Landslides
DOI 10.1007/s10346-025-02513-y
Received: 26 May 2024
Accepted: 28 March 2025

Sujong Lee · Minwoo Roh · Hyun-Woo Jo · Kim Joon · Woo-Kyun Lee

Machine learning-based rainfall-induced landslide susceptibility model and short-term early warning assessment in South Korea

Abstract The extreme rainfall events associated with climate change trigger landslides. Approximately 60% of South Korea comprises mountainous terrain, with steep slopes rendering it particularly prone to landslides. Despite the implementation of early warning systems by the Korea Forest Service (KFS), landslide damage remains substantial, with approximately 2345 hectares affected over the past 5 years, resulting in severe human and economic costs. The current 24-h early warning system, based on Tier 3 administrative division (Town), faces challenges in accurately identifying high-susceptibility landslide areas. Thus, a daily landslide susceptibility model that integrates landslide-associated conditioning factors with meteorological, topographic, and environmental data was designed to assess landslide susceptibility with a spatial resolution of 100 m. Using AutoML, we identified Random Forest as the optimal model for predicting landslide susceptibility. Training the model with landslide data from 2016 to 2022 resulted in an accuracy of 0.93, AUC of 0.98, and F- 1 score of 0.98. A kappa value of 0.85 indicated the effective classification of past landslides using testing data. Location-based validation using 2023 occurrences revealed highly susceptible classifications for 88% of 43 landslides, while spatial scale-based hazard assessment using observed data indicated high hazard for 96% of 607 landslides in Tiers 3 and 4 (Township). Weather forecasting was also found to affect accuracy, with 76% accuracy for forecasts made at 5:00 PM and 41% for forecasts made at 8:00 AM. It was confirmed that further calibration of forecasting data can enhance the performance of the susceptibility model. The designed process thus enhances landslide prevention and preparedness on both local and regional scales, offering a crucial tool for mitigating the impact of landslides in South Korea.

Keywords Machine learning · Landslide susceptibility model · Landslide early warning process · Random Forest · Short-term weather forecasting

Introduction

Landslides are significant natural disasters resulting from multiple factors, including geological conditions and anthropogenic activities, and are particularly triggered by intense rainfall, flooding, rapid snowmelt, and earthquakes. They predominantly occur in mountainous regions and pose a risk to most countries worldwide (Wieczorek 1996; Dai et al. 2002; Highland & Bobrowsky 2008). Climate change exacerbates both the frequency and intensity of rainfall and the associated increase in extreme weather events globally has directly influenced landslide occurrence (Gariano & Guzzetti 2016; Nadim et al. 2006). Although heavy rainfall is the most influential factor, land use and land cover changes due to human activity have also been identified as important factors (Froude & Petley 2018). Particularly in the Northern Hemisphere, regions affected by extreme climatic events and human activity experience high concentrations of landslides. East and South Asia therefore suffer from large-scale and frequent landslides during the rainy season in summer (Petley 2012; Sim et al. 2022). The increase in the number of landslides suggests that they can no longer be solely considered natural disasters, but rather as climate change-induced risks. The negative impacts of landslides include casualties and property damage (Haque et al. 2019); however, the impact also encompasses the loss of forests and conversion of forests into carbon emissions (Geertsema et al. 2009; Liu et al. 2022).

South Korea, situated in the northern Mid-Latitude Region (MLR), has topographical and meteorological conditions that render it prone to landslides (KFS 2024a,b). With more than 60% of its territory covered by mountainous areas and rainfall concentrated in the summer season, landslides are experienced annually nationwide. Due to these topographical and meteorological conditions, the predominant types of landslides in South Korea are shallow landslides and debris flows (Lee et al. 2014). According to statistics from the KFS, the landslide-damaged area reached 13,676 ha over the 30-year period from 1993 to 2022, with more than 1.5 billion dollars allocated for recovery and more than 300 casualties reported. Although the most severe events were induced by typhoons in 2002, 2006, and 2022, massive landslide events occurred in the summer season of 2023 due to heavy localized rainfall (Ham et al. 2014; KMA 2024). Since 1980, the South Korean government and various research institutes have developed landslide models based on geological information systems (GIS) (Choi 1986; Carrara et al. 1999; Kim 2013). Recently, the necessity of an early warning system integrated with the landslide susceptibility model has expanded to prevent and prepare for landslides (Song et al. 2022).

Two primary types of models stand out in the domain of landslide modeling: statistical- and physical-based models (Reichenbach et al. 2018; Spiekermann et al. 2023). Both types use machine learning techniques or process-based algorithms to quantify landslide susceptibility. The former offers the advantage of covering vast areas but is limited by its lower spatial resolution and inability to simulate landslide mechanisms precisely. While the latter provides high spatial resolution and precision but its applicability is limited to specific areas. Statistical models that are focused on susceptibility have been developed using traditional machine learning techniques, including Bayesian probability (S. Lee et al. 2002; Sujatha et al. 2014), logistic regression (Atkinson & Massari 1998; Hemasinghe et al. 2018; Zhu & Huang 2006), random forest (Ng et al. 2021; Ren et al. 2024; Zhang et al. 2017), support vector machines (Ballabio & Sterlacchini 2012;

Published online: 16 April 2025

Dou et al. 2020; S. Lee et al. 2017), and boosting algorithms (Ng et al. 2021; Park & Kim 2019; Sahin 2020). With advancements in technology, recent studies on landslide susceptibility have been conducted using state-of-the-art methods, including automated machine learning (Ma et al. 2024; Tang et al. 2023) and deep learning (Achu et al. 2023; Hussain et al. 2023). Azarafza et al. (2021) and Nikoobakht et al. (2022) confirmed that deep learning-based landslide susceptibility models outperform traditional machine learning methods. However, these studies used these methods to produce static susceptibility maps, offering limited monitoring of landslide susceptibility under more frequent and intense rainfall conditions driven by climate change. Physical models focus on susceptibility using slope stability or hydrological analysis and have been used in landslide susceptibility studies such as the Transient Rainfall Infiltration and Grid-Based Regional Slope-Stability (TRIGRS) (Ciurleo et al. 2019; Dikshit et al. 2019; D. W. Park et al. 2013) and the development of the integrated hydrological–geotechnical model (Federici et al. 2015; Passalacqua et al. 2016). Furthermore, an integrated approach combining statistical and physical models has been introduced recently (Cui et al. 2024; Huang 2023; Yang et al. 2024).

Early warning systems must include hazard monitoring, forecasting and prediction, disaster risk assessment, communication, and preparedness activity, thus aiding systems and processes that enable individuals, communities, governments, businesses, and others to take timely action and reduce disaster risks before a hazardous event occurs (UNDRR 2016). Representative landslide early warning systems have previously been developed based on the rainfall intensity-duration threshold using landslide inventory and meteorological data (Caine 1980; Guzzetti et al. 2008), while Lee et al. (2021) analyzed the cumulative event rainfall duration threshold and applied it to an early warning system in Chuncheon, South Korea.

Due to its steep mountainous terrain and the rapidly changing characteristics of soil layers caused by their shallow depth, almost all of South Korea's territory is vulnerable to landslides (Choe 2001). Therefore, landslide early warning systems have been developed by both the KFS and the Korea Institute of Geoscience and Mineral Resources (KIGAM). The models produced by the KFS have been in development since 2013 and include a static landslide hazard map and dynamic landslide prediction system known as the Korea Landslide Early-warning System (KLES). Although the landslide hazard map considers only internal factors, the landslide prediction system also considers short-term weather forecasting and the soil moisture index (Lee et at. 2015), enabling the KFS to provide real-time landslide hazard alerts. The physical-based model developed by KIGAM focuses on assessing the sediment movement hazards in local areas (KIGAM 2019). Both systems use models to generate daily landslide susceptibility maps, identify high-susceptibility regions before events occur, and include short-term weather forecasting and various geological factors. Consequently, not only hazard areas but also potentially damaged areas can be predicted using these models. However, although both systems rely on robust models that are based on infinite slope stability analysis and can accurately predict hazard areas, there are still opportunities for improvement in terms of prevention and preparedness.

Currently, the KFS provides an early warning system targeting Tier 3 administrative divisions within 12 h in advance of landslides. In 2024, the KFS announced a new advancement plan for a landslide information system aimed at improving spatiotemporal coverage. Specifically, the KFS plans to provide landslide prediction information for up to four administrative divisions within 48 h in advance. Furthermore, the KFS set a plan to provide real-time landslide hazard information within 1 h by integrating the landslide hazard map and KLES. In contrast, the early warning system developed by KIGAM focuses solely on national parks and provides information 24 h in advance. While such systems provide timely and precise landslide warnings to prevent disasters, this information is typically disseminated through the web, only reaching local governments in Tier 3 areas, limiting accessibility for citizens.

This study aims to address these limitations by developing a machine learning-based, precise landslide susceptibility model with a 100 m spatial resolution that covers the entire country. By integrating 3-day weather forecasts, the proposed model predicts landslide susceptibility up to 72 h in advance, surpassing the spatial coverage and lead times of existing systems. The incorporation of daily weather forecasts provides a critical lead time, enabling not only hazard identification and preventive measures but also improving the accessibility of information for citizens. The research process involved (1) data preparation and sampling, (2) selecting the optimal machine learning model using PyCaret and development of the landslide susceptibility model, (3) building a semi-automatic preprocess to acquire 3-day weather forecasting data, (4) calculating the 3-day landslide susceptibility results and disseminating this information to citizens, and (5) validating and assessing the model and early warning results. This approach has the potential to deliver early warnings at finer administrative divisions by generating pixel-based susceptibility results. Consequently, the early warning process using daily weather forecasts allows citizens to more easily access timely information, enabling them to take preventive actions and prepare up to 72 h in advance, thereby enhancing community resilience against increasing landslide susceptibility.

Study area and applied data

Study area

South Korea, a peninsula located in the mid-latitude region of Eastern Asia, is approximately 70% mountainous territory (10,043,000 ha) (Fig. 1). The administrative division of South Korea is divided into four tiers from Tier 1, which includes 17 cities (si) and provinces (do), to Tier 4 (NGII 2015). The administrative divisions of South Korea are described in Table 1.

High mountains are distributed in the northeast and lower mountains in the southwest of the country. Although the average elevation of South Korea is approximately 300 m, which is lower than that of other East Asian countries, its complex geological structure has led to the formation of relatively steep slopes and diverse landform features (NGII 2020). South Korea experiences a monsoon climate that is characterized by cold, dry winters and hot, humid summers, with almost all rainfall concentrated in the summer season. Monthly average precipitation of 111 mm has been recorded over the past 7 years, with the highest precipitation of 482 mm observed in July 2023. Rainfall-induced landslides are reported annually, and the trend in rainfall-induced large-scale landslides appears to be increasing each year. Two-thirds of the bedrock in the study area consists of granite and metamorphic rocks, particularly gneiss, which has been associated with severe landslide

