

## Manuscript number: nhess-2024-90

My co-authors and I would like to express our gratitude to the reviewer for his constructive feedback and suggestions for strengthening our research. The changes we have made to the attached file in response to such feedback and suggestions have been highlighted in blue to facilitate their identification. I would also like to offer my apologies for the length of time it took us to prepare this response. We also record our deep appreciation for the efficient handling of the manuscript.

### Response to Reviewer#1

**General remarks:** I am attaching my full comments in the attached PDF. At the same time, I am summarizing my general comments here for the editor's perusal.

This manuscript presents a valuable reflection of data-driven modelling for robust regional-scale analyses of landslide masses. The authors deserve commendation for their interesting research, which has significant implications for hazard prediction and modelling. However, I have some major comments and concerns. While the study is promising and of great interest to the landslide community, it requires further work. Some aspects of the training and testing regimes are not clear. Furthermore, the choice of certain parameters is not well justified which, in my opinion, must be clarified for readers to understand the logic of choosing said parameters. The English language, particularly in the Introduction, needs improvement. Some sentences read awkwardly and are hard to follow. Improved sentence phrasing is necessary to make the manuscript clearer, especially for non-native English readers. In my opinion, a major revision is required to adapt the manuscript before considering acceptance.

**Response:** Thank you for your detailed comments and for the recognition of the value of our research. We appreciate your commendation and acknowledge the importance of addressing your highlighted concerns. In the revised manuscript, we have focused on the aspects of the training and testing datasets to enhance understanding, as well as provide a stronger justification regarding the choice of predictor variables to ensure the logic is clear to all readers. Additionally, we revised the language throughout the entire manuscript to enhanced readability.

**General comments:** The problem of landslide volume estimation has been a focus for the community for quite some time, through methods such as area-volume scaling, geometrical modelling, numerical simulations, and more. This parameter is crucial as it helps gauge the magnitude of landslides, particularly at regional scales. Most highly accurate methods, like numerical simulations, often struggle at the regional scale. This manuscript offers a valuable reflection of data-driven modelling for delivering robust regional-scale analyses of landslide masses. Kudos to the authors

for this interesting research, which has significant implications for hazard prediction and modelling. However, there are some major comments and curiosities I have. I believe the study is promising and of great interest to the landslide community, but it requires further work. The English language writing can be improved, especially in the Introduction. Some sentences read awkwardly and are hard to follow. Sentence phrasing must be improved to make the manuscript clearer, particularly for non-native English readers.

**Response:** We appreciate the thoughtful feedback and for recognizing the value of our research in the context of landslide volume prediction and acknowledgment of the challenges faced by highly accurate methods at regional scales, and we appreciate that our data-driven modeling approach resonates with the landslide community. We took your concerns regarding the clarity of the English language and improved the phrasing and overall readability, particularly in the Introduction, to ensure it is accessible to all readers.

#### **Specific major comments:**

Comment 1: The Introduction needs to be revisited for editing in both grammar and phrasing of the language. Moreover, the motivation for the importance of volume quantification appears to be a bit lacklustre. I do not see a geomorphological connection as to why volume estimates are important to understand process mechanism and kinematics. Although, the manuscript does not explore said mechanism and kinematics expressions, however, to build a succinct story, a logical connection between the geomorphology and the surface failure should, in my opinion, be expressed to highlight why volume estimations are important as it directly feeds into the story of hazard prediction moving forward.

**Response:** Thank you for your insightful comment. We appreciate your suggestion to improve both the grammar and phrasing to enhance clarity. We also acknowledge the need to strengthen the motivation for volume quantification and its geomorphological significance. In the revised manuscript, we emphasized the connection between volume estimates and the understanding of process mechanisms, illustrating their importance in the context of hazard prediction. The revised Introduction is given below,

“Landslides due to rainfall are phenomena that dislocate a mass of soil from its natural position and slide downward along a slope due to gravity forces. Intense or long-duration rainfall infiltrates the soil and increases the pore pressure, resulting in soil saturation that leads to slope failure. The saturated soil becomes weak and loses cohesion, and the slope fails when rainfall crosses a certain threshold (Bernardie et al., 2014; Martinović et al., 2018; Lee et al.,

2021). The heavy rainfall saturates a slope and triggers a landslide due to the reduction of the soil's shear strength and the increase of pore water pressure (Luino et al., 2022; Chen et al., 2021; Chatra et al., 2019; Lacerda et al., 2014; Tsai and Chen, 2010;). For example, steep slopes with loose soils and even moderate rainfall can lead to the displacement of an enormous quantity of soil mass. On the contrary, in slopes with more stable, cohesive soils, the surface failure might be smaller (Tsai and Chen, 2010). The rainfall quantity and duration influence the volume of the landslides; the higher the intensity and the longer the duration of rainfall, the larger the resulting surface failure (Chen et al., 2017; Bernardie et al., 2014; Chang and Chiang, 2009). The landslide occurrences can also be influenced by human activities that weaken the slope, such as excavation at the slope toe and loading caused by construction and land use such as agriculture, mining etc. (Rosi et al., 2016). The rapid urbanization activities affect the topography through hill cutting, deforestation and water drainage (Rahman et al., 2017); these activities disturb the slope structure and change the water flow, which exacerbates the effect of landslides in regions where human engineering activities are mostly located (Holcombe et al., 2016; Islam et al., 2017; Chen et al., 2019).

To estimate the volume of the soil mass displaceable subsequent to intensive rainfall, is essential to set appropriate mitigation strategies to reduce environmental degradation, infrastructure damage, casualties, and to establish post-disaster resilience policies to restore the socio-economic aspect of communities (Van et al., 2021; Alcántara-Ayala, 2021). This quantification of the volume of landslides due to rainfall (VLDR) is essential for effective risk management (Tacconi et al., 2020), emergency response, engineering design (Cheung, 2021), economic assessment and environmental protection (Alcántara-Ayala and Sassa, 2023). Firstly, to manage landslide risk effectively, the quantification of VLDR can be useful for updating hazard maps to reflect the scale of potential landslides in various regions to facilitate the identification of high-risk zones for monitoring and intervention. In addition, to develop mitigation strategies, such as land stabilization measures and land use planning, planners might put in place strict construction regulations in particular regions that are susceptible to landslides (Mateos et al., 2020). The accurate measurements of VLDR can be used to promote public awareness for safety measures and preparedness (Yang and Adler, 2008). Secondly, estimating precise VLDR is crucial for structural engineers to design a structure that can withstand extreme landslide events. Knowing the exact volume of displaceable material, an engineer can set robust stabilization solutions to prevent future occurrences (Dai and Lee, 2001). Moreover, the VLDR can help design the drainage system to manage water flow by controlling

groundwater and surface runoff to mitigate landslide risks (Dikshit et al., 2019; Kim et al., 2014). Furthermore, to prepare for emergence responses such as resource allocation, evacuation planning, and search and rescue operations, accurate VLDR estimation is necessary to ensure efficient implementation (Fan et al., 2019). To allocate resources effectively, the volume data is needed to determine the expected number of personnel for evacuation, materials sufficient for cleaning up and recovery (Amatya, 2016; Yang and Adler, 2008; Spiker and Gori, 2003). Further, to establish environmental protection measures such as ecosystem impacts, preservation of soil and water quality, and habitat restoration, the estimates of VLDR are essential (Pradhan et al., 2022; Li et al., 2022a; Barik et al., 2017).

To mitigate the economic impacts of landslides, the values of VLDR can be a basis for estimation of property damages, which is critical for settling insurance claims and assessment of financial impacts on communities and government to facilitate efficient budgeting for repairing damaged infrastructure and restoration of affected parts (Klimeš et al., 2017; Dai et al., 2002). The prediction of the VLDR can assist in long-term economic planning for landslide risk by creating disaster preparedness and recovery funds (Winter and Bromhead, 2012). The accurate estimation of the VLDR is an important key for designing strategies for resilience and planning for the protection of the inhabitants of a particular region with certain landslide risks subjected to a predicted quantity of rainfall (Conte et al., 2022). Consequently, for the safety of communities, the selection of infrastructure construction sites must be done in places with low landslide risks (Fan et al., 2017). Further, for the protection of crops, the farmland location, and other land use activities, accurate landslide prediction taking into account real root causes through the analysis of triggering and influencing factors, is crucial to achieve a durable landslide safety management system (Paudel et al., 2003; Lee, 2009; Fan et al., 2017; Chen et al., 2019; Dai et al., 2019; Alcántara-Ayala, 2021). “

Comment 2: Are the training and testing datasets split randomly with keeping the training data fixed or is the split performed geographically? It would be interesting to see a geographically split dataset to see how well the model(s) perform due to apparent differences in the geological and environmental conditions across the study area.

Response: Thank you for your insightful suggestion, which helped us improve the manuscript. In the present study, we opted to split the training and testing data randomly, implementing a 10-fold cross-validation to obtain an optimal model. This choice was made to balance bias and

variance effectively, adhering to a common 70% training and 30% testing split frequently employed in machine learning models (Nguyen et al., 2021), which has been shown to be an optimal data ratio.

While a geographically-based split could offer insight into regional variability, it may introduce challenges for this study, as landslide occurrences in our dataset are unevenly distributed, with about 60% located in the northeast part of the country. Geographically splitting this region as the test set would significantly reduce test data size, which could compromise model reliability and result in a suboptimal training process. To address regional variability without introducing geographic splitting, we incorporated altitude as a predictor variable in the model, recognizing that orographic rainfall in higher-altitude regions impacts soil saturation and may influence landslide susceptibility differently across regions. This approach allows the model to account for environmental differences while maintaining a balanced and representative dataset.

Comment 3: One of my main concerns, or rather my curiosity, is regarding the data set itself.

The volume information, along with the inventory, is particularly noteworthy in this case, as most inventories lack volume data. Keeping this in mind, how do the authors think about the application of such methods in other areas? Now, the authors have created a method that works pretty well within the given region. Instead of finding other regions (which might be difficult and time-consuming) could the authors simply use the model and predict volumes on similar nearby regions where the volumes are not calculated? This could serve as a simple prediction example demonstrating the method's application, without requiring extensive investigation. This approach is important as it helps the authors extend beyond a simple 'exercise' of the method, since it is currently applied only in the study area. Moreover, this would make the claim in Conclusion, Lines 346-349 more credible.

Response: Thank you for your insightful comment. We agree that extending the applicability of our model to other regions is a valuable goal. While a comprehensive analysis of other regions is beyond the scope of this study, we recognize the potential to apply our model to similar regions with similar geological and environmental conditions.

In the present investigation, we selected a test set treated as unknown data to the model, where volume predictions were based solely on predictor variables, and actual volume values were used only to evaluate model performance. Our results indicate that the DNN, EGB, GLM, and RF models performed well, achieving an  $R^2 > 0.8$ . This level of accuracy suggests that the

model could provide reliable volume estimates in adjacent areas with comparable input data. We have clarified this point in the revised manuscript to highlight the model's adaptability.

