standard test method for particle-size analysis of soils. ASTM International, West Conshohocken, PA, 2007. https://doi.org/10.1520/D0422-63R07E02
ASTM D854-10 (2010) Standard test methods for specific gravity of soil solids by water pycnometer. ASTM International, West Conshohocken, PA.https://doi.org/10.1520/D0854-10
Beven KJ, Kirkby MJ (1979) A physically based, variable contributing area model of basin hydrology. Hydrol Sci Bull 24:43–69
Broothaerts N, Kissi E, Poesen J, Van Rompaey A, Getahun K, Van Ranst E, Diels J (2012) Spatial patterns, causes and consequences of landslides in the Gilgel gibe catchment, SW Ethiopia. CATENA 97:127–136
Budimir MEA, Atkinson PM, Lewis HG (2015) A systematic review of landslide probability mapping using logisitic regression. Landslides 12:419–436

Çellek S (2020) Effect of the slope angle and its classification on landslide. Nat Hazard 87:23
Chae BG, Kim WY, Cho YC, Kim KS, Lee CO, Choi YS (2004) Development of a logistic regression model for probabilistic prediction of debris flow. J Eng Geol 14(2):211–222 (**(in Korean with English abstract)**)
Chen F, Yu B, Li B (2018) A practical trial of landslide detection from single-temporal Landsat8 images using contour-based proposals and random forest: a case study of national Nepal. Landslides 15:453–464
Chen SC, Chang CC, Chan HC, Huang LM, Lin LL (2013) Modeling typhoon event-induced landslides using GIS-based logistic regression: A case study of Alisan Forestry Railway, Taiwan. Math Problems Eng Article ID 728304
Chen T, Guestrin C (2016) XGBoost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGK DD International Conference on Knowledge Discovery and Data Mining, Newyork, USALACM, pp 785–794. https://doi.org/10.1145/2939672.2939785
Choi YH, Lee JW, Kim MJ (2011) A Study on development standard calculation program of forest road drainage facilities. J Kor Soc for Sci 100(1):25–33 (**(in korean with English abstract)**)
Cousins BF (1978) Stability charts for simple earth slopes. J Geotech Eng Div 104:267–279
CRED (2023) 2022 Disasters in numbers. Brussels: CRED, Retrieved from https://cred.be/sites/default/files/2022_EMDAT_report.pdf
Dikau R (1989) The application of a digital relief model to landform analysis in geomorphology. In: Three dimensional applications in geographical information systems. CRC Press, pp 51–77
DSBA (2020) November 12, 04-7: Ensemble Learning – XGBoost, Youtube, Retrieved from https://www.youtube.com/watch?v=VHky3d_qZ_E
Dutton AL, Loague K, Wemple BC (2005) Simulated effect of a forest road on near-surface hydrologic response and slope stability. Earth Surf Proc Land 30:325–338
Eker R, Aydin A (2014) Assessment of forest road conditions in terms of landslide susceptibility: a case study in Yığılca Forest Directorate (Turkey). Turk J Agric for 38(2):281–290
Fawcett T (2006) An introduction to ROC analysis. Pattern Recogn Lett 27(8):861–874
Fernandes NF, Guimarães RF, Gomes RAT, Vieira BC, Montgomery DR, Greenberg H (2004) Topographic controls of landslides in Rio de Janeiro: field evidence and modeling. CATENA 55:163–181
Gerrard J, Gardner R (2002) Relationships between landsliding and land use in the Likhu Khola drainage basin, Middle Hills, Nepal. Mount Res Dev 22:48–55
Godt JW, Baum RL, Savage WZ, Salciarini D, Schulz WH, Harp EL (2008) Transient deterministic shallow landslide modeling: requirements for susceptibility and hazard assessments in a GIS framework. Eng Geol 102(3–4):214–226
Hasegawa S, Dahal RK, Yamanaka M, Bhandary NP, Yatabe R, Inagaki H (2009) Causes of large-scale landslides in the Lesser Himalaya of central Nepal. Environ Geol 57:1423–1434
Hox JJ, Bechger TM (1999) An introduction to structural equation modeling. Family Sci Rev 11:354–373
John JC, Douglas S (2012) Landslides types, mechanisms and modeling. Cambridge University Press, p 420
Jotisankasa A, Vathananukij H (2008) Investigation of soil moisture characteristics of landslide-prone slopes in Thailand. In: International Conference on Management of Landslide Hazard in the Asia-Pacific Region 11th -15th November 2008 Sendai Japan, p 12
Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, Ye Q, Liu TY (2017) LightGBM: a highly efficient gradient boosting decision tree. Adv Neural Inf Process Syst 30:1
Kimaro DN, Msanya BM, Kilasara M, Mtakwa PW, Poesen J, Deckers JA (2000) Major factors influencing the occurrence of landslides in the northern slopes of the Uluguru Mountains, Tanzania. Workshop Presentation, pp 67–78
Kim HS (2022) Analysis on major influential factors and Occurrence probability of landslide in forest road. Phd Thesis, Chungbuk National University, p 114
KMA (Korea Meteorological Administration) (2023), Open MET data portal. https://data.kma.go.kr/resources/html/en/aowdp.html.