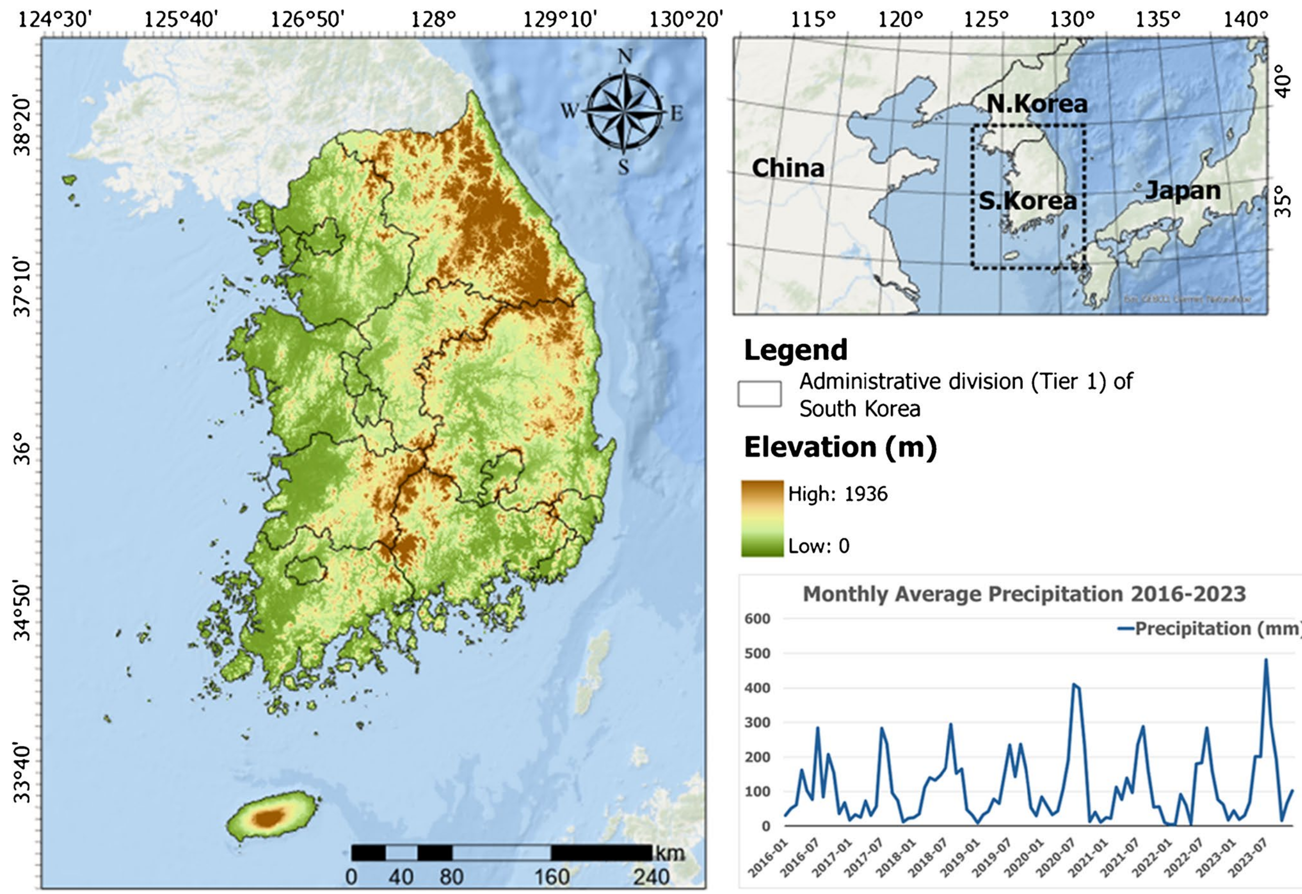


Fig. 1 Location of South Korea, elevation, and average monthly precipitation

Table 1 Description of the administrative division of South Korea

Tier	Romanized Korean	English equivalent
First	si	City
	do	Province
Second	si	City
	gun	County
	gu	District
Third	eup	Town
	myeon	Township
	dong	Township
Fourth	ri	Township

damage (Kim 2009). Over the past 10 years, approximately 2439 ha of the land in South Korea has been damaged by landslides, and the KFS has suggested that local heavy rainfall and illegal land use change have rendered South Korea increasingly susceptible to landslides (KFS 2023). On a governmental level, landslide susceptibility regions have traditionally been set and managed using static maps based on geological characteristics by KFS. However, almost all of the areas damaged by landslides have not been located within these designated susceptible regions. Furthermore, the current early warning system provided by KFS offers alerts within 12 h, which is not sufficient for adequate preparedness and prevention for citizens. This has highlighted the need for an overall alert process that incorporates real-time meteorological information.

Applied data

The representative causes of landslides include physical, natural, and human factors. To consider these factors, various geospatial data, such as topographic, terrain, bedrock, and forest cover maps, can be utilized (Highland & Bobrowsky 2008). In South Korea, the KFS classifies landslide conditioning factors as external or internal (KFS 2021). External factors, also known as physical factors, include rainfall intensity, prolonged rainfall, and earthquakes, while internal or natural factors include soil type, topography, and geological features. Thus, the data describing rainfall-induced landslides used in the susceptibility model were categorized into two groups. Both conditioning factors and landslide inventory data were considered in this study (Table 2). For internal factors, seven of the

Table 2 Factors and landslide inventory data used

Number	Type	Name	Resolution (spatial/temporal)	Variable type	Source
1	Inventory	Landslide inventory	Point/2016–2022		Korea Forest Service (KFS)
2	External factors	Daily rainfall	1 km/daily	Continuous	Korea Meteorological Administration (KMA)
3		5-day cumulative rainfall	1 km/daily	Continuous	
4	Internal factors	Slope	100 m	Continuous	Constructed based on data from NASA SRTM
5		Aspect	100 m	Categorical	
6		Curvature	100 m	Continuous	
7		Flow direction	100 m	Categorical	
8		TRI	100 m	Continuous	
9		TWI	100 m	Continuous	
10		SPI	100 m	Continuous	
11		Land cover status	100 m	Categorical	Ministry of Environment (ME)
12		Soil depth A	100 m	Continuous	Korea Forest Service (KFS)
13		Soil depth B	100 m	Continuous	
14		Bedrock type	100 m	Categorical	

major factors used in the landslide hazard map by KFS were pre-considered. Although the metadata format of each factor varies from vector to raster, all the preprocessed data are converted to a raster format with a 100 m spatial resolution and the same coordinate system (EPSG: 5186).

Landslide inventory data

The landslide inventory dataset provided by the KFS includes information on the occurrence periods, locations, and extent of the damage caused by shallow landslides and debris flows. A landslide inventory dataset from 2016 to 2022 with a total of 4215 landslides was used in this study (Fig. 2). However, the utilized dataset did not include all information concerning specific dates and longitude or latitude for the study period; thus, pre-processing was performed. First, the occurrence period was replaced by the specific dates of the heaviest recorded rainfall during the study period, which was determined by comparing the occurrence period with daily rainfall information. Second, the locations of Tier 3 and 4 administrative divisions provided were converted into latitude and longitude datapoints, with visual interpretation used to adjust inaccurate locations to include steep and forested areas.

A more recent landslide inventory for 2023, which includes a total of 798 landslides and inventories and more detailed location data, was used to validate the model, improving the assessment of early landslide warnings in these administrative divisions.

Meteorological factor

Daily rainfall data and 5-day cumulative rainfall data were applied with two types of meteorological data sources used for the training susceptibility model and short-term hazard assessment: historical rainfall data and short-term weather forecasting. Daily rainfall, such as 24-h rainfall, is fundamental information for predicting landslide occurrences, and heavy rainfall can trigger landslides regardless of geological and hydrological conditions (Dai and Lee 2001; Brand 1985). In South Korea, cumulative rainfall over 5 days has been shown to significantly impact landslide occurrences (Kang et al. 2016). Historical rainfall data from 2016 to 2022 were obtained from the Korea Meteorological Administration (KMA), or more precisely, the Automated Synoptic Observing System (ASOS) and Automatic Weather System (AWS), which cover 657 stations throughout the country. Date and daily rainfall pools were obtained and interpolated with a spatial resolution of 1 km using inverse distance-weighted interpolation. Precipitation, temperature, and wind speed were obtained from 3-day short-term weather forecasting provided by the KMA for each day from 2 am in 3-h intervals (KMA 2023). The objective was to obtain forecasts for Tier 3 administrative divisions and below for the convenience of citizens in preparing for dangerous weather. As the data were retrieved via an open API, preprocessing was conducted from acquisition to interpolation.

Environmental factors

National geospatial data were utilized to reflect both the land cover and soil factors. The utilized land cover map, which was provided by the Ministry of Environment (ME) of South Korea, encompasses various land cover types. Specifically, the Level- 2 land cover map produced in 2022 offers 22 types of land cover with a 5 m spatial resolution. Land cover types were reclassified into nine categories: deciduous forest, coniferous forest, mixed forest, urban areas, agricultural land, grassland, wetland, barren land, and water. The forest location soil map produced by the

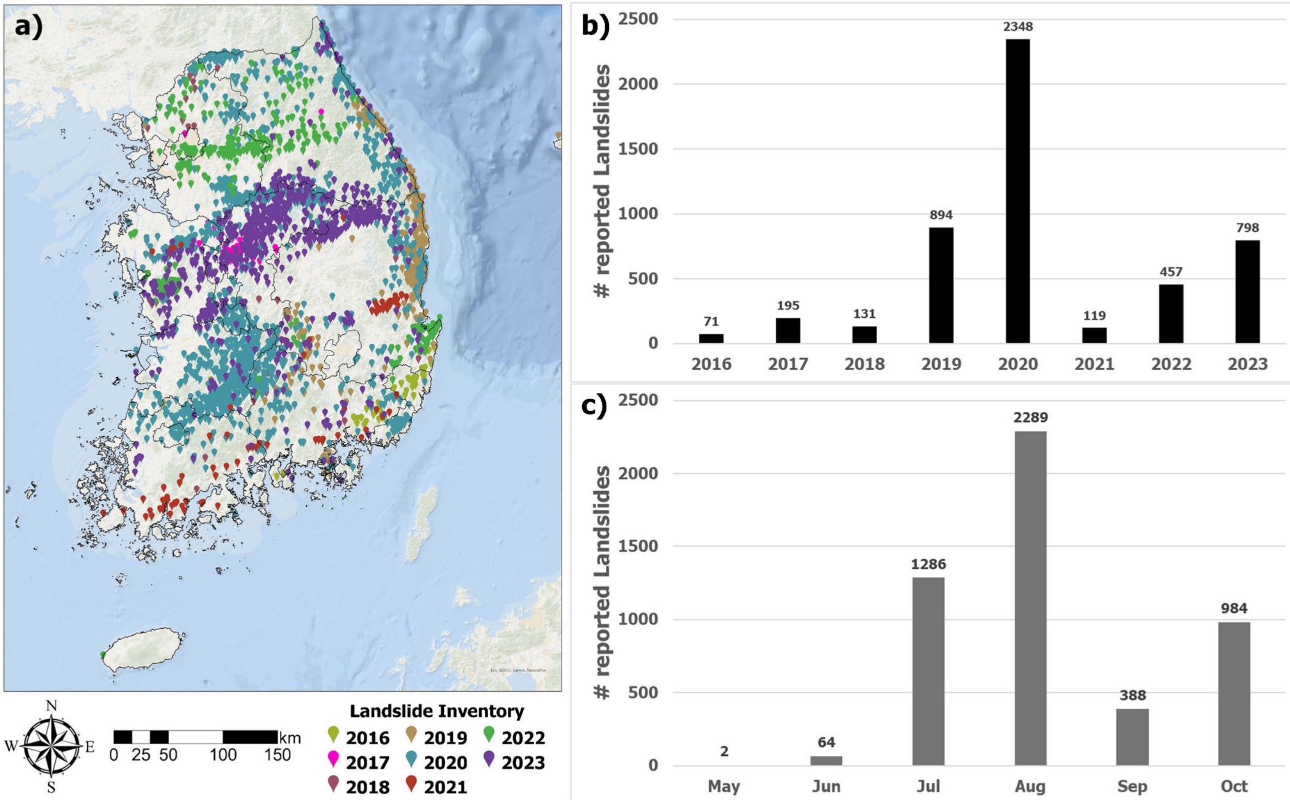


Fig. 2 **a** Landslide inventory data and number of reported landslides: **b** yearly and (**c**) monthly

KFS comprises bedrock types and soil depths, and a forest location soil map that is scaled to 1:25 000 was utilized in this study, allowing maps of parent rock types and the depth of the soil in two layers (A and B) to be created with a spatial resolution of 100 m. The parent rock-type includes igneous, sedimentary, and metamorphic rocks, as well as areas without information, and the soil depth ranges from 0 to 100 cm.

Topographical and hydrological factors

Topographical factors are essential for landslide modeling; thus, digital elevation model (DEM) data from the Shuttle Radar Topography Mission (SRTM) digital elevation database from NASA, which uses the Google Earth Engine platform, were utilized. This dataset is global in scope with a spatial resolution of 30 m. The slope, aspect, curvature, and flow direction were calculated from the DEM using ArcPro geoprocessing tools. Aspect and flow direction were reclassified into the eight cardinal directions. Three types of topographical and hydrological indices were estimated using the respective formulas: the Terrain Ruggedness Index (TRI, Eq. 1), which can quantify objective topographic heterogeneity by calculating the sum of changes in elevation between eight neighboring grid cells (Riley et al. 1999); the Topographic Wetness Index (TWI, Eq. 2), which represents the spatial distribution of soil moisture and surface saturation by quantifying the effects of the local topography on hydrological processes (Beven & Kirkby 1979; Qin et al. 2011); and the Stream Power Index (SPI, Eq. 3), which shows the erosive power of flowing water by calculating the slope and catchment area (Moore et al. 1991). The formulas for the indices are as follows:

$$\mathrm{TRI} = \sqrt{\max^2 - \min^2} \tag{1}$$

$$\mathrm{TWI} = \ln \frac{A_c}{\tan\beta} \tag{2}$$

$$\mathrm{SPI} = A_c \times \tan\beta \tag{3}$$

where max and min indicate the highest and minimum values for cells in 3 × 3 rectangular neighborhoods, A_c is the specific catchment area, and β is the slope angle at the point of interest.

Method

Data preparation and sampling

The landslide susceptibility model is a supervised machine learning technique; thus, a labeling dataset is necessary for training. Two labels are required in binary classification models: occurrence and non-occurrence. As landslide inventories include only landslide occurrence events, non-occurrence events need to be obtained via sampling. However, a concrete and hybrid sampling method for non-occurrence has not yet been established (Ren et al. 2024), and the traditional random sampling method is fraught with uncertainties and errors.