Comment 4: My biggest concern is related to the soil-depth. Now, it is impossible to imagine the calculation of volumes without the depth of the material that has failed as that is the 3<sup>rd</sup> volume calculations. It appears that the soil depth was 'removed' after feature importance analysis for the best performing EGB model. Sure, the depth information might not have been that important in this example of model training for this region, but I would argue that in other regions, particularly if the region contains multiple deep-seated landslides and the failure surface runs deep until the bedrock. I am just not convinced that removing soil depth makes sense, as geomorphologically, depth (which also relates to soil composition) is very important for accurate volume estimation and calculation.

Response: Thank you for the fruitful observation. We agree that soil depth is important in the prediction of the volume of landslides due to rainfall. In this study, the average topsoil depth was considered, and during the training process, the contribution was minor in the prediction of volumes and values below 0.01 were not shown even though those features remained in the model. To remove the confusion caused by the absence of those variables with less contribution on the variable importance plot and to acknowledge that those variables may be more significant in other regions, all variables used to train all models were shown in the updated manuscript. The updated figure with its caption in the revised manuscript is depicted below,

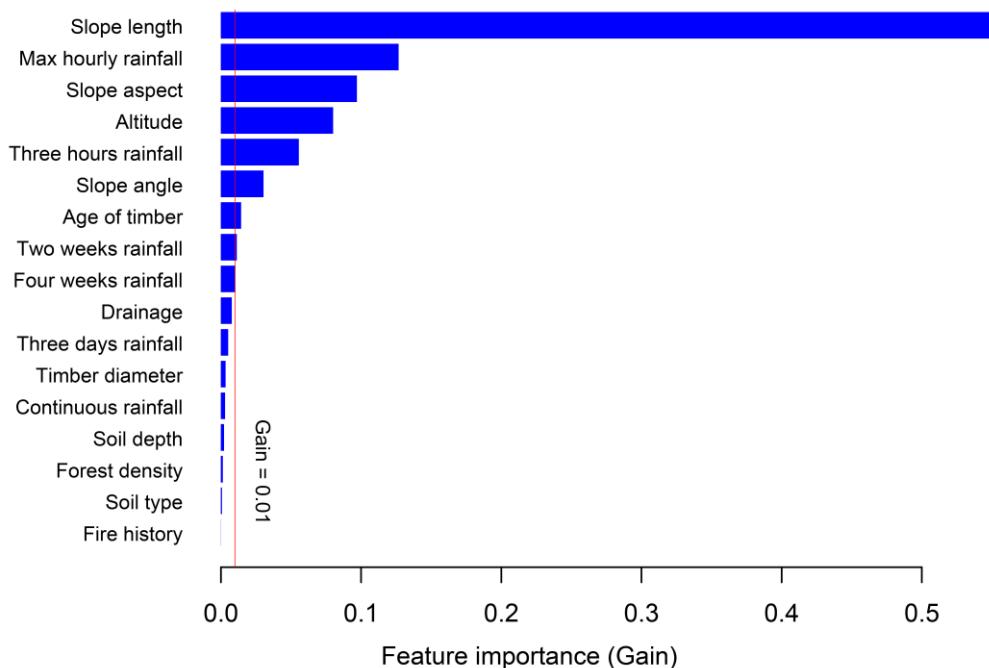


Figure 6. Variable importance for the EGB model.

Comment 5: Another question is pertaining to the type of failure movement. The inventory contains multitude of information but what about the movement types of the landslides? What types of landslides are considered in the inventory? Because clearly shallow and deep landslides would require separate treatments when looking at volume predictions because the material composition, material type, and material depths would be tremendously different. Do the authors combine these landslides together? What is the proportion of these landslide types? Also, are there prevalent debris flows, because volumes of debris flows is another story altogether since entrained volumes due to channelization are different than surface failure volumes. I see that the Discussion can be improved a lot by addressing and discussing these topics and limitations.

Response: We appreciate the reviewer's insightful comments. We agree that landslide movement types are critical for accurate volume predictions, as they exhibit distinct failure mechanisms, material properties, and depositional patterns. As the reviewer correctly noted, our initial dataset contained a variety of landslide types. Upon further examination, we identified that the majority of landslides in our study area were shallow, translational slope failures. Only one deep-seated landslide, with an approximate volume of 33,000 m<sup>3</sup>, was included in the inventory. As observed in prior studies, shallow translational slides are common in granite areas of Korea due to uniform weathering profiles, while metamorphic regions tend to experience larger debris flows due to steeper slopes and irregular weathering profiles (Kim and Chae, 2009). Kim et al. (2001) further noted that in north and northwest part of the country, most landslides are classified as debris flows, though their initiation points often exhibit characteristics of translational slides. Recognizing that shallow and deep-seated landslides exhibit different material properties, failure mechanisms, and volumetric characteristics, we have removed this deep-seated landslide from our analysis to ensure consistency and relevance to our study objectives. We have therefore focused our analysis on this dominant type, as it represents the primary landslide hazard in the region.

This manuscript contains exclusively shallow-seated landslides with volumes below 13,000m<sup>3</sup> with topsoil depth varying between 0.2m and 1m. We have updated the methodology and analysis sections to clarify that our dataset only includes shallow-seated landslides. Additionally, the Discussion section now addresses this limitation, acknowledging that the exclusion of deep-seated landslides and debris flows may affect the generalizability of our findings to other landslide types. This improvement aligns with the study's focus on shallow landslides, allowing for a more accurate assessment of volume predictions within this specific

landslide type. We have also noted in the Discussion that future studies would benefit from separate analyses of deep-seated landslides and debris flows, given the unique volumetric and channelization characteristics of debris flows.

**Comment 6:** The Discussion section is oriented quite too much on the aspects of the different models, conditioning factors, and their roles in the prediction of the volumes. As I mentioned in my previous comment, not much is discussed on the practical questions of scalability, different modes of movements, soil depths, runout volumes of entrained materials etc. These are essential topics as the direct counterpart of statistical models, i.e., numerical models tend to answer these questions. So, a comparison with the literature in that order is missing which I believe would add new levels of arguments to put forward by the authors and cement why their method works well despite lacking/following physical laws.

**Response:** This study aim was to construct a data-driven algorithm that predicts the volume of landslides due to rainfall. The result of nine different tested algorithms revealed a tremendous difference between classical regression models (OLS, RR, and GLM) and other data-driven machine learning models. In this study, apart from SVM regression, DT and KNN, other machine learning models (DNN, DT, RF, and EGB) exhibited high prediction capability with  $R^2$  above 50%. Further, to understand the applicability of the developed models, the trained model was tested using unknown data, with volume predictions generated solely based on the predictor variables; actual volume values were utilized only for evaluating model performance. We found that the DNN, EGB, GLM, and RF models achieved  $R^2 > 0.8$ , indicating that the model could yield reliable volume estimates in adjacent areas with similar geological and environmental conditions. It was noted that the numerical models and machine learning approach mostly used for the landslide volume estimation depend on landslide geometry (Leong and Cheng, 2022; Do et al., 2017; Shirzadi et al., 2017). As of our knowledge, none of the ML models used to predict volume of landslides using multiple predictors (such as, geological, topographical, geomorphological, soil, vegetation, and rainfall factors) on large scale. Therefore, the direct comparison with result of existing numerical and statistical models that solely depend on geometrical features of landslide (such as, surface area or runout length) is out of the scope of this investigation.

**Comment 7:** In Table 1, under Geomorphology, the feature “erosion” is presented. Now, erosion itself can be referred to the volume, which is the main variable that the authors are trying to estimate. So, how is this variable used in the training regime?

Or is this erosion feature different than the output of volume? Also, there are summary statistics of the erosion under Table 2. Why is that? My concern is that the authors are not clear as to what 'erosion' refers to in the data-driven model construct. If it is in fact similar to volumes, then the predictor variable and output variables are more or less the same. This needs further in-dept clarification.

Response: We appreciate the reviewer comment regarding the confusion originating from the use of 'erosion' as a predictor variable. We agreed that the term may have caused confusion.

In the preprint, the feature named 'erosion' was incorporated as a categorical variable with 'Yes' and 'No' values, indicating whether minor erosion events (such as gradual surface degradation due to wind or water) occurred prior to the landslide event. This differs from the volume variable, which is our dependent variable and represents the total mass of displaced material due to a landslide. Importantly, volume was not used as a predictor in the model; rather, it serves solely as the target output. To avoid ambiguity, we have removed the 'erosion' variable from the predictor variable list in Table 1 and accordingly updated Table 2 in the revised manuscript.

Table 2: Summary statistics continuous variables.

Variable	units	N	Min	Mean	Median	Max	Std dev
Max Hourly rain	mm	455	0	48	48	78	20
Continuous rainfall	mm	455	0	285	327	550	106
Three hours rainfall	mm	455	0	88	80	171	60
Twelve Hours rainfall	mm	455	0	150	99	447	95
One day rainfall	mm	455	0	202	162	538	112
Three days rain	mm	455	0	280	284	550	86
Seven days rain	mm	455	0.5	323	330	634	88
Two weeks rain	mm	455	0.5	385	400	663	90
Three weeks rain	mm	455	86	504	533	914	115
Four weeks rain	mm	455	108	587	561	1135	160
Soil depth	m	455	0.2	0.6	0.75	.75	0.19
Soil type	-	455	1.5	1.6	1.5	1.7	0.087
Timber diameter	m	455	0.15	0.27	0.23	0.35	0.086
Age of tree	Years	455	10	34	35	60	14
Slope length	m	455	1.8	21	13	180	23
Slope angle	Degree (o)	455	10	34	34	65	7.9
Altitude	m	455	9	391	272	1324	273

Comment 8: Table 1: Descriptions should be written properly for each feature/variable. At the moment, the descriptions read more like a summary of the sub-groups, written

altogether. Please provide descriptions individually for each feature properly. For example, Slope angle, slope aspect, and slope length are all written in one statement. Make them three individual statements to make it clearer to understand. Also, the descriptions are not clear enough. For volume landslide due to rainfall". This is not a description. It is a reasoning to justify a claim. Please provide appropriate descriptions.

**Response:** Thank you for your comment. We have updated Table 1 to enhance clarity by providing separate descriptions for each feature. Each description is now specific to the individual feature, detailing its relevance to landslide volume estimation. While rainfall parameters, such as rainfall on the day of the event and rainfall in prior days were grouped, as they represent related precipitation metrics, all other features have been distinctly separated. The revised version of Table 1, with improved feature descriptions, is shown below.

Table 1. Landslide influencing and triggering factors.