Korea Forest Service (2021) Comprehensive measures for national landslide prevention, p 1–56
Kutner MH, Nachtsheim CJ, Neter J (2004) Applied linear regression models, 4th ed. McGraw-Hill Education, p 701
Lin HD, Jiang YS, Wang CC, Chen HY (2016) Assessment of apparent cohesion of unsaturated lateritic soil using an unconfined compression test. In: Proceedings of the 2016 world congress on advances in civil, environmental, and materials research (ACEM16), Jeju, Korea, 28 August–1 September 2016
Luce CH, Wemple BC (2000) Special issue: hydrologic and geomorphic effects of forest roads. Earth Surf Proc Land 26:111–232
Mehrotra R, Namuduri K, Ranganathan N (1992) Gabor filter-based edge detection. Pattern Recogn 25(12):1479–1494
Moon SW, Yun HS, Seo YS (2020) Physical properties and friction characteristics of fault cores in South Korea. Econ Environ Geol 53(1):71–85
Moore DS, Notz WI, Flinger MA (2013) The basic practice of statistics, 6th ed. WH Freeman and Company, New York, NY, p 138
Nugraha H, Wacano D, Dipayana GA, Cahyadi A, Mutaqinc BW, Larasati A (2015) Geomorphometric characteristics of landslides in the Tinalah watershed, Menoreh Mountains, Yogyakarta, Indonesia. Procedia Environ Sci 28:578–586
Pham BT, Pradhan B, Bui DT, Prakash I, Dholakia MB (2016) A comparative study of different machine learning methods for landslide susceptibility assessment: a case study of Uttarakhand area (India). Environ Model Softw 84:240–250
Prokhorenkova L, Gusev G, Vorobev A, Dorogush AV, Gulin A (2017) CatBoost: unbiased boosting with categorical features. In: 32nd Conferences on neural information processing systems, Montreal, Canada, p 31
Osman N, Barakbah SS (2006) Parameters to predict slope stability soil water and root profiles. Ecol Eng 28:90–95
Owen RC (1981) Soil strength and microclimate in the distribution of shallow landslides. J Hydrol 20:17–26
Qu M, Bai Y, Hu Q, He L, Qiu E, Wan X (2021) A comprehensive prediction method for the saturated internal friction angle of sliding zone soils based on landslide engineering requirements. Geotech Eng 25:4144–4158
Quan HC, Lee BG, Lee CS, Ko JW (2011) The landslide probability analysis using logistic regression analysis and artificial neural network methods in Jeju. J Kor Soc Geospat Inf Sci 19(3):33–40 (**(in Korean with English abstract)**)
Regmi NR, Giardino JR, Vitek JD (2010) Modeling susceptibility to landslides using the weights of evidence approach: Western Colorado, USA. Geomorphology 115:172–187
Ryu SG (2008) Effects of Multicollinearity in logit model. J Kor Soc Transp 26(1):113–126 (**(in korean with English abstract)**)
Sanchez G (2013) PLS path modeling with R. Trowchez Editions, Berkeley: 222
Schmidt K, Roering J, Stock J, Dietrich W, Montgomery D, Schaub T (2001) The variability of root cohesion as an influence on shallow landslide susceptibility in the Oregon Coast Range. Can Geotech J 38:995–1024
Šimundić AM (2009) Measures of diagnostic accuracy: basic definitions. J Int Feder Clin Chem Lab Med 19(4):203–211
Ullman JB, Bentler PM (2013) Structural equation modeling. In: JA Schinka and WF Velicer (eds), Handbook of Psychology (vol 2), Research Methods in Psychology, Hoboken, NJ, Wiley, pp 661–690
Vanacker V, Molina A, Govers G, Poesen J, Dercon G, Deckers S (2005) River channel response to short-term human-induced change in landscape connectivity in Andean ecosystems. Geomorphology 72:340–353
Wang LH, Guo M, Sawada K, Lin J, Zhang J (2016) A comparative study of landslide susceptibility maps using logistic regression, frequency ratio, decision tree, weights of evidence and artificial neural network. Geosci J 20(1):117–136
Wemple BC, Jones JA, Grant GE (1996) Channel network extension by logging roads in two basins, Western Cascades, Oregon. J Am Water Resour Assoc 32(6):1195–1207
Wobus CW, Hodges KV, Whipple KX (2003) Has focused denudation sustained active thrusting at the Himalayan topographic front? Geology 31:861
Wright S (1921) Correlation and causation. J Agric Res 20:557–585
Xiao T, Segoni S, Chen L, Yin K, Casagli N (2020) A step beyond landslide susceptibility maps: a simple method to investigate and explain the different outcomes obtained by different approaches. Landslides 17:627–640
Yoon YG (2020) Feature extraction and analysis of electrocardiogram using LightGBM. Master's thesis of Korea University, p 28

Yung Y (2008) Structural equation modeling and path analysis using PROC TCALIS in SAS 9.2. in Proceedings of the SAS Global Forum 2008 Conference Cary NC SAS Institute Inc. Paper 384
Zikmund WG (2000) Business research methods (6th ed). Fort Worth: Harcourt College Publishers: 513

Publisher's Note