In this study, a spatiotemporal random sampling method was designed to reflect the landslide conditioning factors for the entire study area (Fig. 3). The designed sampling method considers both meteorological and other factors that can induce landslides. Sampling begins by setting the sample size ($N = 8082$) and considering the landslide inventory data. The sample is then divided into two groups ($n = 4041$), and samples are obtained for each group by considering the spatiotemporal conditions of the landslide inventory data. The first group was sampled from the same location as the inventory data but on different dates, whereas the other group was sampled from different locations on the same date. Dates within the inventory period were selected, and locations were extracted from a 1-km grid that was constructed for South Korea. In the final data preprocessing step, outliers located outside the study area boundary were removed, resulting in a total of 11,862 labeling datasets, including 4041 occurrences and 7821 non-occurrences.

Automated machine learning and Random Forest

Automated machine learning (AutoML) is an end-to-end machine learning process that is accessible for implementing state-of-the-art machine learning approaches (Hutter et al. 2019; Kanti Karmaker et al. 2021). PyCaret is a representative AutoML library implemented in a Python environment that offers an end-to-end pipeline with low code and performs time-consuming procedures from data preprocessing to modeling functions (Ali 2020; Chauhan et al. 2020; Sarangpure et al. 2023). PyCaret automatically detects data types and can thus distinguish between numerical and categorical data during preprocessing, after which it splits the data into training and testing sets. The modeling function of PyCaret provides more than 15 classification algorithms and an optimization function that is based on a custom grid search with cross-validated results using user-defined fold and hyperparameter candidates. The initial model selection and optimization are implemented by focusing on the evaluation and improvement of various performance criteria: accuracy, area under the receiver operating characteristic curve (AUC), precision, recall, F- 1, and kappa value. Following optimization, users can access the analysis functions for model explainability and interpretability.

Random Forest is a parallel ensemble learning technique that involves the aggregation of numerous decision tree models in a process called bagging, which minimizes variance and overfitting when dealing with complex and sizable datasets (Breiman 2001). The key algorithms used by Random Forest depend on the diversity and randomness of the dataset. Random Forest is executed by constructing various training datasets, known as bootstraps, for each decision tree, which is achieved through random sampling

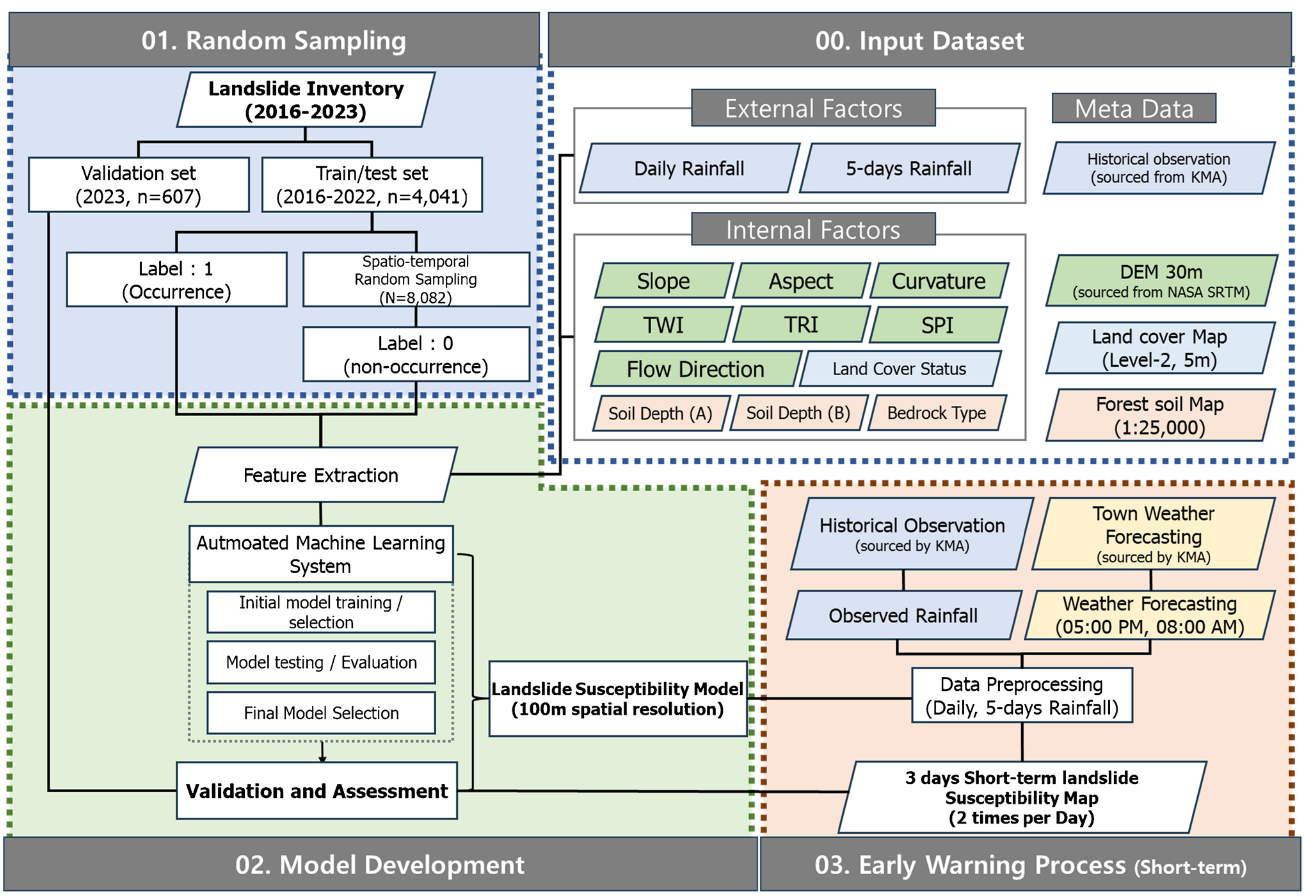


Fig. 3 Research flow used in the development of the landslide susceptibility model and early warning process

and replacement (Abellán et al. 2018). In addition, each decision tree employs a random subspace approach during the selection of optimal variables for the decision tree branch points. This method involves the random selection of a smaller number of variables than those included in the original dataset and continues until a fully grown tree is achieved. This is followed by a majority vote mechanism that is based on the outcomes of each decision tree. A detailed description of Random Forest is provided in Breiman (2001).

Short-term early warning process

A short-term early warning process was designed to predict landslide susceptibility within 3 days (or 72 h), which involves updating a landslide susceptibility map twice daily, at 9:00 am and 6:00 pm, using forecasting data from 8:00 am and 5:00 pm. For the 9:00 am update, the early warning assessment begins from the reference date (D), whereas for the 6:00 pm update, it starts from the day following the reference date ($D + 1$). Specifically, two types of meteorological data are used to operate the model: daily and 5-day cumulative rainfall data. Weather forecasting data is utilized as daily rainfall data, and observed data from ASOS and AWS are applied for calculating the 5-day cumulative rainfall data. The combination of input data to estimate the 5-day cumulative rainfall differs slightly during the updating time as illustrated in Fig. 4. The short-term early warning process includes the preprocessing of town weather forecasting and observed data and the operation of a landslide susceptibility model. All the series of preprocesses are implemented in a Python environment at each of two forecasting times (8:00 am and 5:00 pm) daily and semi-automatically. Preprocessing involves the following steps:

1. Forecasting data acquired from the KMA API hub is implemented through a semi-automated process in a Python environment, using data retrieved at 8:00 am and 5:00 pm daily.
2. Hourly forecasting is converted into daily data through time-series merging.
3. Data acquired from 3831 stations located in Tier 3 administrative divisions are subjected to spatial interpolation to generate raster data.
4. Observed data by ASOS and AWS from the KMA Data Portal are acquired at 8:00 am daily. This includes observation data from four days ago up to the previous day.
5. Data acquired from 657 stations are subjected to spatial interpolation to generate raster data.
6. Five-day cumulative rainfall is calculated by merging with observation data.
7. The landslide susceptibility model is operated with the changing external (meteorological) factors.

Validation and assessment

A process for validating the model and assessing the results was also designed to verify the effectiveness of the landslide early warning system for South Korea (Fig. 5). This process requires an adequate landslide inventory and detailed information. Obtaining sufficient landslide inventory data for South Korea is challenging; however, over 4000 landslide events were obtained and utilized to train the model, with more recent landslide inventory data from 2023 used for validation and assessment. This dataset initially contained information on 798 events across the entire territory; however, after preprocessing steps such as the removal of duplicates, this number was reduced to 609. While the inventory data include Tier 3 and 4 locations along with dates of landslide occurrence, some entries involved more detailed locations for recovery planning. Inventory data with more detailed locations were used to validate the model, with 609 entries used to assess the early warning results.

Observed rainfall data were used to confirm the prediction performance and validate the model. After computing a daily susceptibility map for past events, location-based validation was performed using detailed coordinate information. Each pixel of the landslide susceptibility map assigns susceptibility indices ranging from 0 to 1, which are categorized into five levels at 0.2 intervals from very low to very high. Consequently, the location-based validation results displayed both the susceptibility index and grade.

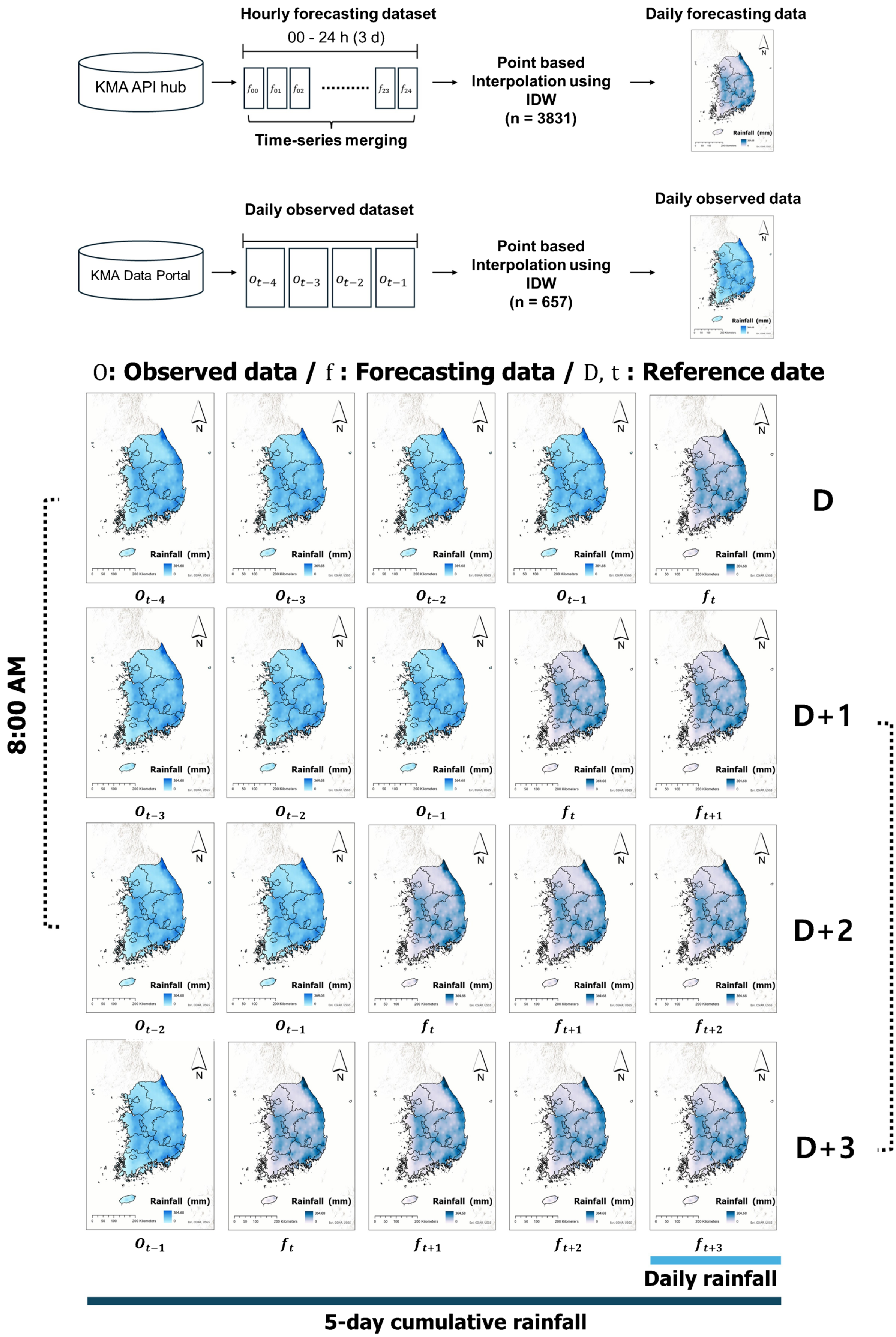
The landslide hazard criteria for different spatial scales were applied by aggregating pixel-based results into detailed administrative boundaries using a categorization method (KFS 2024a, b). Unlike the universal landslide hazard assessment framework proposed by van Westen et al. (2006, 2008), which integrates susceptibility with magnitude and frequency, the KFS framework employs a simplified matrix-based method. This method combines static landslide susceptibility maps with weather forecasting information at the watershed scale or Tier 4 administrative divisions. Using these criteria, spatio-temporal zonal statistics were calculated for Tier 3 and Tier 4 boundaries to analyze the distribution of susceptibility levels and assign hazard categories (Fig. 5). The categories were validated against historical landslide-affected divisions to confirm the model reliability.