Group	Features	Description	Reference
Vegetation	Fire history	The burning of the vegetation intensifies the mass movement of soil near the uncovered burned stem of trees and free movement on uncovered soil due to post-fire rainfall and storms. The sliding may also be due to loss of vegetation, altered soil property and structure, which lead to soil degradation and infiltration which increase pore pressure, and change in hydrology by concentrating water flow in places that exacerbate landslides.	Highland and Bobrowsky, 2008; Culler et al., 2021; Hyde et al., 2016; Stoof et al., 2012
	Age of tree	Mature forests have more resistance to shallow landslides due to highly developed roots, which improve soil cohesion and leaves that prevent direct contact of raindrops with the soil surface.	Sato et al., 2023; Lann et al., 2024
	Forest density	The presence of forest reduces the likelihood of landslides about three times compared to grassland. Grassland has been revealed to be three times more vulnerable to shallow landslides than broadleaf and, coniferous and in secondary forests.	Lann et al., 2024; Greenwood et al., 2004; Turner et al., 2010; Scheidl et al., 2020; Asada et al., 2023
	Timber	Tree spacing and size had been used to	Cohen and

Group	Features	Description	Reference
Geomorphology	diameter (m)	investigate the effect of root and tree in shallow landslide control. The high root density generally enhances slope stability, and specific tree placement and root sizes between 5 to 20 mm are effective in landslide prevention.	Schwarz., 2017; Wang et al., 2016
	Drainage	The drainage has a significant effect on the slope stability and promotes the efficient control of the influence of rainfall on the ground water fluctuation. The presence of drainage increases the threshold of landslides due to rainfall.	Yan et al., 2019; Sun et al., 2010 ; Wei et al., 2019 ; Korup et al., 2007
	Slope angle (degree)	The steeper slopes have lower presence of landslide due to low transportable materials. Slopes between 20-40 degrees are most vulnerable to greater landslides as rainfall intensity and duration increase. Here, we considered the average angle of the terrain at the landslide location, which provides valuable insight into the region's overall steepness and geomorphic characteristics, which are crucial factors influencing landslide susceptibility and risk modeling.	Duc, 2013 ; Qiu et al., 2016 ; Donnarumma et al., 2013
	Slope aspect	The effect of rainfall on slope differs by slope angle and slope aspect which lead to unevenly distributed occurrence of landslides.	Panday and Dong, 2021; Cellek, 2021
	Slope length (m)	The volume increases as the slope length increases. There exists a complex interplay between rainfall, length of slope and slope angle on the occurrence of landslides.	Turner et al., 2010
	Soil depth (m)	Soil properties, depth, and texture have significant differences in infiltration rates, which have different influences on the occurrence of landslides.	Kitutu et al., 2009; McKenna et al., 2012
	Soil type	Higher rainfall intensity affects the occurrence of landslides differently, particularly in certain soil types that have shorter saturation and failure times.	Liu et al., 2021
Location	Altitude	Regional variability of elevation and mountain steepness affect the quantity of rainfall and associated landslides.	Hyun et al, 2010, Yoon and Bae, 2013; Park,

Group	Features	Description	Reference
Rainfall	Maximum hourly rainfall	The rainfall infiltrates the slope and increases pore water pressure that reduces soil shear strength, which leads to soil saturation that causes surface failure.	2015 Um et al., 2010 Wieczorek, 1987; Dai and Lee, 2001; Smith et al., 2023
	Continuous rainfall	Sudden intense rainfall concentrated in short periods of time is responsible for shallow landslide and debris flow.	Zhang et al., 2019
	Three hours rainfall		
	Three days rainfall		Ran et al., 2022
	Two weeks rainfall	The antecedent rainfalls increase moisture in the soil and weaken soil cohesion.	Zhang et al., 2019 ; Bernardie et al., 2014; Chen et al., 2015a; Gariano et al., 2017
	Four weeks rainfall		

Comment 9: Lines 311-312: It would be nice explain why the random forest works well with smaller volumes. The connection between the machine learning predictions and the scale of the estimated volumes should be explained more intricately to provide a grounded understanding. Does the EGB model predict larger volumes more accurately than the rest, like Random Forest? If so, then why? Please explain these aspects.

Response: Thank you for your insightful comment. Random Forest tends to perform well with smaller volumes due to its ability to capture complex relationships and interactions in the data without overfitting. RF uses multiple decision trees as base models, builds each tree on a random subset of samples and features, and computes averages as predictions to get the final result (Breiman, 2001). The model's random sampling of both observations and features allows it to build diverse trees; this enhances the generalization capabilities, particularly when the dataset is small. This characteristic helps the RF to maintain accuracy by reducing variance. It was noticed that the difference between of  $R^2$  on training and testing sets was small compared to other models.

In contrast, the EGB model may predict larger volumes more accurately because it employs an iterative process to improve predictions. It uses a decision tree as the base model and builds them sequentially in such a way that each new tree corrects prediction errors made by previous trees using gradient descent, allowing for fine-tuning of predictions over iterations and to minimize loss functions effectively (Chen and Guestrin, 2016). This iterative correction can capture complex patterns in larger datasets that may not be evident in smaller ones. The fact that the RF predictions are averages of multiple decision trees may cause the difference since predicting averages will be less than predictions produced sequentially (Sagi and Rokach, 2018).

Furthermore, as volume size increases, the relationships between features can become more intricate, and EGB's ability to handle these complexities may lead to superior performance in those scenarios. However, Random Forest remains advantageous when data is scarce because it is less prone to overfitting compared to some boosting methods, which may struggle with limited data. A clear understanding of these dynamics provides valuable insights into the varying performance of different models across different volume scales, emphasizing the importance of choosing the right algorithm based on dataset characteristics. This has been highlighted in the discussion section of the revised manuscript.

**Minor comments:**

Comment 1: Line 31: "high", should be "height".

Response: Thank you for your observation. The identified error has been corrected in the revised manuscript. The entire sentence has been modified as,

"Landslides due to rainfall are phenomena that dislocate a mass of soil from its natural position and slide downward along a slope due to gravity forces."

Comment 2: Line 36: "resulting volume of landslides". Change this to "resulting surface failure".

Response: We have made the modification in the revised manuscript, replacing "resulting volume of landslides" with "resulting surface failure" for improved clarity.

Response 3: Line 38: "fragilize". Not sure if such a word is used commonly to express the weakening of slopes. I'd rather opt for 'weaken'.

Response: Thank you for your suggestion. Accordingly, we have replaced "fragilize" with "weaken" in the revised manuscript.

Comment 4: Similar English issues are found in Section 2 (Study area). Please address the language issues.

Response: Thank you for the fruitful suggestion and observations. We have addressed the language issues throughout the entire manuscript, including the Study Area section. The modification made in the study area section is reflected in the text below:

"The region for testing the model is South Korea, characterized by mountainous (63% of total land) relief, especially in the eastern part of the country (Lee et al., 2022). South Korea is located on the southern part of the Korean Peninsula, bordered by the Yellow Sea to the west coast and the East Sea (Sea of Japan) to the East. According to the Korean Meteorological Administration (2020), the country has a temperate climate characterized by four distinct seasons: hot and humid summers, cold winters, and springs and falls with moderate temperatures. The annual rainfall ranges between 1000 mm to 1400mm and 1800mm for the central region and southern region, respectively (Jung et al., 2017; Alcantara and Ahn, 2020). During the summer, heavy rainfall from June to September leads to significant surface runoff, increases landslide risk, and causes approximately 95% of all landslides each year (Lee et al., 2020; Park and Lee, 2021). In addition, the landslides may be aggravated by typhoons, which mostly occur in August and September, and it is anticipated that frequency will increase due to climate change (Kim and Park, 2021). The rainfall trend analysis from 1971 to 2100 predicted the increase in rainfall of 271.23mm, which indicates the growing risk of landslides associated with climate change (Lee, 2016). Temperature variations are influenced by its geographical location, the average summer temperatures range between 25 and 30°C, while winter temperatures can drop to -10°C in some parts of the country (Korea Meteorological Administration, 2020). The South Korean geologically is mainly composed of granitic and metamorphic rocks, such as gneiss, schist, and granite, which influence the stability of the landscape (Jung et al., 2024). The geomorphology is characterized by rugged mountains, river valleys, and coastal plains, with the Taebaek Mountains running along the eastern edge (Kim et al., 2020). In addition, the influence of rainfall, environmental, geomorphology, and

geological factors increase the vulnerability to landslides across the country, especially in the northeastern mountainous region, as depicted in Figure 1.

The predominant soil types in South Korea include clay, sandy, and loamy soils, each with different characteristics affecting water infiltration, retention and erosion (Kang et al., 2022; Lee et al., 2023). Clay soils, being more stable, can become highly saturated, increasing landslide risk during heavy rains. On the other hand, sandy soils are more prone to shallow landslides due to fast saturation, leading to instability. Regions with steep topography and poorly consolidated soil (loose) are mostly at risk, especially after prolonged rainfalls (Kim et al., 2015).

Coastal areas are exposed to sea-level rise and coastal erosion, which can further complicate the landscape and increase landslide susceptibility. The combination of heavy summer rainfall, geological composition, and geomorphological factors makes South Korea particularly vulnerable to shallow landslides. Thus, continuous monitoring and research are vital to understand the complex interactions between climate, geology, soil types, and landslide occurrences in this region (Park, 2022). Understanding the combination of environmental, geological stability, and geomorphological features is crucial for developing effective disaster management strategies and enhancing public safety in landslide-prone areas. As climate change continues to impact rainfall patterns, South Korea faces ongoing challenges in mitigating landslide risks and protecting vulnerable communities.”

Comment 5: Figure 2. Font size of plot (b) is different than the rest, and also stretched. Please make all font sizes uniform.

Response: Thank you for your suggested improvements. Figure 2 (now Figure 3), titled “Workflow for the Prediction of Volume of Landslide Due to Rainfall,” has been revised to ensure uniform font sizes throughout the plot. The updated figure is provided below,

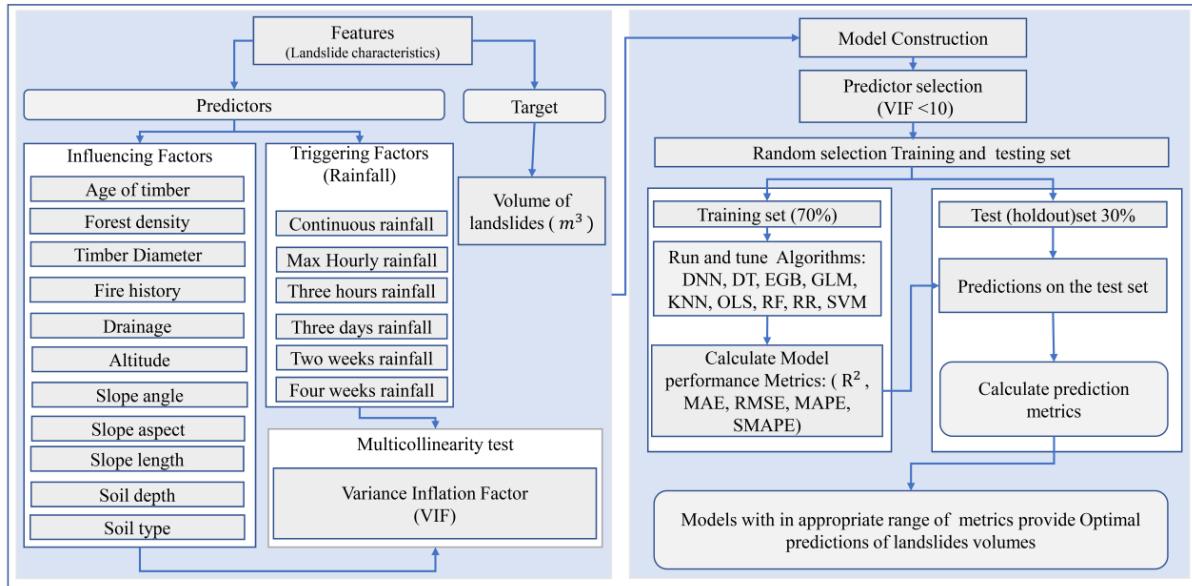


Figure 3. Workflow for the prediction of the volume of landslides due to rainfall.

Comment 6: Line 111: Replace ‘joined’ with ‘combined’.

Response: Thank you for your comment. As suggested, the term "joined" has been replaced with "combined" in the revised manuscript.

Comment 7: Line 128: “flown away”? I am not sure if using this term is accurate. Generally, we refer to them as “removed material” from the surface. Can you please double-check this?

Response: Thank you for your valuable comment. The suggested modifications were incorporated in the revised manuscript as,

“The estimation of the volume of removed material by landslides is important as it helps to assess risks the estimated damage can cause down at the toe of the failed slope, such as blocking transportation network, burying crops or farmland, the damage-built environment near landslide risks area, and post-disaster recovery planning (Evans et al., 2007; Rotaru et al., 2007; Intrieri et al., 2019).”

Comment 8: Is the slope angle the average angle of the terrain where the landslide was located or is the angle of reach? In my opinion, the angle of reach would make more sense as landslides that are closer to each other will exhibit different angles of reach but the same adjacent landslides would bear the same average slope angle as you are

averaging based on the terrain. Please make it clear as to which one you have considered and why.

**Response:** Thank you for your comment. The slope angle referenced in the manuscript pertains to the average angle of the terrain at the landslide location. This measurement provides valuable insight into the overall steepness and geomorphic characteristics of the area, which are crucial factors influencing landslide susceptibility and risk modeling (Donnarumma et al., 2013). On the other hand, the angle of reach refers to the angle at which a landslide material travels after detaching from the slope, which is important for assessing mobility and potential impact (Corominas, 1996). However, this is a different metric and not the focus of our analysis. While the angle of reach considers the mobility of landslides, the average slope angle is critical for assessing the risk of landslide occurrence. We acknowledge your point regarding the differences in angle of reach among closely situated landslides, but in our study, the average slope angle is more relevant for evaluating landslide volume. We have clarified this distinction in the revised manuscript (Table 1) to ensure a better understanding.

Comment 9: Line 136: What do you mean by ‘composing material’? This is not clear.

**Response:** Thank you for your insightful comment. The term "composing material" refers to soil composition properties, which significantly impact slope stability. These properties, including soil permeability indices, influence water infiltration and saturation levels, both of which are critical factors in landslide susceptibility (Chen et al., 2015a). The revised sentence is as follows,

“The slope stability depends on soil composition properties, including soil permeability indices that affect water infiltration and saturation level (Chen et al., 2015a).”

Comment 10: Lines 140-142: Please check the English grammar here. The sentence can be improved a lot.

**Response:** Thank you for your comment. We revised the sentence in the updated manuscript. Additionally, we conducted a thorough review of the manuscript to identify and correct similar issues throughout.

Comment 11: Line 341: Change to “Among the tested models,”

Response: Thank you for your comment. The sentence has been modified in the revised manuscript as,

“Among the tested models, extreme gradient boosting (EGB) produced the most accurate prediction.”

Comment 12: Conclusion- Line 349: Change from “can be a better tool” to “can be a good tool”.

Response: Thank you for your comment. As suggested, the sentence has been revised in the updated manuscript as,

“Therefore, this model can be a good tool for planning for resilience and infrastructure pre-construction risk assessment to ensure the new infrastructure is placed in stable regions free from severe landslides.”

## References

Alcantara, A. L., and Ahn, K. H. (2020). Probability distribution and characterization of daily precipitation related to tropical cyclones over the Korean Peninsula. *Water*, 12(4), 1214.

Alcántara-Ayala, I. (2021). Integrated landslide disaster risk management (ILDRiM): the challenge to avoid the construction of new disaster risk. *Environmental Hazards*, 20(3), 323-344.

Alcántara-Ayala, I., & Sassa, K. (2023). Landslide risk management: from hazard to disaster risk reduction. *Landslides*, 20(10), 2031-2037.

Amatya, S. C. (2016). Landslide disaster management in Nepal: a near-future perspective. Nepal-Japan Friendship Association of Water Induced Disaster (NFAD), Japan Department of Water Induced Disaster Management (DWIDM).

Asada, H., & Minagawa, T. (2023). Impact of vegetation differences on shallow landslides: a case study in Aso, Japan. *Water*, 15(18), 3193.

Barik, M. G., Adam, J. C., Barber, M. E., & Muhunthan, B. (2017). Improved landslide susceptibility prediction for sustainable forest management in an altered climate. *Engineering geology*, 230, 104-117.

Bernardie, S., Desramaut, N., Malet, J.-P., Gourlay, M., and Grandjean, G. (2014). Prediction of changes in landslide rates induced by rainfall. *Landslides*, 12(3), 481–494. doi:10.1007/s10346-014-0495-8

Breiman, L. (2001). Random forests. *Machine learning*, 45, 5-32.

Chang, K. T., and Chiang, S. H. (2009). An integrated model for predicting rainfall-induced landslides. *Geomorphology*, 105(3-4), 366-373.

Chatra, A. S., Dodagoudar, G. R., & Maji, V. B. (2019). Numerical modelling of rainfall effects

on the stability of soil slopes. *International Journal of Geotechnical Engineering*.

Chen, C. W., Oguchi, T., Hayakawa, Y. S., Saito, H., and Chen, H. (2017). Relationship between landslide size and rainfall conditions in Taiwan. *Landslides*, 14, 1235-1240.

Chen, L., Guo, Z., Yin, K., Shrestha, D. P., & Jin, S. (2019). The influence of land use and land cover change on landslide susceptibility: a case study in Zhushan Town, Xuan'en County (Hubei, China). *Natural hazards and earth system sciences*, 19(10), 2207-2228.

Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 785-794).

Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 785-794).

Chen, X., Zhang, L., Zhang, L., Zhou, Y., Ye, G., & Guo, N. (2021). Modelling rainfall-induced landslides from initiation of instability to post-failure. *Computers and geotechnics*, 129, 103877.

Chen, Z., Luo, R., Huang, Z., Tu, W., Chen, J., Li, W., ... and Ai, Y. (2015a). Effects of different backfill soils on artificial soil quality for cut slope revegetation: Soil structure, soil erosion, moisture retention and soil C stock. *Ecological engineering*, 83, 5-12.

Cheung, R. W. (2021). Landslide risk management in Hong Kong. *Landslides*, 18(10), 3457-3473.

Cohen, D., & Schwarz, M. (2017). Tree-root control of shallow landslides. *Earth Surface Dynamics*, 5(3), 451-477.

Conte, E., Pugliese, L., and Troncone, A. (2022). A simple method for predicting rainfall-induced shallow landslides. *Journal of Geotechnical and Geoenvironmental Engineering*, 148(10), 04022079.

Corominas, J. (1996). The angle of reach as a mobility index for small and large landslides. *Canadian Geotechnical Journal*, 33(2), 260-271.

Culler, E. S., Livneh, B., Rajagopalan, B., & Tiampo, K. F. (2021). A data-driven evaluation of post-fire landslide susceptibility. *Natural Hazards and Earth System Sciences Discussions*, 2021, 1-24. <https://doi.org/10.5194/nhess-23-1631-2023>

Dai, F. C., & Lee, C. F. (2001). Frequency-volume relation and prediction of rainfall-induced landslides. *Engineering geology*, 59(3-4), 253-266.

Dai, F. C., Lee, C. F., & Ngai, Y. Y. (2002). Landslide risk assessment and management: an overview. *Engineering geology*, 64(1), 65-87.

Dai, K., Xu, Q., Li, Z., Tomás, R., Fan, X., Dong, X., ... and Ran, P. (2019). Post-disaster assessment of 2017 catastrophic Xinmo landslide (China) by spaceborne SAR interferometry. *Landslides*, 16, 1189-1199.

Dikshit, A., Satyam, N., & Pradhan, B., 2019. Estimation of rainfall-induced landslides using the TRIGRS model. *Earth Systems and Environment*, 3, 575-584.

Do, X. K., Regmi, R. K., Nguyen, H. P. T., & Jung, K. (2017). Study on the formation and geometries of rainfall-induced landslide dams. *KSCE Journal of Civil Engineering*, 21, 1657-1667.

Donnarumma, A., Revellino, P., Grelle, G., & Guadagno, F. M. (2013). Slope angle as indicator parameter of landslide susceptibility in a geologically complex area. *Landslide Science and Practice: Volume 1: Landslide Inventory and Susceptibility and Hazard Zoning*, 425-433.

Duc, D. M. (2013). Rainfall-triggered large landslides on 15 December 2005 in Van Canh district, Binh Dinh province, Vietnam. *Landslides*, 10(2), 219-230.

Evans, S. G., Mugnozza, G. S., Strom, A., and Hermanns, R. L. (Eds.). (2007). *Landslides from massive rock slope failure* (Vol. 49). Springer Science and Business Media.

Fan, X., Xu, Q., Liu, J., Subramanian, S. S., He, C., Zhu, X., & Zhou, L. (2019). Successful early warning and emergency response of a disastrous rockslide in Guizhou province, China. *Landslides*, 16, 2445-2457.

Fan, X., Xu, Q., Scaringi, G., Dai, L., Li, W., Dong, X., ... & Havenith, H. B. (2017). Failure mechanism and kinematics of the deadly June 24th 2017 Xinmo landslide, Maoxian, Sichuan, China. *Landslides*, 14, 2129-2146. <https://doi.org/10.1007/s10346-017-0907-7>

Gariano, S. L., Rianna, G., Petrucci, O., and Guzzetti, F. (2017). Assessing future changes in the occurrence of rainfall-induced landslides at a regional scale. *Science of the total environment*, 596, 417-426.

Greenwood, J. R., Norris, J. E., & Wint, J. (2004). Assessing the contribution of vegetation to slope stability. *Proceedings of the Institution of Civil Engineers-Geotechnical Engineering*, 157(4), 199-207.

Highland, L. and Bobrowsky, P. (2008). *The Landslide Handbook: A Guide to Understanding Landslides*, United States Geological Survey, Reston, VA, Circular 1325, <https://pubs.usgs.gov/circ/1325/> (last access: 6 March 2023).

Holcombe, E. A., Beesley, M. E., Vardanega, P. J., & Sorbie, R. (2016). Urbanisation and landslides: hazard drivers and better practices. In *Proceedings of the Institution of Civil Engineers-Civil Engineering* (Vol. 169, No. 3, pp. 137-144). Thomas Telford Ltd.

Hyde, K. D., Riley, K., & Stoof, C. (2016). Uncertainties in predicting debris flow hazards following wildfire. *Natural hazard*

Hyun, Y. K., Kar, S. K., Ha, K. J., & Lee, J. H. (2010). Diurnal and spatial variabilities of monsoonal CG lightning and precipitation and their association with the synoptic weather conditions over South Korea. *Theoretical and applied climatology*, 102, 43-60.