Results

Landslide susceptibility model and performance

A labeling dataset, consisting of 11,862 data points that include information on conditioning factors obtained through feature extraction, was utilized in PyCaret. Particularly, external factors were extracted spatiotemporally based on the dates of occurrence and non-occurrence. A 7:3 ratio was used to split the training and testing data, and stratified k-fold cross-validation ($k = 3$) was applied for the initial model selection. A total of 14 machine learning models were trained and evaluated and the initial top 5 models were selected based on their accuracy rankings. The results demonstrated superior performance for the ensemble-based algorithms during training, with Random Forest ranking highest (Table 3). The performance indicators for Random Forest consistently outperformed the other models, whereas the boosting-based algorithms exhibited high precision scores. These findings suggest that Random Forest, which uses bagging algorithms to minimize variance and overfitting, may be more suitable for simulating past rainfall-induced landslide events than boosting algorithms.

Optimization of Random Forest was also implemented using a grid search and stratified k-fold cross-validation (k



Hourly forecasting dataset
00 - 24 h (3 d)
KMA API hub
f_{00} f_{01} f_{02} .... f_{23} f_{24}
Time-series merging
Point based Interpolation using IDW (n = 3831)
Daily forecasting data
Daily observed dataset
KMA Data Portal
o_{t-4} o_{t-3} o_{t-2} o_{t-1}
Point based Interpolation using IDW (n = 657)
Daily observed data
O: Observed data / f : Forecasting data / D, t : Reference date
8:00 AM
D
D+1
D+2
D+3
5:00 PM
o_{t-4} o_{t-3} o_{t-2} o_{t-1} f_t
o_{t-3} o_{t-2} o_{t-1} f_t f_{t+1}
o_{t-2} o_{t-1} f_t f_{t+1} f_{t+2}
o_{t-1} f_t f_{t+1} f_{t+2} f_{t+3}
Rainfall (mm)
Daily rainfall
5-day cumulative rainfall

◂**Fig. 4** Combination of observed and forecasting data for early warning process

= 3) on the defined hyperparameter candidates, resulting in improved performance indicators (Table 4). Specifically, the optimization function provided by PyCaret was applied, and the optimized results based on the criteria of accuracy and kappa showed improved outcomes among the six criteria used for optimization. Despite these improvements, a tradeoff between recall and precision was observed; however, the performance indicators suggest that the landslide susceptibility model can effectively reconstruct past landslides and predict future landslides.

According to the confusion matrix, Cohen's kappa value was calculated at 0.845, indicating that the classifications made by the optimized model were accurate (Fig. 6). Approximately 93% of the landslide occurrences and 94% of the non-occurrences were correctly classified. The minimal difference observed between the falsely classified results may be attributed to mislocated data. The receiver operating characteristic (ROC) curve was generated for both landslide occurrences (class 1) and non-occurrences (class 0) using an independent testing dataset, separate from the training data used for model development (Fig 6). The *X*-axis represents the true positive rate, which indicates the ability to correctly predict an actual positive case in each class. The curve skews toward the top-left corner, with an AUC value of 0.981, demonstrating that the Random Forest-based model is highly effective at classifying both past actual landslide occurrences and non-occurrence. This result suggests that the differences in the factors between the two distributions of occurrence and non-occurrence are clearly and accurately reflected in the model.

The feature importance of Random Forest is calculated by reducing the impurities in the nodes using entropy. Analysis of the results revealed that the external meteorological factors significantly influence the occurrence of landslides, with daily rainfall being the most influential, followed by the 5-day cumulative rainfall (Fig. 7a). Although the gap in importance scores for external and internal factors is considerable, internal factors also significantly impact landslide occurrence. The SHapely Additive exPlanations (SHAP) method is a permutation feature importance technique for sensitivity analysis that quantitatively indicates the contribution of each feature to a prediction result while maintaining consistency (Lundberg et al. 2017). The *X*-axis represents the impact of each factor's contribution to the model's output. The color indicates the range of each factor value from low (blue) to high (red), with higher density on the line indicating a higher distribution of values (Fig. 7b). The analysis was conducted under the assumption of independence of each factor, and SHAP values were calculated to determine the impact of specific factors on the outcome while keeping other factors fixed. The results suggest a positive relationship between meteorological factors and the prediction of landslide occurrence, indicating that higher daily rainfall and 5-day cumulative rainfall significantly increase the susceptibility to landslides. A low value of TRI and slope contributed to non-occurrence, whereas the depth of soil B and SPI indicated a slightly positive contribution to landslide occurrence.

Short-term early warning results

Approximately 800 landslide events occurred in South Korea in 2023, with almost all disaster events occurring in July. A landslide susceptibility map was therefore created using a short-term early warning process from the end of June to August. Although a prototype susceptibility map was calculated once daily until the end of June, the susceptibility was computed twice daily in July using a semi-automated short-term warning process and the obtained early warning information has been provided on the OJeong Resilience Institute Website (OJERI@KU) since June 2023, with simple descriptions included. A total of 54 landslide susceptibility maps have been produced, with 22 announcements made over 14 days corresponding to severe landslides. The susceptibility index and level can be seen on the map. The results were stored along with forecast rainfall data for validation and assessment (Table 5).

Results of validation and assessment

Location-based validation revealed that nearly 35% of the landslide events were categorized as very high, with 54% classified as high (Table 6). Moderate and low events accounted for 9% and 2% of cases, respectively, and no events were classified as very low. Regarding the susceptibility indices, events classified as very high ranged from 0.8 to 0.95, while high events ranged from 0.6 to 0.79. The mean susceptibility index value was 0.73, with a range of 0.95 to 0.36. Most events occurred on July 14 and 15, with others recorded in June and August. The figures show significant location-based results for July 14 and 15 (Figs. 8 and 9). Calculation with 100 m resolution allowed more precise prediction of each estimated landslide occurrence event. The number of events per level statistics for July closely mirrors the total statistics. The lowest level on June 30 coincided with lower daily rainfall. Occurrences in moderate susceptibility areas were noted, even for regions experiencing extremely high rainfall, suggesting potential limitations in terms of accurate coordinate information and internal factors, which contribute to the predictive accuracy of the model.

Results based on the utilization of three different meteorological datasets are presented via a spatial scale-based assessment (Table 7). A total of 607 landslide events occurred in Tier 3 and 4 administrative divisions from the end of June to August. Each row in the event count columns indicates the number of Tier 3 and 4 divisions in each category.

In the case of observed data, approximately 96% of the regions in which landslides occurred were classified as having high or very high categories, with approximately 98% of areas categorized as being at hazard of landslides classified as moderately high or higher. Very high was most frequently observed on July 15 and 14, accounting for approximately 90% of the category.

In terms of weather forecasting data, different results were obtained for the data captured at different times (Figs. 10 and 11). The 5:00 PM data were acquired the day before the target date, and the forecast lasted from the next day to 3 days later. Approximately 93% of the regions were classified as moderate or high using this method, with only 7% classified as low

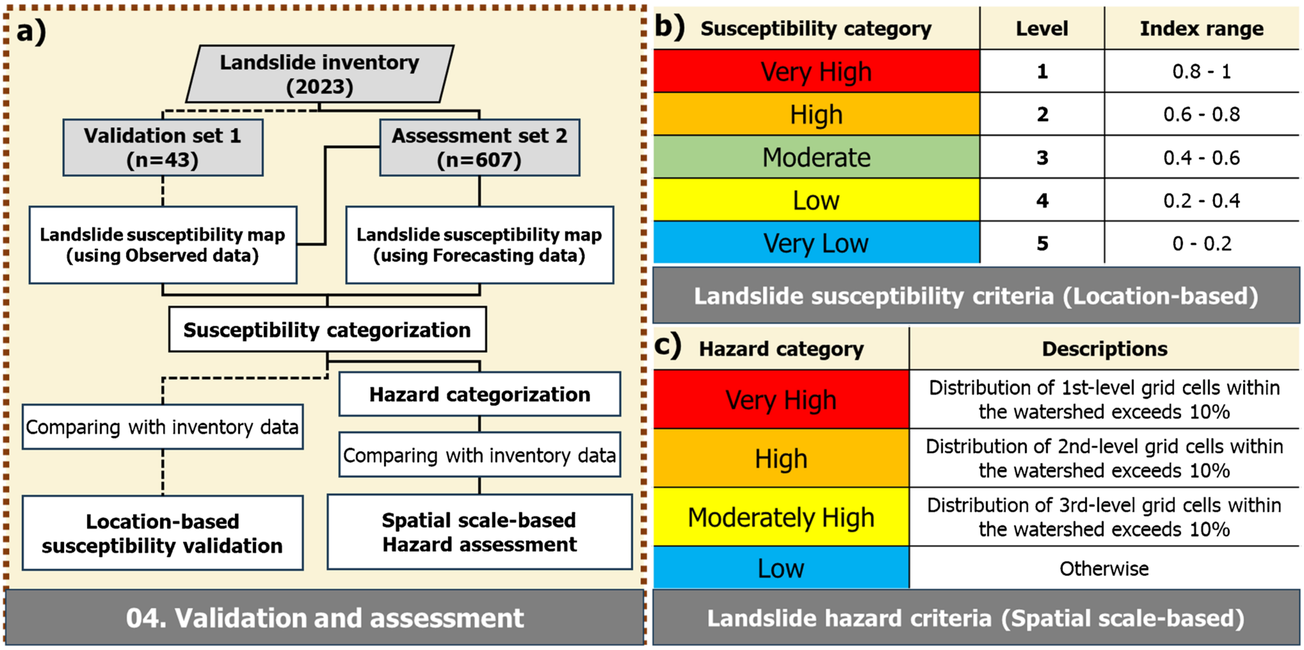


Fig. 5 **a** Workflow used in validation and assessment. **b** Susceptibility criterion for each index range. **c** Hazard criteria with descriptions

Table 3 Results of initial model selection by PyCaret

Model	Accuracy	AUC	Recall	Precision	F- 1
Random Forest Classifier	**0.929**	**0.979**	**0.923**	**0.872**	**0.897**
Extra Tree Classifier	0.926	0.981	0.899	0.883	0.890
Light Gradient Boosting Machine	0.925	0.976	0.921	0.864	0.891
Extreme Gradient Boosting	0.924	0.974	0.919	0.862	0.889
Gradient Boosting Classifier	0.921	0.972	0.919	0.855	0.886

Data in bold emphasis indicate the initial performance metrics of the Random Forest model

Table 4 Optimization results for Random Forest with hyperparameter descriptions

	Major hyperparameter	Value	Accuracy	AUC	Recall	Precision	F- 1
Optimized model	n_estimators	1150	0.930	0.981	0.931	0.869	0.898
	max_depth	70					
	Criterion	Entropy					
	max_feature	sqrt					
Hyperparameter description and candidates	n_estimators: maximum number of iterations			Min, 50; max, 2000; interval, 100 (default, 100)			
	max_depth: maximum tree depth			Min, 10; max, 110; interval, 10 (default, 0)			

category. The landslide events on July 14 and 15 accounted for approximately 83% of the moderately high and high categories, respectively.