Intrieri, E., Carlà, T., and Gigli, G. (2019). Forecasting the time of failure of landslides at slope-scale: A literature review. *Earth-science reviews*, 193, 333-349.

Islam, M. A., Islam, M. S., & Islam, T. (2017, September). Landslides in Chittagong hill tracts and possible measures. In *Proceedings of the international conference on disaster risk mitigation*, Dhaka, Bangladesh (Vol. 23).

Jung, M. J., Jeong, Y. J., Shin, W. J., & Cheong, A. C. S. (2024). Isotopic distribution of bioavailable Sr, Nd, and Pb in Chungcheongbuk-do Province, Korea. *Journal of Analytical Science and Technology*, 15(1), 46.

Jung, Y., Shin, J. Y., Ahn, H., and Heo, J. H. (2017). The spatial and temporal structure of

extreme rainfall trends in South Korea. *Water*, 9(10), 809.

Kang, M. W., Yibeltal, M., Kim, Y. H., Oh, S. J., Lee, J. C., Kwon, E. E., & Lee, S. S. (2022). Enhancement of soil physical properties and soil water retention with biochar-based soil amendments. *Science of the total environment*, 836, 155746.

Kim, D. E., Seong, Y. B., Weber, J., & Yu, B. Y. (2020). Unsteady migration of Taebaek Mountain drainage divide, Cenozoic extensional basin margin, Korean Peninsula. *Geomorphology*, 352, 107012.

Kim, H. G., & Park, C. Y. (2021). Landslide susceptibility analysis of photovoltaic power stations in Gangwon-do, Republic of Korea. *Geomatics, Natural Hazards and Risk*, 12(1), 2328-2351.

Kim, J., Lee, K., Jeong, S., & Kim, G. (2014). GIS-based prediction method of landslide susceptibility using a rainfall infiltration-groundwater flow model. *Engineering geology*, 182, 63-78.

Kim, M. S., Onda, Y., Kim, J. K., & Kim, S. W. (2015). Effect of topography and soil parameterisation representing soil thicknesses on shallow landslide modelling. *Quaternary International*, 384, 91-106.

Kim, W. Y., & Chae, B. G. (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.

Kim, Won-Yeong, Kim, Kyeong-Soo, Chae, Byeong-Gon, & Cho, Yong-Chan. (2001). Quantitative Prediction of Landslide Probability in Gyeonggi Province, Korea. *Journal of the Korean Society of Engineering Geology*, 33-44.

Kitutu, M. G., Muwanga, A., Poesen, J., and Deckers, J. A. (2009). Influence of soil properties on landslide occurrences in Bududa district, Eastern Uganda. *African journal of agricultural research*, 4(7), 611-620.

Klimeš, J., Stemberk, J., Blahut, J., Krejčí, V., Krejčí, O., Hartvich, F., & Kycl, P. (2017). Challenges for landslide hazard and risk management in 'low-risk' regions, Czech Republic—landslide occurrences and related costs (IPL project no. 197). *Landslides*, 14, 771-780.

Korup, O., Clague, J. J., Hermanns, R. L., Hewitt, K., Strom, A. L., and Weidinger, J. T. (2007). Giant landslides, topography, and erosion. *Earth and Planetary Science Letters*, 261(3-4), 578-589.

Lacerda, W. A., Palmeira, E. M., Netto, A. L. C., & Ehrlich, M. (Eds.). (2014). Extreme rainfall induced landslides: an international perspective. Oficina de Textos.

Lann, T., Bao, H., Lan, H., Zheng, H., & Yan, C. (2024). Hydro-mechanical effects of vegetation on slope stability: A review. *Science of the Total Environment*, 171691.

Lee, D. B., Kim, Y. N., Sonn, Y. K., & Kim, K. H. (2023). Comparison of Soil Taxonomy (2022) and WRB (2022) Systems for classifying Paddy Soils with different drainage grades in South Korea. *Land*, 12(6), 1204.

Lee, D. H., Cheon, E., Lim, H. H., Choi, S. K., Kim, Y. T., & Lee, S. R. (2021). An artificial neural network model to predict debris-flow volumes caused by extreme rainfall in the central region of South Korea. *Engineering Geology*, 281, 105979.

Lee, D. H., Kim, Y. T., and Lee, S. R. (2020). Shallow landslide susceptibility models based on

artificial neural networks considering the factor selection method and various non-linear activation functions. *Remote Sensing*, 12(7), 1194.

Lee, J. U., Cho, Y. C., Kim, M., Jang, S. J., Lee, J., & Kim, S. (2022). The effects of different geological conditions on landslide-triggering rainfall conditions in South Korea. *Water*, 14(13), 2051.

Lee, M. J. (2016). Rainfall and landslide correlation analysis and prediction of future rainfall base on climate change. In *Geohazards Caused by Human Activity*. IntechOpen.

Lee, S. G. (2009). The Effects of Landslide in South Korea and Some Issues for Successful Management and Mitigation. *한국토양비료학회 학술발표회 초록집*, 181-191.

Leong, E. C., & Cheng, Z. (2022). A geometry-modelling method to estimate landslide volume from source area. *Landslides*, 19(8), 1971-1985.

Liu, Y., Deng, Z., & Wang, X. (2021). The effects of rainfall, soil type and slope on the processes and mechanisms of rainfall-induced shallow landslides. *Applied Sciences*, 11(24), 11652.

Luino, F., De Graff, J., Biddoccu, M., Faccini, F., Freppaz, M., Roccati, A., ... & Turconi, L. (2022). The Role of soil type in triggering shallow landslides in the alps (Lombardy, Northern Italy). *Land*, 11(8), 1125.

Martinović, K., Gavin, K., Reale, C., and Mangan, C. (2018). Rainfall thresholds as a landslide indicator for engineered slopes on the Irish Rail network. *Geomorphology*, 306, 40-50.

Mateos, R. M., López-Vinielles, J., Poyiadji, E., Tsagkas, D., Sheehy, M., Hadjicharalambous, K., ... & Herrera, G. (2020). Integration of landslide hazard into urban planning across Europe. *Landscape and urban planning*, 196, 103740.

McKenna, J. P., Santi, P. M., Amblard, X., and Negri, J. (2012). Effects of soil-engineering properties on the failure mode of shallow landslides. *Landslides*, 9, 215-228.

Nguyen, Q. H., Ly, H. B., Ho, L. S., Al-Ansari, N., Le, H. V., Tran, V. Q., ... & Pham, B. T. (2021). Influence of data splitting on performance of machine learning models in prediction of shear strength of soil. *Mathematical Problems in Engineering*, 2021(1), 4832864.

Panday, S., & Dong, J. J. (2021). Topographical features of rainfall-triggered landslides in Mon State, Myanmar, August 2019: Spatial distribution heterogeneity and uncommon large relative heights. *Landslides*, 18(12), 3875-3889.

Park, C. Y. (2015). The classification of extreme climate events in the Republic of Korea. *Journal of the Korean association of regional geographers*, 21(2), 394-410.

Park, S. J., & Lee, D. K. (2021). Predicting susceptibility to landslides under climate change impacts in metropolitan areas of South Korea using machine learning. *Geomatics, Natural Hazards and Risk*, 12(1), 2462-2476.

Park, S.J. (2022). Assessment of disaster risks induced by climate change, using machine learning techniques (Doctoral dissertation, 서울대학교 대학원).

Paudel, P. P., Omura, H., Kubota, T., and Morita, K. (2003). Landslide damage and disaster management system in Nepal. *Disaster Prevention and Management: An International Journal*, 12(5), 413-419.

Pradhan, S., Toll, D. G., Rosser, N. J., & Brain, M. J. (2022). An investigation of the combined effect of rainfall and road cut on landsliding. *Engineering Geology*, 307, 106787. <https://doi.org/10.1016/j.enggeo.2022.106787>

Qiu, H., Regmi, A. D., Cui, P., Cao, M., Lee, J., and Zhu, X. (2016). Size distribution of loess slides in relation to local slope height within different slope morphologies. *Catena*, 145, 155-163.

Rahman, M. S., Ahmed, B., & Di, L. (2017). Landslide initiation and runout susceptibility modeling in the context of hill cutting and rapid urbanization: a combined approach of weights of evidence and spatial multi-criteria. *Journal of Mountain Science*, 14(10), 1919-1937.

Ran, Q., Wang, J., Chen, X., Liu, L., Li, J., & Ye, S. (2022). The relative importance of antecedent soil moisture and precipitation in flood generation in the middle and lower Yangtze River basin. *Hydrology and Earth System Sciences*, 26(19), 4919-4931.

Rosi, A., Peternel, T., Jemec-Auflič, M., Komac, M., Segoni, S., and Casagli, N. (2016). Rainfall thresholds for rainfall-induced landslides in Slovenia. *Landslides*, 13, 1571-1577.

Rotaru, A., Oajdea, D., and Răileanu, P. (2007). Analysis of the landslide movements. *International journal of geology*, 1(3), 70-79.

Sagi, O., & Rokach, L. (2018). Ensemble learning: A survey. *Wiley interdisciplinary reviews: data mining and knowledge discovery*, 8(4), e1249.

Sato, T., Katsuki, Y., & Shuin, Y. (2023). Evaluation of influences of forest cover change on landslides by comparing rainfall-induced landslides in Japanese artificial forests with different ages. *Scientific reports*, 13(1), 14258.

Scheidl, C., Heiser, M., Kamper, S., Thaler, T., Klebinder, K., Nagl, F., ... and Seidl, R. (2020). The influence of climate change and canopy disturbances on landslide susceptibility in headwater catchments. *Science of the total environment*, 742, 140588.

Shirzadi, A., Shahabi, H., Chapi, K., Bui, D. T., Pham, B. T., Shahedi, K., & Ahmad, B. B. (2017). A comparative study between popular statistical and machine learning methods for simulating volume of landslides. *Catena*, 157, 213-226.

Smith, H. G., Neverman, A. J., Betts, H., & Spiekermann, R. (2023). The influence of spatial patterns in rainfall on shallow landslides. *Geomorphology*, 437, 108795.

Spiker, E. C., & Gori, P. (2003). National landslide hazards mitigation strategy, a framework for loss reduction (No. 1244). US Geological Survey.

Stoof, C. R., Vervoort, R. W., Iwema, J., Van Den Elsen, E., Ferreira, A. J. D., & Ritsema, C. J. (2012). Hydrological response of a small catchment burned by experimental fire. *Hydrology and Earth System Sciences*, 16(2), 267-285.

Sun, H. Y., Wong, L. N. Y., Shang, Y. Q., Shen, Y. J., and Lü, Q. (2010). Evaluation of drainage tunnel effectiveness in landslide control. *Landslides*, 7, 445-454.

Tacconi Stefanelli, C., Casagli, N., & Catani, F. (2020). Landslide damming hazard susceptibility maps: a new GIS-based procedure for risk management. *Landslides*, 17, 1635-1648.