However, the results of forecasting data at 8:00 AM indicated moderate or high for approximately 53% of the regions. The results suggest that landslide events on July 14 and 15 accounted

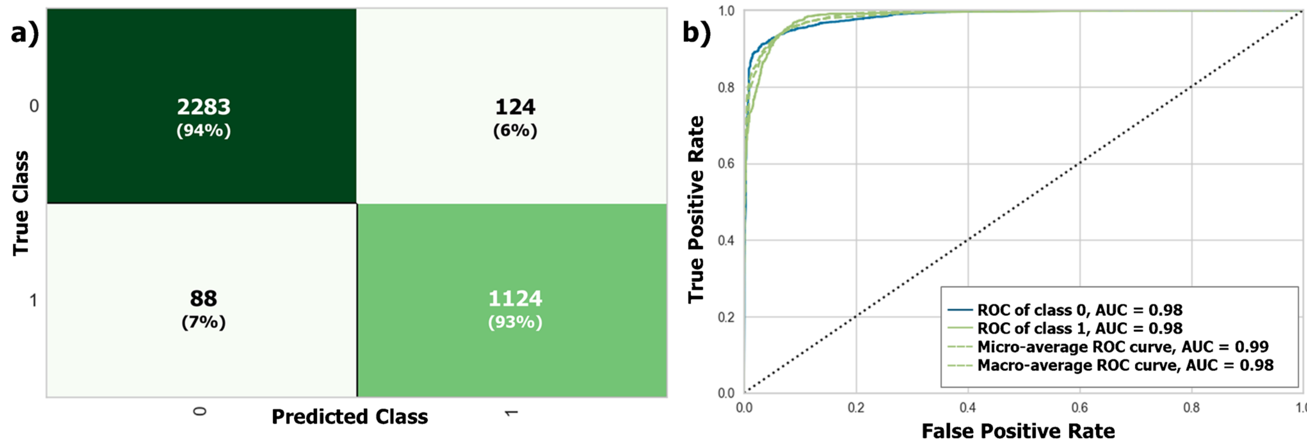


Fig. 6 **a** Confusion matrix and (**b**) ROC curves obtained for the landslide susceptibility model

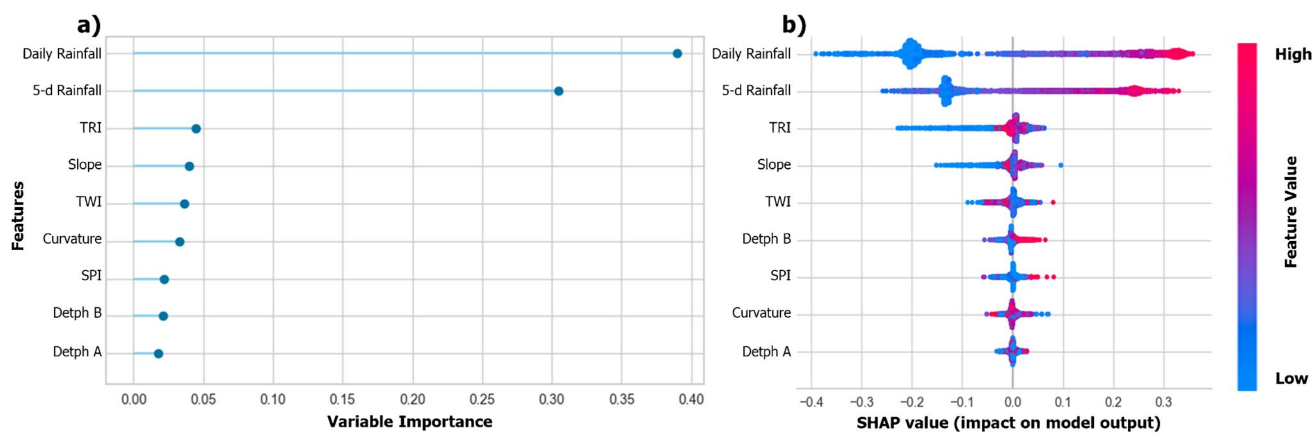


Fig. 7 Variable importance results for optimized landslide susceptibility model based on continuous variables: **a** feature importance in model and (**b**) SHAP summary plots

Table 5 Landslide damage history and early warning status in 2023

Index	Date	Number of occurrences	Forecasting	
			6:00 pm	9:00 am
1	2023.06.30	4	○	–
2	2023.07.11	2	○	○
3	2023.07.12	2	○	○
4	2023.07.14	307	○	○
5	2023.07.15	334	○	○
6	2023.07.16	7	○	○
7	2023.07.17	1	○	○
8	2023.07.18	21	○	○
9	2023.07.24	6	○	○
10	2023.08.10	84	○	○

for almost all percentages in the low-hazard category. These findings indicate a limitation of the 8:00 AM forecasting data, with the inactivity before 8:00 meaning that not all daily rainfall data are included.

Discussion

Comparison of the proposed approach with the current model

A comprehensive approach was implemented to develop a landslide susceptibility model and early warning process aimed at enhancing the prevention and preparedness efforts in South Korea. By integrating the model with early warning, daily landslide susceptibility maps were successfully computed for 2023. Landslide susceptibility model targeting in South Korea has generally focused solely on internal factors, such as geological and topographical variables, or local study areas with massive landslides, including smaller-scale regions and the capital (Hakim et al. 2022; Park & Lee 2021; Pradhan et al. 2023; Wang et al. 2023). The results indicate that the proposed model is capable of daily

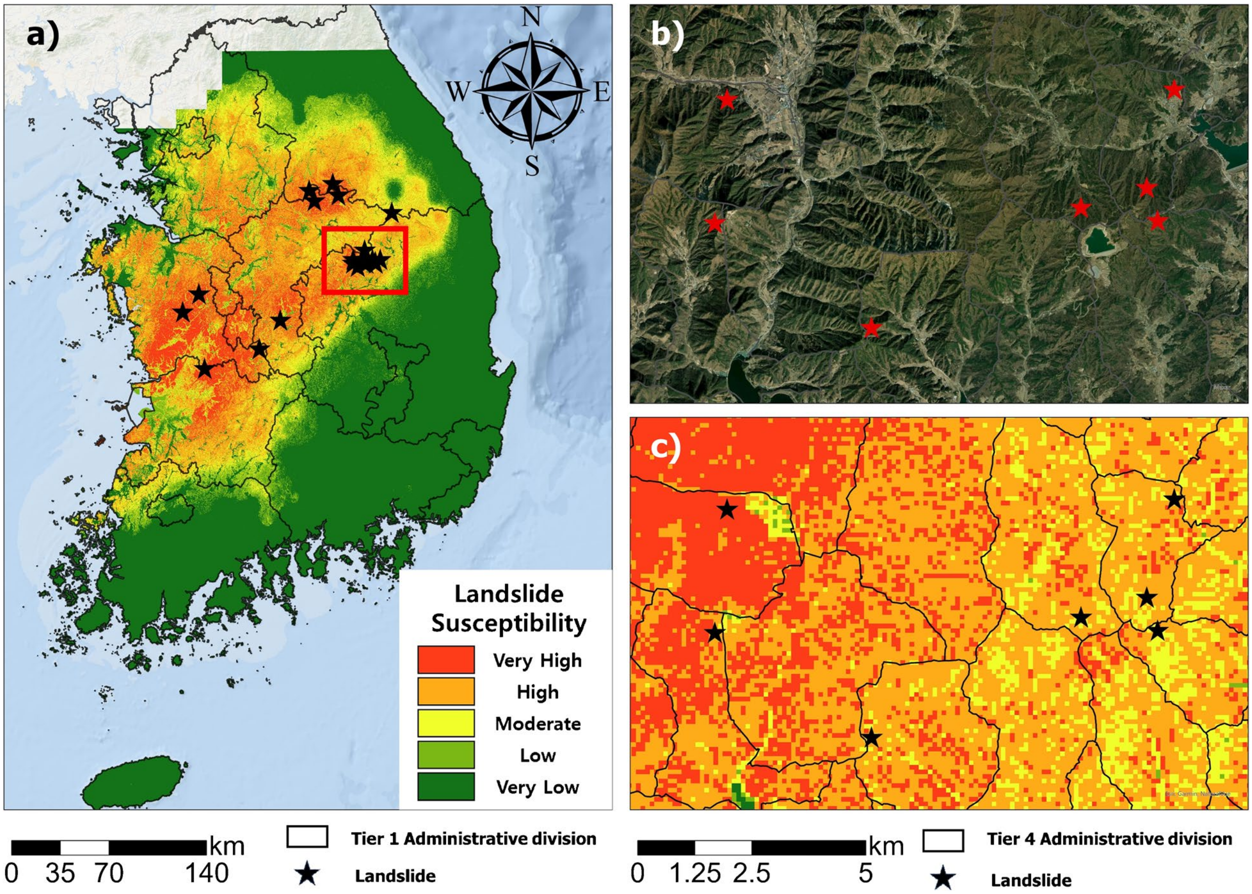


Fig. 8 Landslide susceptibility validation results for July 14, 2023: **a** landslide susceptibility map obtained using observed data, **b** Maxar image showing landslide events (sourced from ESRI in ArcPro basemap), and **c** landslide events on susceptibility map

monitoring with advanced spatiotemporal resolution that covers the entire territory of South Korea, particularly Tier 3 or 4 administrative divisions.

Model development executed using PyCaret indicated that Random Forest was most suitable for the landslide susceptibility model. The final, optimized model showed significant results, with an accuracy of 0.93, recall score of 0.93, and F- 1 score of 0.90, surpassing those of other machine learning-based models targeting South Korea (Kadavi et al. 2019; S. M. Lee & Lee 2024). According to the feature and permutation importance analyses, external factors exhibit a positive relationship with the occurrence of landslide occurrence, whereas moderate significance was indicated for internal factors such as TRI, slope, and TWI.

Application of the early warning process to the landslide susceptibility model allowed the generation of a daily landslide susceptibility map for 2023 on the OJERI website. Validation and assessment were conducted following aggregation of the landslide inventory for 2023. Location-based validation indicated that approximately 89% of the actual landslides were classified as high or very high susceptibility. The spatial-scale-based assessment also provided important insights.

Applicability to the early warning system

The results demonstrated significant accuracy in predicting actual landslide occurrences when observed rainfall data were used. The forecasting results at 5:00 PM also exhibited moderately significant performance. However, forecasting at 8:00 AM was limited because of insufficient rainfall data availability; the 8:00 AM forecast data includes rainfall from 08:00 to 24:00, and although this has the advantage of being more recent than the 5:00 PM data from the previous day, the mean rainfall obtained was generally lower than that obtained using 5:00 PM data (Fig. 12). The fact that the forecasting data at 8:00 AM is slightly lower than the observed data also indicates that the use of this data is limited when attempting to obtain the most recent daily rainfall data.

To address this issue, a simple calibration method was designed to compute complementary rainfall forecasting data, which applies cell statistics to both the day-before and morning forecasting data. The day-before forecasting data include complete rainfall information, whereas only recent rainfall trends and information are included in the morning forecasting data. The maximum function for both datasets was used to obtain the cell statistics. Calibrated rainfall data have the advantage of maintaining the spatial distributions of both datasets, and the mean calibrated daily rainfall was slightly higher (Fig. 12). However, the maximum value tends

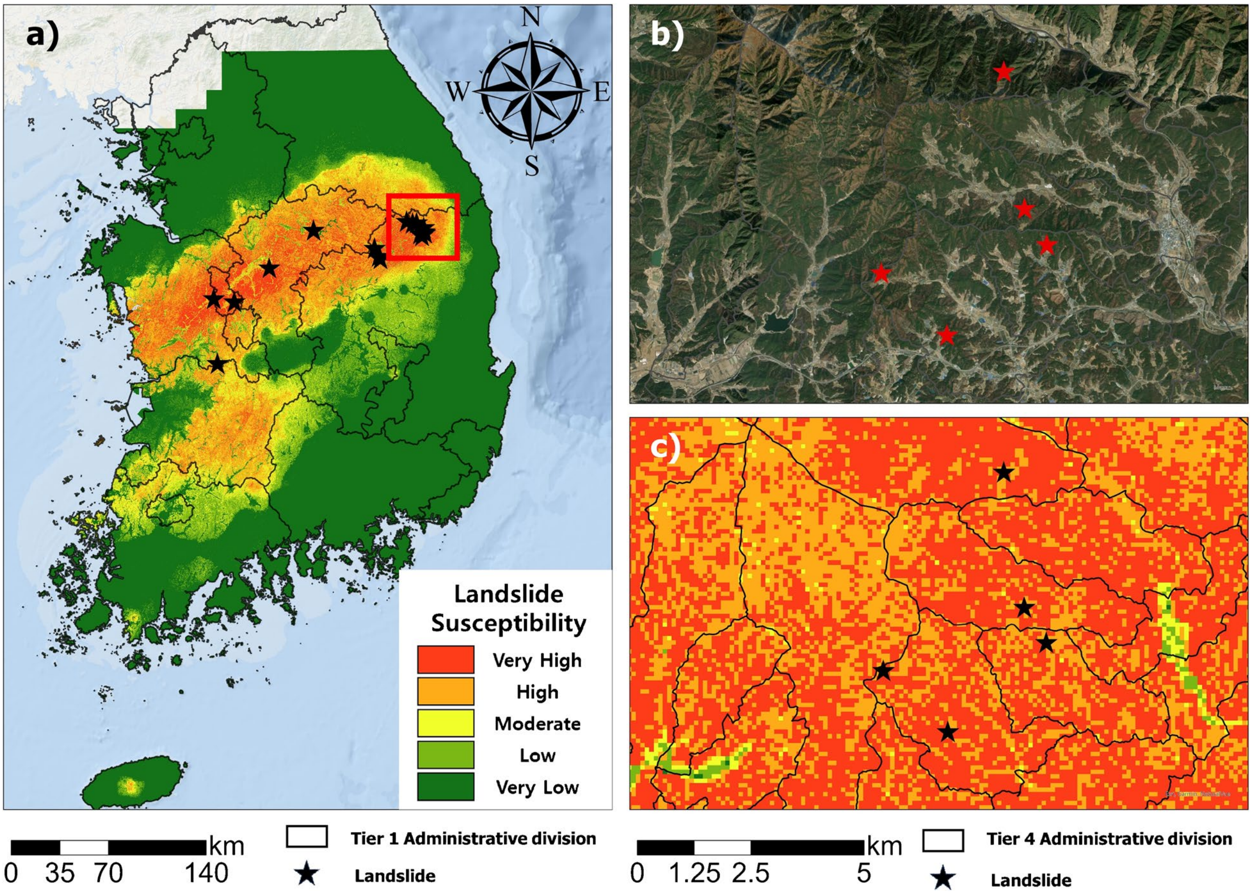


Fig. 9 Landslide susceptibility validation results for July 15, 2023: **a** landslide susceptibility map obtained using observed data, **b** Maxar image showing landslide events (sourced from ESRI in ArcPro basemap), and **c** landslide events on susceptibility map

to align with the observed data trends, apart from some outliers. Although the calibrated data were not the same as the observed data, comprehensive rainfall distribution was obtained from the spatial distribution. This finding suggests that the calibration method for forecasting data could mitigate against the aforementioned limitations.