Tsai, T. L., & Chen, H. F. (2010). Effects of degree of saturation on shallow landslides triggered by rainfall. *Environmental Earth Sciences*, 59, 1285-1295.

Turner, T. R., Duke, S. D., Fransen, B. R., Reiter, M. L., Kroll, A. J., Ward, J. W., ... & Bilby, R. E. (2010). Landslide densities associated with rainfall, stand age, and topography on forested landscapes, southwestern Washington, USA. *Forest Ecology and Management*, 259(12), 2233-2247.

Um, M. J., Yun, H., Cho, W., & Heo, J. H. (2010). Analysis of orographic precipitation on Jeju-Island using regional frequency analysis and regression. *Water resources management*, 24, 1461-1487.

Van Tien, P., Luong, L. H., Duc, D. M., Trinh, P. T., Quynh, D. T., Lan, N. C., ... & Loi, D. H. (2021). Rainfall-induced catastrophic landslide in Quang Tri Province: the deadliest single landslide event in Vietnam in 2020.

Wang, D., Hollaus, M., Schmaltz, E., Wieser, M., Reifelshammer, D., & Pfeifer, N. (2016). Tree stem shapes derived from TLS data as an indicator for shallow landslides. *Procedia Earth and Planetary Science*, 16, 185-194.

Wei, Z. L., Shang, Y. Q., Sun, H. Y., Xu, H. D., and Wang, D. F. (2019). The effectiveness of a drainage tunnel in increasing the rainfall threshold of a deep-seated landslide. *Landslides*, 16, 1731-1744.

Wieczorek, G. (1987). In central Santa Cruz Mountains, California. Debris flows/avalanches: process, recognition, and mitigation, 7, 93.

Winter, M. G., & Bromhead, E. N. (2012). Landslide risk: some issues that determine societal acceptance. *Natural Hazards*, 62, 169-187.

Yan, L., Xu, W., Wang, H., Wang, R., Meng, Q., Yu, J., and Xie, W. C. (2019). Drainage controls on the Donglingxing landslide (China) induced by rainfall and fluctuation in reservoir water levels. *Landslides*, 16, 1583-1593.

Yang, H., & Adler, R. F. (2008). Predicting global landslide spatiotemporal distribution: integrating landslide susceptibility zoning techniques and real-time satellite rainfall estimates. *International Journal of Sediment Research*, 23(3), 249-257.

Yoon, S. S., & Bae, D. H. (2013). Optimal rainfall estimation by considering elevation in the Han River Basin, South Korea. *Journal of Applied Meteorology and Climatology*, 52(4), 802-818.

Zhang, K., Wang, S., Bao, H., & Zhao, X. (2019). Characteristics and influencing factors of rainfall-induced landslide and debris flow hazards in Shaanxi Province, China. *Natural Hazards and Earth System Sciences*, 19(1), 93-105.

## Response to Reviewer #2

### General remarks

**Comment 1:** In the introduction, the authors should explain more about why volume estimations are crucial for understanding and managing landslide hazards.

Response: Thank you for your valuable comment. As suggested, we have revised the introduction to highlight the critical role of landslide volume estimations in understanding and managing landslide hazards. The revised section of the introduction is provided below.

“To estimate the volume of the soil mass displaceable subsequent to intensive rainfall, is essential to set appropriate mitigation strategies to reduce environmental degradation, infrastructure damage, casualties, and to establish post-disaster resilience policies to restore the socio-economic aspect of communities (Van et al., 2021; Alcántara-Ayala, 2021). This quantification of the volume of landslides due to rainfall (VLDR) is essential for effective risk management (Tacconi et al., 2020), emergency response, engineering design (Cheung, 2021), economic assessment and environmental protection (Alcántara-Ayala and Sassa, 2023). Firstly, to manage landslide risk effectively, the quantification of VLDR can be useful for updating hazard maps to reflect the scale of potential landslides in various regions to facilitate the identification of high-risk zones for monitoring and intervention. In addition, to develop mitigation strategies, such as land stabilization measures and land use planning, planners might put in place strict construction regulations in particular regions that are susceptible to landslides (Mateos et al., 2020). The accurate measurements of VLDR can be used to promote public awareness for safety measures and preparedness (Yang and Adler, 2008). Secondly, estimating precise VLDR is crucial for structural engineers to design a structure that can withstand extreme landslide events. Knowing the exact volume of displaceable material, an engineer can set robust stabilization solutions to prevent future occurrences (Dai and Lee, 2001). Moreover, the VLDR can help design the drainage system to manage water flow by controlling groundwater and surface runoff to mitigate landslide risks (Dikshit et al., 2019; Kim et al., 2014). Furthermore, to prepare for emergence responses such as resource allocation, evacuation planning, and search and rescue operations, accurate VLDR estimation is necessary to ensure efficient implementation (Fan et al., 2019). To allocate resources effectively, the volume data is needed to determine the expected number of personnel for evacuation, materials sufficient for cleaning up and recovery (Amatya, 2016; Yang and Adler, 2008; Spiker and Gori, 2003). Further, to establish environmental protection measures such as ecosystem impacts,

preservation of soil and water quality, and habitat restoration, the estimates of VLDR are essential (Pradhan et al., 2022; Li et al., 2022a; Barik et al., 2017).

To mitigate the economic impacts of landslides, the values of VLDR can be a basis for estimation of property damages, which is critical for settling insurance claims and assessment of financial impacts on communities and government to facilitate efficient budgeting for repairing damaged infrastructure and restoration of affected parts (Klimeš et al., 2017; Dai et al., 2002). The prediction of the VLDR can assist in long-term economic planning for landslide risk by creating disaster preparedness and recovery funds (Winter and Bromhead, 2012). The accurate estimation of the VLDR is an important key for designing strategies for resilience and planning for the protection of the inhabitants of a particular region with certain landslide risks subjected to a predicted quantity of rainfall (Conte et al., 2022). Consequently, for the safety of communities, the selection of infrastructure construction sites must be done in places with low landslide risks (Fan et al., 2017). Further, for the protection of crops, the farmland location, and other land use activities, accurate landslide prediction taking into account real root causes through the analysis of triggering and influencing factors, is crucial to achieve a durable landslide safety management system (Paudel et al., 2003; Lee, 2009; Fan et al., 2017; Chen et al., 2019; Dai et al., 2019; Alcántara-Ayala, 2021). ”

**Comment 2:** The literature review section should be expanded to incorporate more recent studies on landslide volume prediction models, providing a comprehensive overview of the current state of research in this field.

Response: We appreciate this suggestion. We have done an extensive literature review to include recent studies on landslide volume prediction models, offering a more comprehensive overview of the current state of research in this field. Accordingly, the literature review section in the introduction has been updated as,

“The prediction of VLDR has gained the interest of many researchers to understand the mechanism and interaction between triggering and aggravating factors. Saito et al. (2014) studied the relationship between rainfall-triggered landslides to test whether the volume of landslides across Japan that occurred between 2001 and 2011 can be directly predicted from rainfall metrics. The findings revealed that larger landslides occurred when rainfall exceeded certain thresholds, but there were significant discrepancies between peaks of rainfall metrics and maximum landslide volumes, and the total rainfall was the suitable predictor of landslides.

Dai and Lee (2001) established the frequency-volume relation for landslides in Hong Kong and noticed that the relation for shallow landslides above  $4\text{m}^3$  followed the power law. The 12-hour rolling rainfall contributed most to the prediction of the volume of landslides. Ju et al. (2023) constructed an area-volume power law model for the estimation of the volume of landslides using high-resolution LiDAR data collected between 2010 and 2020 in Hong Kong. The aim was to estimate accurately the volume of landslides on small-scale landslides. The reliance on localized datasets limits the model's applicability in regions with different geological settings, and the model does not consider all variabilities of landslide characteristics. Razakova et al. (2020) calculated landslide volume using remote sensed data with the aim of assessing the efficiency of aerial photographs in environmental impact assessment and ground-based measurement. The study did not take into account the effect of vegetation and topography and only focused on a single landslide case, which may be a source of bias due to differences in soil composition and environmental factors. Hovius et al. (1997) analyzed multiple sets of aerial photos and frequency-magnitude relation for landslides in New Zealand. The finding pinpointed that the landslides frequency-magnitude followed power law and infrequent large magnitude contributed to the landscape change. The study also noticed the importance of soil composition in the size of the landslides. This work had a limitation due to the reliance on aerial photos only, which cannot provide accurate measurement in regions of dense forest, and the climatic conditions, which are landslide triggering factors, were not considered, and this may affect the generality of the findings. Guzzetti et al. (2008) applied statistical methods on regional landslide inventories and antecedent rainfall data ranging between 10 min to 35 days. The findings revealed that the slope angle and soil type significantly influence landslide volume estimates, and the rainfall intensity is more important than duration. Chatra et al., 2019) applied numerical methods to study the effect of rainfall duration and intensity on the generation of pore pressure in the soil; the finding revealed a higher instability in loose soil compared to medium soil slopes. The work only treated the interaction of soil and rainfall without considering the environmental factors and human activity, which might also influence mass failure. Recently, the application of GIS technologies has been increasing in the identification of regions susceptible to landslides (landslide zonation) (Chen and Zhang, 2021; Gutierrez-Martin, 2020; Li et al., 2022b). These methods are essential in emergency management because they provide a general overview of zones with a higher probability of landslide occurrence; however, they do not put emphasis on the determination of the approximate value of the volume of failing mass in relation to excessive rainfall events.”

**Comment 3:** The study area section should be enhanced with more detailed information on landslide-triggering factors. Additionally, it would be beneficial to incorporate a figure showing representative rainfall characteristics prior to the recorded landslide events in different parts of the Korean Peninsula. This would help better understand the unique rainfall patterns of the region responsible for landslides.

**Response:** We appreciate this fruitful comment. As recommended, we have included a rainfall plot within the data subsection in the revised manuscript to ensure logical coherence. Furthermore, we have combined the study area and data into a single section with two subsections for improved clarity. The updated information on the study area and the figure illustrating the rainfall characteristics prior to the recorded landslide events are provided below,

“The region for testing the model is South Korea, characterized by mountainous (63% of total land) relief, especially in the eastern part of the country (Lee et al., 2022). South Korea is located on the southern part of the Korean Peninsula, bordered by the Yellow Sea to the west coast and the East Sea (Sea of Japan) to the East. According to the Korean Meteorological Administration (2020), the country has a temperate climate characterized by four distinct seasons: hot and humid summers, cold winters, and springs and falls with moderate temperatures. The annual rainfall ranges between 1000 mm to 1400mm and 1800mm for the central region and southern region, respectively (Jung et al., 2017; Alcantara and Ahn, 2020). During the summer, heavy rainfall from June to September leads to significant surface runoff, increases landslide risk, and causes approximately 95% of all landslides each year (Lee et al., 2020; Park and Lee, 2021). In addition, the landslides may be aggravated by typhoons, which mostly occur in August and September, and it is anticipated that frequency will increase due to climate change (Kim and Park, 2021). The rainfall trend analysis from 1971 to 2100 predicted the increase in rainfall of 271.23mm, which indicates the growing risk of landslides associated with climate change (Lee, 2016). Temperature variations are influenced by its geographical location, the average summer temperatures range between 25 and 30°C, while winter temperatures can drop to -10°C in some parts of the country (Korea Meteorological Administration, 2020). The South Korean geologically is mainly composed of granitic and metamorphic rocks, such as gneiss, schist, and granite, which influence the stability of the landscape (Jung et al., 2024). The geomorphology is characterized by rugged mountains, river

valleys, and coastal plains, with the Taebaek Mountains running along the eastern edge (Kim et al., 2020). In addition, the influence of rainfall, environmental, geomorphology, and geological factors increase the vulnerability to landslides across the country, especially in the northeastern mountainous region, as depicted in Figure 1.

The predominant soil types in South Korea include clay, sandy, and loamy soils, each with different characteristics affecting water infiltration, retention and erosion (Kang et al., 2022; Lee et al., 2023). Clay soils, being more stable, can become highly saturated, increasing landslide risk during heavy rains. On the other hand, sandy soils are more prone to shallow landslides due to fast saturation, leading to instability. Regions with steep topography and poorly consolidated soil (loose) are mostly at risk, especially after prolonged rainfalls (Kim et al., 2015).

Coastal areas are exposed to sea-level rise and coastal erosion, which can further complicate the landscape and increase landslide susceptibility. The combination of heavy summer rainfall, geological composition, and geomorphological factors makes South Korea particularly vulnerable to shallow landslides. Thus, continuous monitoring and research are vital to understand the complex interactions between climate, geology, soil types, and landslide occurrences in this region (Park, 2022). Understanding the combination of environmental, geological stability, and geomorphological features is crucial for developing effective disaster management strategies and enhancing public safety in landslide-prone areas. As climate change continues to impact rainfall patterns, South Korea faces ongoing challenges in mitigating landslide risks and protecting vulnerable communities.”

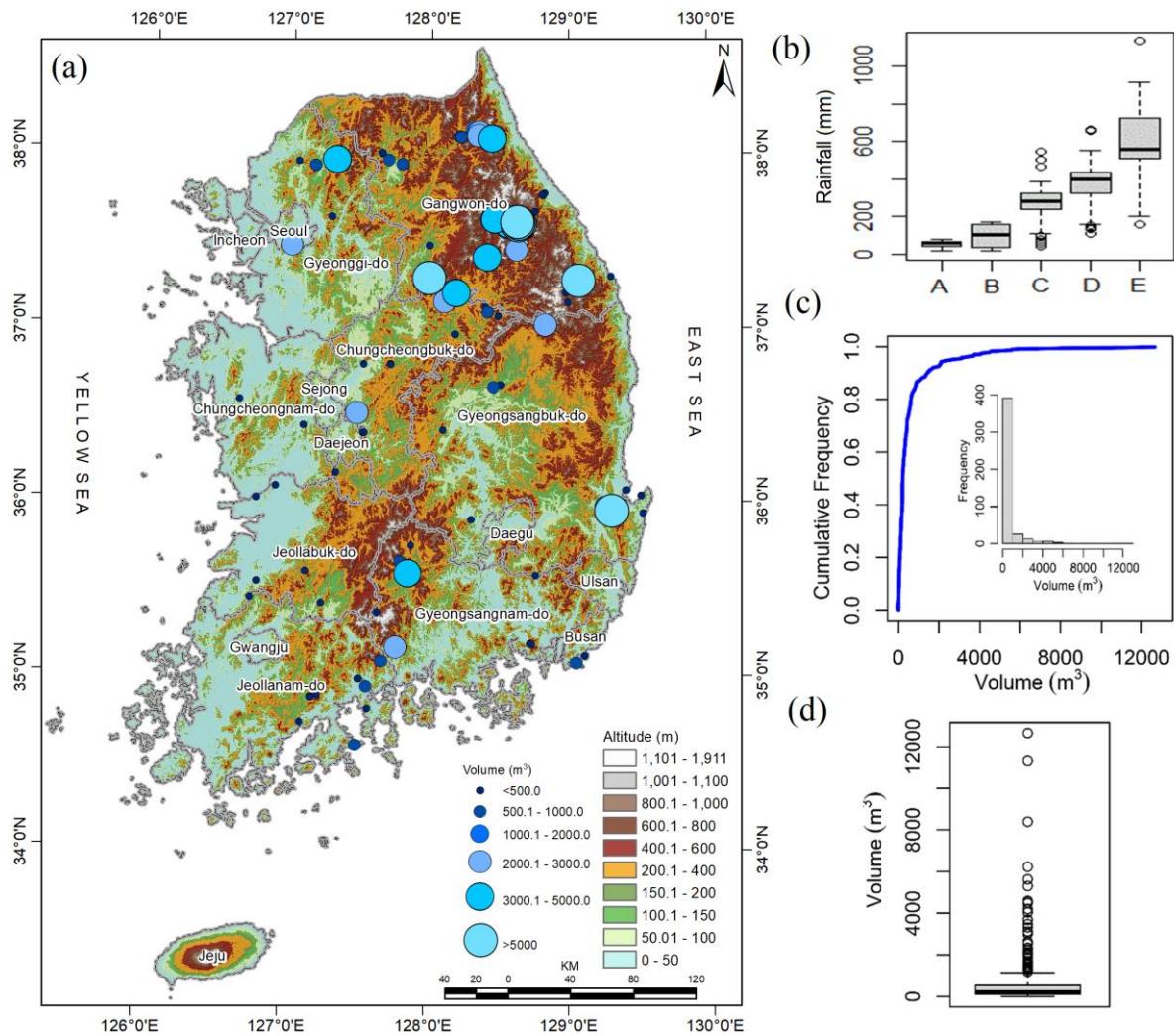


Figure 1. (a) Spatial distribution of landslides in South Korea, (b) temporal variation of rainfall, i.e., A: Maximum hourly rainfall, B: Four weeks rainfall, C: Three hours rainfall, D: Three days rainfall and E: Two weeks rainfall, (c) cumulative frequency distribution of volume of landslides and (d) box plot of volume of landslides.

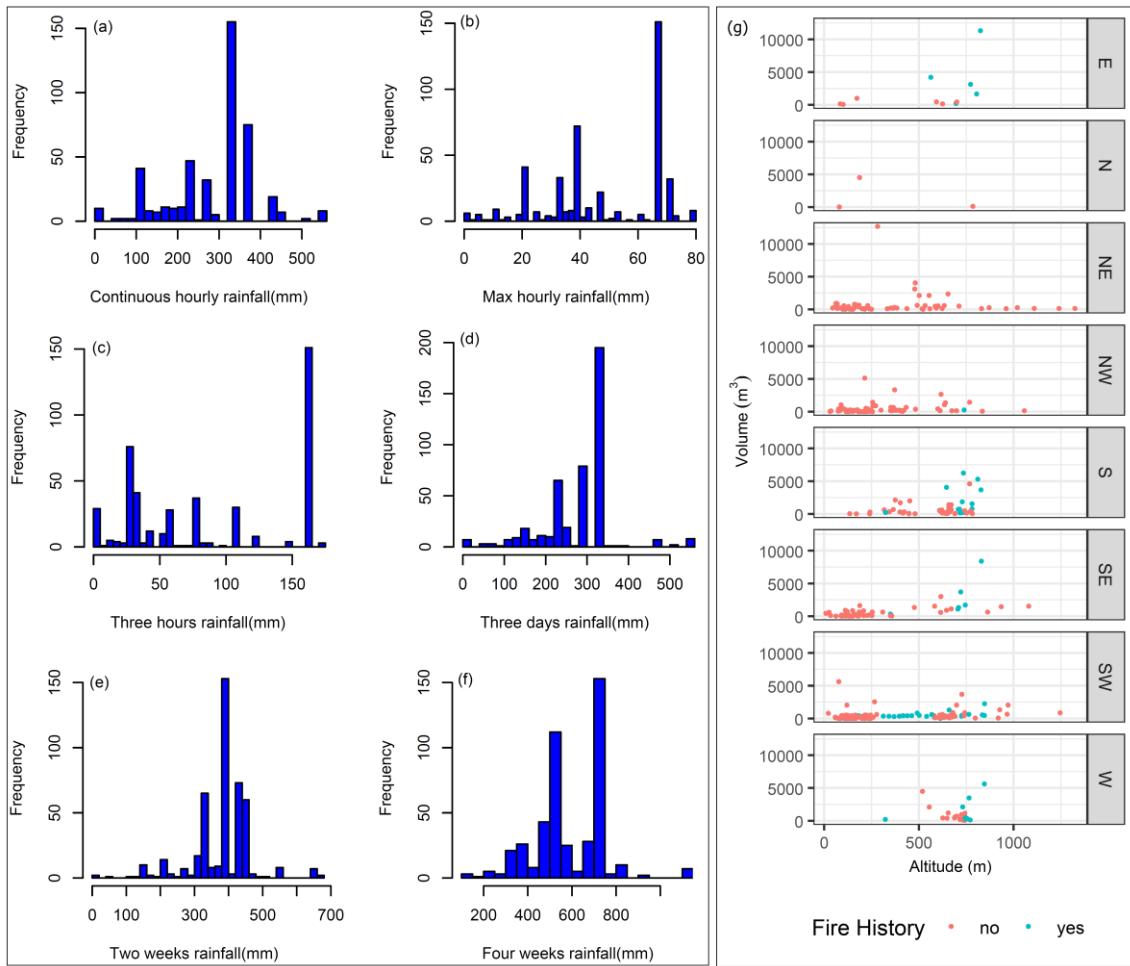


Figure 2. (a-f) Histograms of rainfall data, and (g) the scatter plot showing the variation of landslide volumes with respect to slope aspect, fire history and altitude.

Comment 4: Figure 2 needs to be updated. In the predictor variables, the authors should clearly specify which factors are influencing factors and which are triggering factors.