Table 6 Location-based landslide susceptibility results

Susceptibility category	Level	Number of events	Mean/standard deviation of susceptibility index
Very high	1	15	0.88/0.05
High	2	23	0.69/0.06
Moderate	3	4	0.56/0.03
Low	4	1	0.36/0
Very low	5	0	-
Total		43	0.73/0.13

Implications for further studies

Landslide inventory data in South Korea may contain inaccurate coordinate information and occurrence dates, limiting their reliability and usefulness. In addition, the current sampling method for non-occurrences relies on simple temporal and spatial conditions, which can affect the representativeness of non-occurrence data. This lack of representativeness may lead to insufficient statistical differences between occurrences and non-occurrences, particularly for internal factors. It can also introduce high bias or variance in the model's training and prediction of landslide susceptibility. To address this, non-occurrences should be selected more strategically, for instance by identifying regions within watershed boundaries that share similar environmental conditions but have not experienced landslides. These issues were evident in our dataset. Labeled data for internal factors showed a smaller statistical difference than external factors (Table 8). Feature and permutation importance analyses revealed that internal factors had relatively lower significance than meteorological data, highlighting the need for higher-quality and more comprehensive landslide inventory data to enhance model robustness. In contrast, external factors demonstrated significant differences in labeled data, with meteorological data achieving the highest variable importance scores. For example, the spatial distribution

Table 7 Spatial scale-based landslide hazard region validation results

Hazard category	Forecasting (5:00 PM)		Forecasting (8:00 AM)		Observed data	
	Number of events	Ratio (%)	Number of events	Ratio (%)	Number of events	Ratio (%)
Very high	246	41	83	14	481	79
High	213	35	164	27	103	17
Moderately high	101	17	74	12	14	2
Low	47	7	286	47	9	2
Total	607	100	607	100	607	100

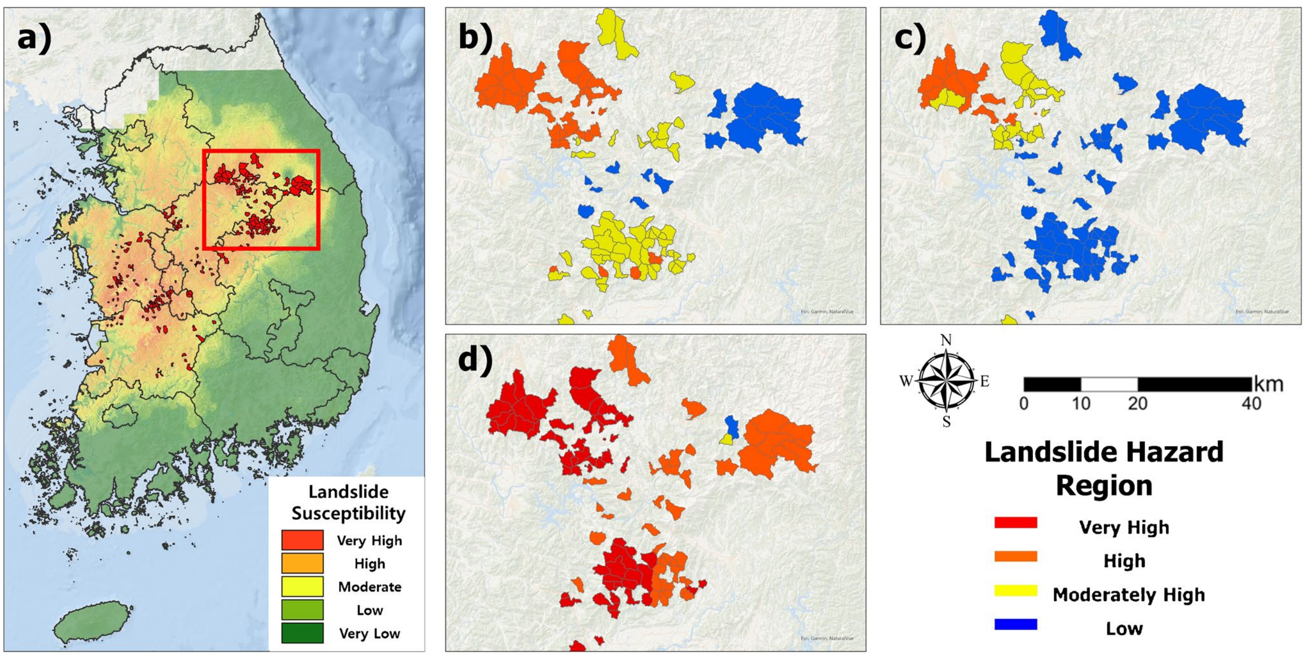


Fig. 10 Spatial scale-based hazard assessment results for July 14, 2023: **a** actual landslide occurrences (Tiers 3 and 4), **b** landslide hazard results for the studied region using forecasting data from 5:00 PM and **c** 8:00 AM, **d** landslide hazard region results using observed data

of susceptibility results closely aligned with rainfall patterns during validation and high-category regions corresponded to areas of intense rainfall. Taken together, these findings indicate that the current model setup may be overly dependent on external factors, especially meteorological data. To mitigate this dependency and improve overall reliability, it is essential to obtain more accurate and detailed inventory data and to adopt more refined sampling strategies for non-occurrences.

Regarding the early warning process, advancements in data acquisition are necessary. In this study, town weather forecasting data updated every 3 h starting at 2:00 AM were utilized. Forecasts for 8:00 AM and 5:00 PM were aggregated to produce calibrated data. However, to address temporal gaps and better capture spatiotemporal variations in weather forecasting, nowcasting data provided at 1-h intervals should be appropriately integrated. Incorporating hourly data would enable more frequent updates and improve the accuracy of short-term weather forecasting. These refinements could enhance the precision of landslide susceptibility assessments and contribute to a more reliable early warning process.

To advance beyond the current modeling and early warning process, it is necessary to integrate annual land cover change maps and population density data at the Tier 4 administrative divisions into the early warning process. Annual land cover changes have increased along the forest boundaries in South Korea, such as the conversion of forested areas into solar panels or orchards, posing serious landslide threats to nearby residents.

This study focused on rainfall-induced landslide susceptibility across the entire country of South Korea using a statistical-based approach. However, integrating physical models that incorporate factors such as groundwater depth and soil moisture indices could enable real-time monitoring of critical regions by

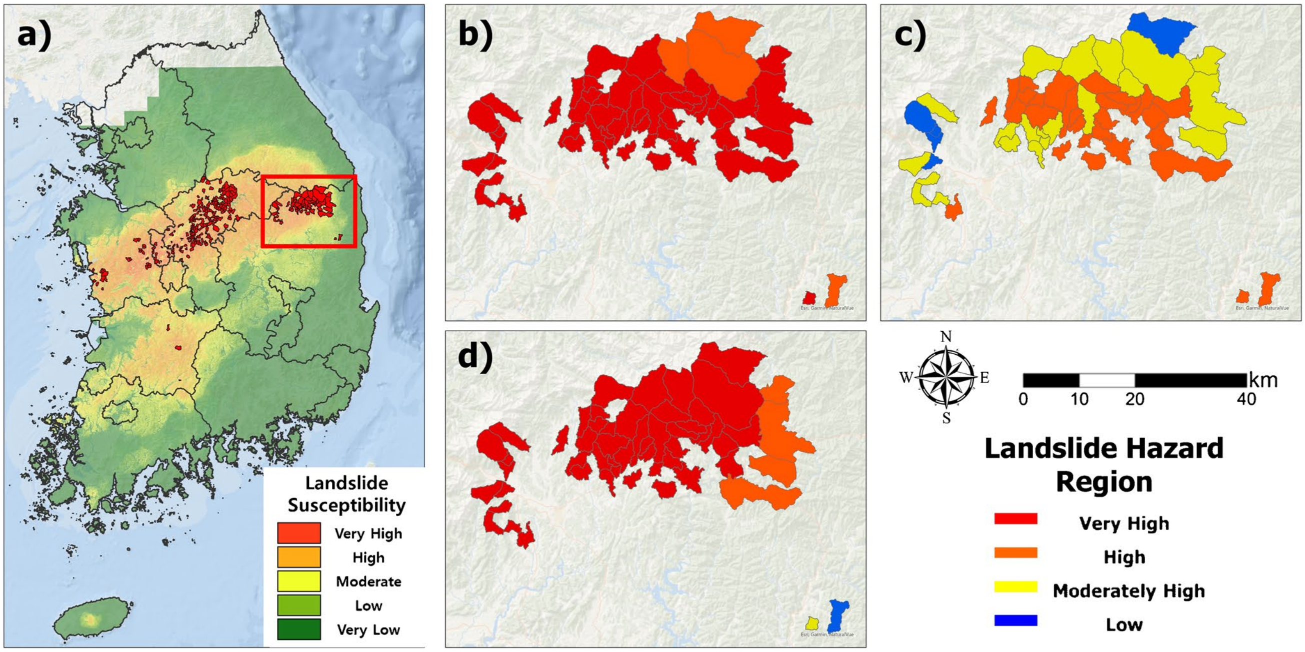


Fig. 11 Spatial scale-based hazard assessment results for July 15, 2023: **a** actual landslide occurrences (Tiers 3 and 4), **b** landslide hazard results for the studied region using forecasting data from 5:00 PM and **c** 8:00 AM, **d** landslide hazard region results using observed data

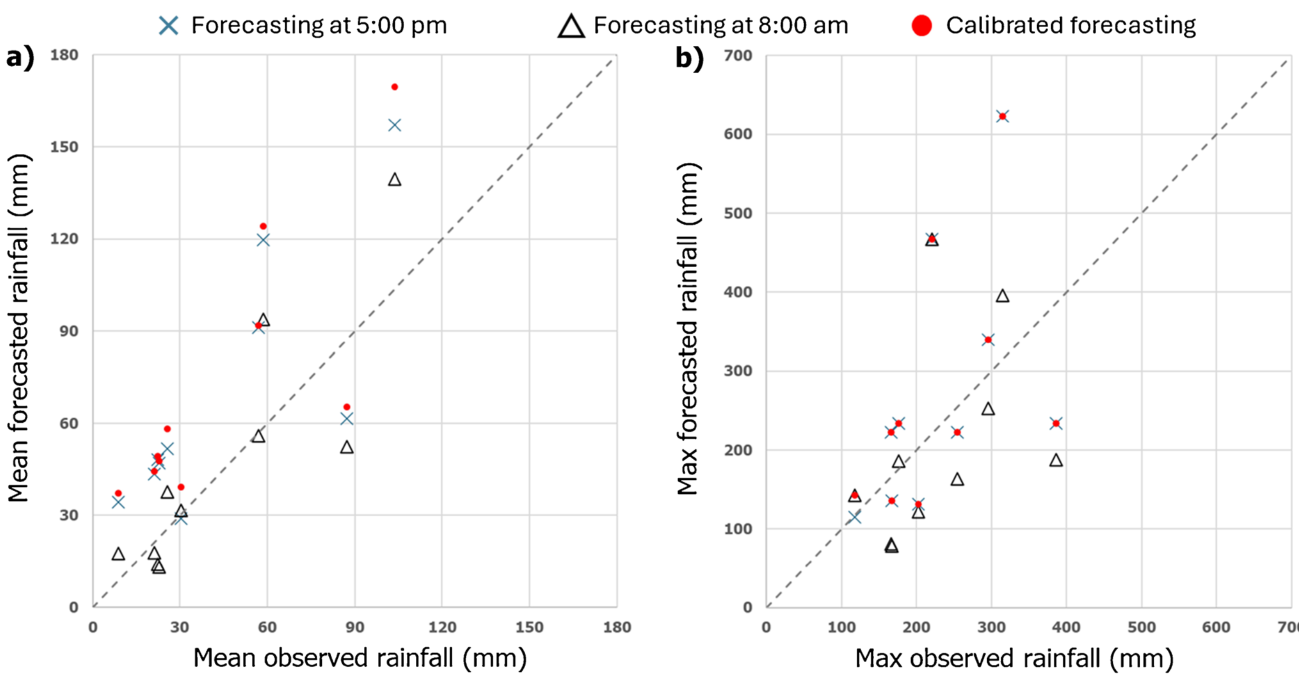


Fig. 12 **a** Mean comparison and (**b**) max comparison of three types of daily forecasted rainfall with observed rainfall during the validation period in 2023

reflecting dynamic hydrological changes. For instance, real-time groundwater monitoring can provide timely alerts on changing subsurface conditions that indicate the potential for landslides. Additionally, coupling the statistical-based susceptibility model with debris flow models would enhance precision in identifying high-susceptibility areas and predicting the extent of potential damage during landslide events. This combined approach would facilitate more refined and reliable landslide hazard assessments,

Table 8 Descriptive table of labeling dataset for continuous data

	Label 0 (non-occurrences)				Label 1 (occurrences)			
	Mean	Standard deviation	Skewness	Count	Mean	Standard deviation	Skewness	Count
Daily rainfall	28.51393	46.48307	2.190328	7,821	177.2725	68.92661	0.060316	4041
5-day cumulative rainfall	72.3078	83.34647	1.709713		277.6858	102.8667	0.663938	
Slope	11.14623	7.27883	0.578596		11.15703	6.309346	0.648028	
Curvature	0.003946	0.355036	0.230202		− 0.00063	0.347895	0.119984	
Flow direction	4.338448	2.265304	0.006458		4.358575	2.303862	− 0.01872	
TRI	60.48009	35.22989	0.466909		60.09793	30.21611	0.619617	
TWI	7.658859	2.166141	2.088068		7.412664	1.796901	1.986022	
SPI	1242.076	42,469.55	77.16831		459.6342	5075.304	38.26775	
Soil depth A	9.001023	8.956944	0.516436		9.057906	8.900696	0.549544	
Soil depth B	31.20176	28.71493	0.150764		32.17248	28.96691	0.126342	

thereby contributing to proactive disaster prevention and mitigation efforts.

Since 2022, the landslide susceptibility model has been updated annually. Specifically, updates focus on the input data and model training processes. Input data updates include the renewal of labeling data and internal factors such as land cover status. The renewed labeling data is then integrated as training data. Following the development of a new susceptibility model based on the random forest algorithm, a daily early warning process is implemented during the summer seasons, validated by real-time occurrences.

This study demonstrates the feasibility of developing a landslide susceptibility model and generating a nationwide landslide hazard map by integrating it with a daily early warning process based on weather forecasting. Although there may be some limitations in utilizing the results for public announcements, the produced warning results are valuable to citizens. Moreover, since our results are at a 100-m spatial resolution, they are suitable for community- and regional-level prevention and preparedness efforts (Highland & Bobrowsky 2008) by administrative bodies in Tier 3 or 4 administrative divisions and watershed areas. Notably, the data used in our study were sourced from public government datasets and open-source platforms. The internal factors can be substituted with open geospatial information such as FAO Harmonized Soil Data, ESA WorldCover, and NASA SRTM data. The model operation can be integrated with the early warning process as an individual module in Python. Given the availability of proper landslide inventory and short-term early warning data, this method can be applied globally, particularly in mid-latitude countries with similar meteorological and geological characteristics. However, applying this model in other regions or countries may present challenges. The model's performance heavily relies on the quality and quantity of input data, including landslide inventories and weather forecasts. Therefore, careful acquisition of conditioning factors and region-specific adjustments is essential for applying this model to other regions. Despite these limitations, the framework presented in this study remains flexible and scalable, allowing for adaptation to varying environmental conditions and data availability.

Conclusions

This study developed a landslide susceptibility model and short-term early warning process. This method can produce nationwide daily landslide susceptibility maps at a 100-m resolution and provide up to 3 days of early warning information. A total of 4041 landslide inventory data points, along with corresponding non-occurrence data, were combined using a spatiotemporal random sampling method. Thirteen landslide conditioning factors were considered at a 100-m resolution. Random Forest was identified as the best-performing model, achieving an accuracy of 0.9298, AUC of 0.9809, and F- 1 of 0.9894. The early warning results for 2023 using weather forecasting data demonstrated promising outcomes: 88% of location-based occurrences and 96% of Tier 3 or 4 administrative divisions were classified as high category. However, accuracy varied with forecast timing, achieving 76% for the 5:00 PM forecast and dropping to 41% for the 8:00 AM forecast. This variability may be attributed to differences in rainfall patterns and the timeliness of data integration, underscoring the need to incorporate nowcasting or observed rainfall data to enhance accuracy further. While the current model shows strong performance, additional improvements are needed. These include enhancing the quality of landslide inventory data, refining non-occurrence sampling methods to represent areas without landslide occurrence better, and integrating remote sensing and socio-economic data to capture a broader range of influencing factors. The modular structure of the susceptibility model and early warning process allows for continuous updates and the integration of state-of-the-art technologies. This ensures that the system remains up-to-date and can adapt to new data and methodologies. In conclusion, this study provides a solid foundation for landslide prevention and preparedness in South Korea. By harmonizing the daily susceptibility model with the early warning process, timely and accurate hazard information can be delivered to citizens and administrative divisions, contributing to proactive landslide management.

Acknowledgements

This study was conducted with the support of the R&D Program for Forest Science Technology (project No. 2021345B10 - 2323-CD01) provided by the Korea Forest Service (Korea Forestry Promotion Institute) and the Green Society Program funded by Hyundai Motor Chung Mong-Koo Foundation.

Funding

This research was supported by Korea Environment Industry & Technology Institute (KEITI) through the "Climate Change R&D Project for New Climate Regime (RS- 2022-KE002294)," funded by the Korea Ministry of Environment (ME).

Data availability

All data used in this work are either publicly available or available from the authors upon reasonable request.

Declarations

Conflict of interest The authors declare no competing interests.

References

Abellán J, Mantas CJ, Castellano JG, Moral-García S (2018) Increasing diversity in random forest learning algorithm via imprecise probabilities. Expert Syst Appl 97:228–243. https://doi.org/10.1016/J.ESWA.2017.12.029

Achu AL, Thomas J, Aju CD, Remani PK, Gopinath G (2023) Performance evaluation of machine learning and statistical techniques for modelling landslide susceptibility with limited field data. Earth Sci Inf 16(1):1025–1039

Ali M (2020) PyCaret: an open source, low-code machine learning library in Python. PyCaret version 2

Atkinson PM, Massari R (1998) Generalised linear modelling of susceptibility to landsliding in the central apennines. Italy Comput Geosci 24(4):373–385. https://doi.org/10.1016/S0098-3004(97)00117-9

Azarafza M, Azarafza M, Akgün H, Atkinson PM, Derakhshani R (2021) Deep learning-based landslide susceptibility mapping. Sci Rep 11(1):24112

Ballabio C, Sterlacchini S (2012) Support vector machines for landslide susceptibility mapping: the Staffora River basin case study. Italy Math Geosci 44(1):47–70. https://doi.org/10.1007/S11004-011-9379-9/METRICS

Beven KJ, Kirkby MJ (1979) A physically based, variable contributing area model of basin hydrology. Hydrol Sci Bull 24(1):43–69. https://doi.org/10.1080/02626667909491834

Brand, E. W. (1985) Predicting the performance of residual soil slopes. In Proceedings 11th Int. Conf. Soil Mech. & Found. Engineering. San Francisco (Vol. 5, pp. 2541–2578).

Breiman L (2001) Random forests. Mach Learn 45(1):5–32. https://doi.org/10.1023/A:1010933404324

Caine N (1980) The rainfall intensity - duration control of shallow landslides and debris flows. Geografiska Annaler: Series A, Phys. Geogr. 62(1–2):23–27. https://doi.org/10.1080/04353676.1980.11879996

Carrara A, Guzzetti F, Cardinali M, Reichenbach P (1999) Use of GIS technology in the prediction and monitoring of landslide hazard. Nat Hazards 20(2–3):117–135. https://doi.org/10.1023/A:1008097111310/METRICS

Chauhan K, Jani S, Thakkar D, Dave R, Bhatia J, Tanwar S, Obaidat MS (2020) Automated machine learning: the new wave of machine learning. 2nd International Conference on Innovative Mechanisms for Industry Applications, ICIMIA 2020 - Conference Proceedings, 205–212. https://doi.org/10.1109/ICIMIA48430.2020.9074859

Choe G (2001) Current status and causes of landslides in South Korea. Magazine of the Korean Society of Hazard Mitigation 1(3):7–14 (**(in Korean)**)

Choi K (1986) Landslides occurrence and its prediction in Korea. Ph. D. Dissertation (in Korean with English abstract), Kangwon National University

Ciurleo M, Mandaglio MC, Moraci N (2019) Landslide susceptibility assessment by TRIGRS in a frequently affected shallow instability area. Landslides 16(1):175–188. https://doi.org/10.1007/S10346-018-1072-3/FIGURES/8

Cui H, Ji J, Hürlimann M, Medina V (2024) Probabilistic and physically-based modelling of rainfall-induced landslide susceptibility using integrated GIS-FORM algorithm. Landslides 21(6):1461–1481

Dai FC, Lee CF (2001) Frequency–volume relation and prediction of rainfall-induced landslides. Eng Geol 59(3–4):253–266

Dai FC, Lee CF, Ngai YY (2002) Landslide risk assessment and management: an overview. Eng Geol 64(1):65–87. https://doi.org/10.1016/S0013-7952(01)00093-X

Dikshit A, Satyam N, Pradhan B (2019) Estimation of rainfall-induced landslides using the TRIGRS model. Earth Systems and Environment 3(3):575–584. https://doi.org/10.1007/S41748-019-00125-W/FIGURES/9

Dou J, Yunus AP, Bui DT, Merghadi A, Sahana M, Zhu Z, 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(3):641–658. https://doi.org/10.1007/S10346-019-01286-5/TABLES/6

Federici B, Bovolenta R, Passalacqua R (2015) From rainfall to slope instability: an automatic GIS procedure for susceptibility analyses over wide areas. Geomat Nat Haz Risk 6(5–7):454–472. https://doi.org/10.1080/19475705.2013.877087

Froude MJ, Petley DN (2018) Global fatal landslide occurrence from 2004 to 2016. Nat/ Hazard Earth Sys 18(8):2161–2181. https://doi.org/10.5194/NHESS-18-2161-2018

Gariano SL, Guzzetti F (2016) Landslides in a changing climate. Earth Sci Rev 162:227–252. https://doi.org/10.1016/J.EARSCIREV.2016.08.011

Geertsema M, Highland L, Vaugeouis L (2009) Environmental impact of landslides. Landslides – disaster risk reduction 589–607. https://doi.org/10.1007/978-3-540-69970-5_31

Guzzetti F, Peruccacci S, Rossi M, Stark CP (2008) The rainfall intensity-duration control of shallow landslides and debris flows: an update. Landslides 5(1):3–17. https://doi.org/10.1007/S10346-007-0112-1/FIGURES/8

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 Icheon. South Korea

J Environ Manage 305:114367. https://doi.org/10.1016/J.JENVMAN.2021.114367
Ham D, Hazard SH-J (2014) Review of landslide forecast standard suitability by analysing landslide-inducing rainfall. Journal of the Korean Society of Hazard Mitigation 14(3):299–310. https://doi.org/10.9798/KOSHAM.2014.14.3.299
Haque U, da Silva PF, Devoli G, Pilz J, Zhao B, Khaloua A, Wilopo W, Andersen P, Lu P, Lee J, Yamamoto T, Keellings D, Jian-Hong W, Glass GE (2019) The human cost of global warming: deadly landslides and their triggers (1995–2014). Sci Total Environ 682:673–684. https://doi.org/10.1016/J.SCITOTENV.2019.03.415
Hemasinghe H, Rangali RSS, Deshapriya NL, Samarakoon L (2018) Landslide susceptibility mapping using logistic regression model (a case study in Badulla District, Sri Lanka). Procedia Engineering 212:1046–1053. https://doi.org/10.1016/J.PROENG.2018.01.135
Highland L, Bobrowsky P (2008) The landslide handbook - a guide to understanding landslides. https://pubs.usgs.gov/circ/1325/
Huang PC (2023) Establishing a shallow-landslide prediction method by using machine-learning techniques based on the physics-based calculation of soil slope stability. Landslides 20(12):2741–2756
Hussain MA, Chen Z, Zheng Y, Zhou Y, Daud H (2023) Deep learning and machine learning models for landslide susceptibility mapping with remote sensing data. Remote Sensing 15(19):4703
Hutter F, Kotthoff L, Vanschoren J (2019) Automated machine learning. 219. https://doi.org/10.1007/978-3-030-05318-5
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(4):1–17. https://doi.org/10.1007/S12665-019-8119-1/TABLES/5
Kang WS, Ma HS, Jeon KS (2016) Influences of cumulative number of days of rainfall on occurrence of landslide. Journal of Korean Society of Forest Science 105(2):216–222 (**(in Korean)**)
Kanti Karmaker S, Hassan M, Smith MJ, Mahadi Hassan M, Xu L, Zhai C, Veeramachaneni K, Karmaker SK, Hassan MM, Ginn S, Smith MJ, Xu L, Veeramachaneni K, Zhai C (2021) AutoML to date and beyond: challenges and opportunities. ACM Comput Surv 54(8):175. https://doi.org/10.1145/3470918
Kim GH (2013) Cover story-the role of forest science and technology in preparing for and mitigating mountainous natural disasters. Disaster Prevention Review 15(3):48–55 (**(in Korean)**)
Kim WY, Chae BG (2009) Characteristics of rainfall, geology and failure geometry of the landslide areas on natural terrains. Korea the Journal of Engineering Geology 19(3):331–344 (**(in Korean)**)
Korea Forest Service (2021) Understanding landslides properly. https://sansatai.forest.go.kr/ . Accessed 1 May 2023. (in Korean)
Korea Forest Service (2023) The comprehensive plan for nationwide landslide prevention in 2023. https://sansatai.forest.go.kr/ . Accessed 1 Aug 2023. (in Korean)
Korea Forest Service (2024a) Landslide prevention sector implementation plan. https://sansatai.forest.go.kr/ . Accessed 30 Jan 2024. (in Korean)
Korea Forest Service (2024b) The comprehensive plan for nationwide landslide prevention in 2024. https://sansatai.forest.go.kr/ . Accessed 30 Apr 2024. (in Korean)
Korea Institute of Geoscience and Mineral Resources (2019) Landslide early warning and risk control technology of geo-environmental hazards for climate change adaptation (GP2017–017–2019). Ministry of Science and ICT, 455 p (in Korean with English Summary).
Korea Meteorological Administration (2023) Meteorological administration short-term forecast inquiry service open API utilization guide. (in Korean)
Korea Meteorological Administration (2024) 2023 Weather YearBook. Accessed 30 Apr 2024. https://www.kma.go.kr/kma/. Accessed 1 Apr 2024. (in Korean)
Lee SM, Lee SJ (2024) Landslide susceptibility assessment of South Korea using stacking ensemble machine learning. Geoenvironmental Disasters 11(1):1–17. https://doi.org/10.1186/S40677-024-00271-Y/FIGURES/7
Lee S, Choi J, Min K (2002) Landslide susceptibility analysis and verification using the Bayesian probability model. Environ Geol 43(1–2):120–131. https://doi.org/10.1007/S00254-002-0616-X/METRICS
Lee JS, Kim YT, Song YK, Jang DH (2014) Landslide triggering rainfall threshold based on landslide type. Journal of the Korean Geotechnical Society 30(12):5–14 (**(in Korean)**)
Lee C, Kim D, Woo C, Kim YS, Seo J, Kwon H (2015) Construction and operation of the national landslide forecast system using soil water index in Republic of Korea. J Korean Soc Hazard Mitig 15(6):213–221 (**(in Korean)**)
Lee S, Hong SM, Jung HS (2017) a support vector machine for landslide susceptibility mapping in Gangwon Province. Korea Sustainability 9(1):48. https://doi.org/10.3390/SU9010048
Lee WY, Park SK, Sung HH (2021) The optimal rainfall thresholds and probabilistic rainfall conditions for a landslide early warning system for Chuncheon. Republic of Korea Landslides 18(5):1721–1739. https://doi.org/10.1007/S10346-020-01603-3
Liu J, Fan X, Tang X, Xu Q, Harvey EL, Hales TC, Jin Z (2022) Ecosystem carbon stock loss after a mega earthquake. CATENA 216(A):106393. https://doi.org/10.1016/J.CATENA.2022.106393
Lundberg SM, Allen PG, Lee S-I (2017) A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, 30. https://github.com/slundberg/shap
Ma J, Lei D, Ren Z, Tan C, Xia D, Guo H (2024) Automated machine learning-based landslide susceptibility mapping for the Three Gorges Reservoir Area. China Mathematical Geosciences 56(5):975–1010
Moore ID, Grayson RB, Ladson AR (1991) Digital terrain modelling: a review of hydrological, geomorphological, and biological applications. Hydrol Process 5(1):3–30. https://doi.org/10.1002/HYP.3360050103
Nadim F, Kjekstad O, Peduzzi P, Herold C, Jaedicke C (2006) Global landslide and avalanche hotspots. Landslides 3(2):159–173. https://doi.org/10.1007/S10346-006-0036-1/TABLES/12
National Geographic Information Institute (2015) Toponymic guidelines for maps and other editors for international use (2nd ed.). http://www.ngii.go.kr/en. Accessed 15 Jan 2024. (in Korean)
National Geographic Information Institute (2020) The national atlas of Korea II 2020. http://nationalatlas.ngii.go.kr/. Accessed 15 Jan 2024. (in Korean)
Ng CWW, Yang B, Liu ZQ, Kwan JSH, Chen L (2021) Spatiotemporal modelling of rainfall-induced landslides using machine learning. Landslides 18(7):2499–2514. https://doi.org/10.1007/S10346-021-01662-0/FIGURES/11
Nikoobakht S, Azarafza M, Akgün H, Derakhshani R (2022) Landslide susceptibility assessment by using convolutional neural network. Appl Sci 12(12):5992
Park S, Kim J (2019) Landslide susceptibility mapping based on random forest and boosted regression tree models, and a comparison of their performance. Appl Sci 9(5):942. https://doi.org/10.3390/APP9050942
Park S-J, Lee D-K (2021) Predicting susceptibility to landslides under climate change impacts in metropolitan areas of South Korea using machine learning. Geomat Nat Haz Risk 12(1):2462–2476. https://doi.org/10.1080/19475705.2021.1963328
Park DW, Nikhil NV, Lee SR (2013) Landslide and debris flow susceptibility zonation using TRIGRS for the 2011 Seoul landslide event. Nat Hazard Earth Sys 13(11):2833–2849. https://doi.org/10.5194/NHESS-13-2833-2013
Passalacqua R, Bovolenta R, Federici B, Balestrero D (2016) A physical model to assess landslide susceptibility on large areas: recent developments and next improvements. Procedia Engineering 158:487–492. https://doi.org/10.1016/J.PROENG.2016.08.477
Petley D (2012) Global patterns of loss of life from landslides. Geology 40(10):927–930. https://doi.org/10.1130/G33217.1
Pradhan B, Dikshit A, Lee S, Kim H (2023) An explainable AI (XAI) model for landslide susceptibility modeling. Appl Soft Comput 142:110324. https://doi.org/10.1016/J.ASOC.2023.110324
Qin C-Z, Zhu A-X, Pei T, Bao L-L, Scholten T, Behrens T, Zhou C-H (2011) An approach to computing topographic wetness index based on maximum downslope gradient. Precision Agric 12:32–43. https://doi.org/10.1007/s11119-009-9152-y

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. https://doi.org/10.1016/J.EARSCIREV.2018.03.001
Ren T, Gao L, Gong W (2024) An ensemble of dynamic rainfall index and machine learning method for spatiotemporal landslide susceptibility modeling. Landslides 21(2):257–273. https://doi.org/10.1007/S10346-023-02152-1/FIGURES/14
Riley S, DeGloria S, Elliot R (1999) Index that quantifies topographic heterogeneity. Download.Osgeo.Org. Retrieved May 13, 2024, from http://download.osgeo.org/qgis/doc/reference-docs/Terrain_Ruggedness_Index.pdf
Sahin EK (2020) Assessing the predictive capability of ensemble tree methods for landslide susceptibility mapping using XGBoost, gradient boosting machine, and random forest. SN Applied Sciences 2(7):1–17. https://doi.org/10.1007/S42452-020-3060-1/TABLES/1
Sarangpure N, Dhamde V, Roge A, Doye J, Patle S, Tamboli S (2023) Automating the machine learning process using PyCaret and Streamlit. 2023 2nd International Conference for Innovation in Technology, INOCON 2023. https://doi.org/10.1109/INOCON57975.2023.10101357
Sim KB, Lee ML, Wong SY (2022) A review of landslide acceptable risk and tolerable risk. Geoenvironmental Disasters 9(1):1–17. https://doi.org/10.1186/S40677-022-00205-6/FIGURES/10
Song Y, Division GH, Resources M (2022) State-of-the-art on development and operation of landslide early warning system for climate change response. J Geol Soc Korea 4036(4):509–525
Spiekermann RI, van Zadelhoff F, Schindler J, Smith H, Phillips C, Schwarz M (2023) Comparing physical and statistical landslide susceptibility models at the scale of individual trees. Geomorphology 440:108870. https://doi.org/10.1016/J.GEOMORPH.2023.108870
Sujatha ER, Kumaravel P, Rajamanickam GV (2014) Assessing landslide susceptibility using Bayesian probability-based weight of evidence model. B Eng Geo Environ 73(1):147–161. https://doi.org/10.1007/S10064-013-0537-9/TABLES/5
Tang G, Fang Z, Wang Y (2023) Global landslide susceptibility prediction based on the automated machine learning (AutoML) framework. Geocarto Int 38(1):2236576
United Nations Office for Disaster Risk Reduction (2016) Report of the open-ended intergovernmental expert working group on indicators and terminology relating to disaster risk reduction
Van Westen CJ, Van Asch TW, Soeters R (2006) Landslide hazard and risk zonation—why is it still so difficult? Bull Eng Geol Env 65:167–184
Van Westen CJ, Castellanos E, Kuriakose SL (2008) Spatial data for landslide susceptibility, hazard, and vulnerability assessment: an overview. Eng Geol 102(3–4):112–131
Wang L, Wang Y, Xiao T, Liu Z, Kim J-C, Lee S (2023) Comparative study of deep neural networks for landslide susceptibility assessment: a case study of Pyeongchang-gun. South Korea Sustainability 16(1):245. https://doi.org/10.3390/SU16010245
Wieczorek GF (1996) Landslide triggering mechanisms. In: Turner AK, Schuster RL (eds) Landslides: investigation and mitigation. Transportation Research Board, National Research Council, Washington DC, pp. 76–90
Yang L, Cui Y, Xu C, Ma S (2024) Application of coupling physics–based model TRIGRS with random forest in rainfall-induced landslide-susceptibility assessment. Landslides, 1–15.
Zhang K, Wu X, Niu R, Yang K, Zhao L (2017) The assessment of landslide susceptibility mapping using random forest and decision tree methods in the Three Gorges Reservoir area. China Environ Earth Sci 76(11):1–20. https://doi.org/10.1007/S12665-017-6731-5/TABLES/8
Zhu L, Huang JF (2006) GIS-based logistic regression method for landslide susceptibility mapping in regional scale. J Zhejiang Univ Sci 7(12):2007–2017. https://doi.org/10.1631/JZUS.2006.A2007/METRICS

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

Sujong Lee · Minwoo Roh · Woo-Kyun Lee (✉)
Department of Environmental Science and Ecological Engineering, Korea University, Seoul 02841, Republic of Korea
Email: leewk@korea.ac.kr

Hyun-Woo Jo · Kim Joon · Woo-Kyun Lee
OJEong Resilience Institute (OJERI), Korea University, Seoul 02841, Republic of Korea

Hyun-Woo Jo
Agriculture Forestry and Ecosystem Services (AFE) Group, Biodiversity and Natural Resources (BNR) Program, International Institute for Applied Systems Analysis (IIASA), Schlossplatz 1, A- 2361 Laxenburg, Austria

Woo-Kyun Lee
Division of Environmental Science and Ecological Engineering, Korea University, Seoul 02841, Republic of Korea