Response: Thank you for this observation. We agreed with the reviewer that the workflow had some missing information. Accordingly, the workflow figure was updated to reflect the reviewer's comment, and the model training and testing part was restructured to make it clearer. The updated Figure 3 (previously Figure 2 in the pre-print) is provided below,

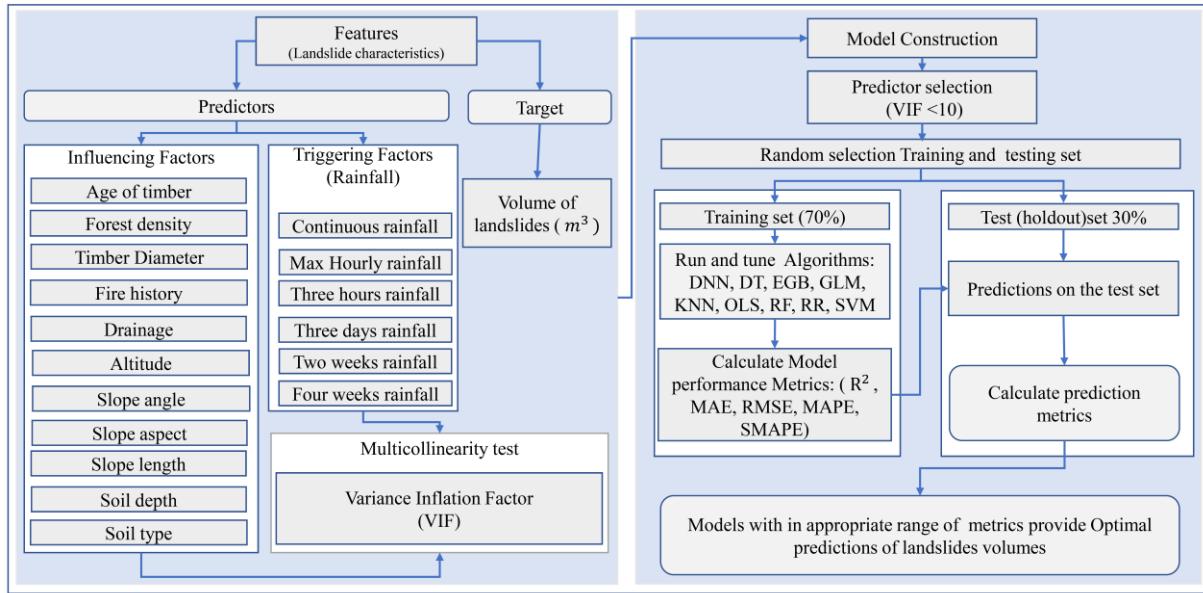


Figure 3. Workflow for the prediction of the volume of landslides due to rainfall.

Comment 5: A more detailed discussion of the input variables considered for volume prediction is required to better understand their roles as influencing and triggering factors. Additionally, the manuscript should provide further justification for the selection of these predictor variables.

Response: Thank you for your insightful comment. The details about the input variable are summarized in the data part of the manuscript, and Table 1 provides the reference justifying the reason for considering the stated feature as an input variable of the model. Accordingly, we have thoroughly revised section 2.2 (i.e., Data) in the revised manuscript as follows,

The landslide inventory dataset contains 450 landslide record information from 2011 to 2012, collected from different locations in South Korea by Korean Forest Services. This dataset tabulates information on landslide geometry, such as runout length, width, depth, and volume of the affected area, along with geomorphological composition, vegetation, and antecedent rainfall prior to landslide events. The details regarding landslide predisposing and triggering factors are summarized in Table 1.

Table 1. Landslide influencing and triggering factors.

Group	Features	Description	Reference
Vegetation	Fire history	The burning of the vegetation intensifies the mass movement of soil near the uncovered burned stem of trees and free movement on	Highland and Bobrowsky, 2008; Culler et al., 2021;

Group	Features	Description	Reference
Geomorphology	uncovered soil due to post-fire rainfall and storms. The sliding may also be due to loss of vegetation, altered soil property and structure, which lead to soil degradation and infiltration which increase pore pressure, and change in hydrology by concentrating water flow in places that exacerbate landslides.	Hyde et al., 2016; Stoof et al., 2012	
	Age of tree	Mature forests have more resistance to shallow landslides due to highly developed roots, which improve soil cohesion and leaves that prevent direct contact of raindrops with the soil surface.	Sato et al., 2023; Lann et al., 2024
	Forest density	The presence of forest reduces the likelihood of landslides about three times compared to grassland. Grassland has been revealed to be three times more vulnerable to shallow landslides than broadleaf and, coniferous and in secondary forests.	Lann et al., 2024; Greenwood et al., 2004; Turner et al., 2010; Scheidl et al., 2020; Asada et al., 2023
	Timber diameter (m)	Tree spacing and size had been used to investigate the effect of root and tree in shallow landslide control. The high root density generally enhances slope stability, and specific tree placement and root sizes between 5 to 20 mm are effective in landslide prevention.	Cohen and Schwarz., 2017; Wang et al., 2016
	Drainage	The drainage has a significant effect on the slope stability and promotes the efficient control of the influence of rainfall on the ground water fluctuation. The presence of drainage increases the threshold of landslides due to rainfall.	Yan et al., 2019; Sun et al., 2010 ; Wei et al., 2019 ; Korup et al., 2007
	Slope angle (degree)	The steeper slopes have lower presence of landslide due to low transportable materials. Slopes between 20-40 degrees are most vulnerable to greater landslides as rainfall intensity and duration increase. Here, we considered the average angle of the terrain at the landslide location, which provides valuable insight into the region's overall steepness and geomorphic characteristics, which are crucial factors influencing landslide	Duc, 2013 ; Qiu et al., 2016 ; Donnarumma et al., 2013

Group	Features	Description	Reference
susceptibility and risk modeling.			
	Slope aspect	The effect of rainfall on slope differs by slope angle and slope aspect which lead to unevenly distributed occurrence of landslides.	Panday and Dong, 2021; Cellek, 2021
	Slope length (m)	The volume increases as the slope length increases. There exists a complex interplay between rainfall, length of slope and slope angle on the occurrence of landslides.	Turner et al., 2010
	Soil depth (m)	Soil properties, depth, and texture have significant differences in infiltration rates, which have different influences on the occurrence of landslides.	Kitutu et al., 2009; McKenna et al., 2012
	Soil type	Higher rainfall intensity affects the occurrence of landslides differently, particularly in certain soil types that have shorter saturation and failure times.	Liu et al., 2021
Location	Altitude	Regional variability of elevation and mountain steepness affect the quantity of rainfall and associated landslides.	Hyun et al, 2010, Yoon and Bae, 2013; Park, 2015 Um et al., 2010
	Maximum hourly rainfall	The rainfall infiltrates the slope and increases pore water pressure that reduces soil shear strength, which leads to soil saturation that causes surface failure.	Wieczorek, 1987; Smith et al., 2023; Dai and Lee, 2001
Rainfall	Continuous rainfall	Sudden intense rainfall concentrated in short periods of time is responsible for shallow landslide and debris flow.	Zhang et al., 2019
	Three hours rainfall		
	Three days rainfall		Ran et al., 2022
	Two weeks rainfall	The antecedent rainfalls increase moisture in the soil and weaken soil cohesion.	Zhang et al., 2019; Bernardie et al., 2014; Chen et al., 2015a; Gariano et al., 2017
	Four weeks rainfall		

Location parameters such as altitude, latitude and longitude are essential elements that determine the microclimate of a given region, influencing rainfall patterns (Hyun et al., 2010; Yoon and Bae, 2013; Park, 2015). The northeastern region is characterized by high-elevation terrain, such as Taebaek, and Sobaek ranges, which dry air and lead to orographic precipitation

(Yun et al., 2009). The windward mountain versants receive a substantial amount of rainfall, which can increase the likelihood of landslides (Jin et al., 2022). This variation of rainfall with respect to the direction highlights the importance of including slope aspect variables in landslide studies (Kunz and Kottmeier, 2006). Figure 2(g) depicts the relationship between the slope aspect and the volume of landslides and slope aspect, altitude and fire history and shows that larger volumes were localized in regions that faced forest fire and altitudes between 500 and 1000m. Additionally, the topographical features such as slope length and slope angle affect the size of the landslide (Panday and Dong, 2021), slope failure due to over-saturation from groundwater and rainfall infiltration that destabilize the slope (Kafle et al., 2022). Furthermore, slope length, slope angle and slope aspect play an important role in the determination of the volume of geological material uprooted by landslides (Zaruba and Mencl, 2014; Khan et al., 2021). The slope stability depends on soil composition properties, including soil permeability indices that affect water infiltration and saturation level (Chen et al., 2015a). From surveyed regions, three main soil types, namely, sandy loam, loam, and silt loam, were observed, and their coefficient of permeability is 1.7, 1.65 and 1.5, respectively (Lee et al., 2013). Moreover, to reduce the infiltration drainage network that channeling rainwater terrain drains soil and reduces the saturation, which minimizes the likelihood of landslide occurrence as a result of groundwater discharge and rainfall water flow (Hovius et al., 1997; Wei et al., 2019). Furthermore, the vegetation protects the topsoil from the direct impact of raindrops hitting the ground, which causes erosion due to the force of gravity and reduces infiltration (Omwega, 1989; Keefer, 2000). The absence of vegetation allows rainwater to seep away fine topsoil, causing shallow landslides (Gonzalez-Ollauri and Mickovski, 2017). On the contrary, vegetation improves soil cohesion and prevents potential shallow landslides due to soil-root interaction (Gong et al., 2021; Phillips et al., 2021). The density of vegetation (forest) and leafage type (broad, pines or mixture) directly affects the quantity of raindrop intercepted and prevented from directly hitting the soil, which emphasizes the vegetation's landslides mitigation role. Further, the occurrence of forest fires can contribute to the occurrence of landslides due to the burning of vegetation covering the area, changing soil properties and increasing soil pH (Lee et al., 2013).

The rainfall, a triggering factor of landslides, is the immediate cause of slope instability and failure due to infiltration that leads to saturation resulting from increased pore water pressure that reduces soil shear strength (Yune et al., 2010; Khan et al., 2012; Kim et al., 2021; Lee et al., 2021). The antecedent rainfall increases the moisture in the soil, which accelerates

the soil saturation; the cumulative effect is essential to understand the saturation levels (Ran et al.,2022). In this study, rainfall variables are grouped based on time, namely, continuous rainfall, which is the accumulative value of rainfall on the day of a landslide from rainfall start hour to the landslide event, maximum hourly rainfall, rainfall during the fixed period such as three hours, one day, three days, two weeks etc.

Comment 6: I recommend providing clearer details on the geometry of the landslide inventory.

Response: Thank you for this recommendation. The landslide inventory provided by the Korea Forest Service (KFS) contains 455 landslide records (point locations) information from 2011 to 2012 within the triggering area. This dataset tabulates information on landslide geometry, such as runout length, width, depth, and volume of the affected area, along with geomorphological composition, vegetation, and antecedent rainfall prior to landslide events, which are integral to understanding the spatial extent and impact of each landslide. These geometric details have been incorporated into our analysis to represent landslide characteristics accurately. Accordingly, the data section has been revised in the updated manuscript as follows:

“The landslide inventory dataset contains 455 landslide record information from 2011 to 2012, collected from different locations in South Korea by Korean Forest Services. This dataset tabulates information on landslide geometry, such as runout length, width, depth, and volume of the affected area, along with geomorphological composition, vegetation, and antecedent rainfall prior to landslide events.”

Comment 7: A brief discussion on why nine data-driven models were chosen is recommended in the methods section. While these models have become quite common, providing a rationale for their selection will help justify their use in the study.

Response: Thank you for your valuable feedback. The inclusion of model selection is essential to justify the selection basis. In the present study, we aimed at predicting the volume of landslides using models that minimize error with interpretability and scalability. Since one model can not have all properties at the same time, we decided to select some of the models with those properties. The OLS, GLM, and DT were selected for their high interpretability, which helps to understand the influence of individual features on predictions (Gelman, 2007; Breiman, 2017). On the other hand, the EGB, RF, SVM, RR, and KNN were chosen due to

their robust performance in capturing complex patterns in data, which is essential for accurate predictions of landslide volumes (Chen and Guestrin, 2016; Liaw and Wiener, 2002; Hastie et al., 2009). Additionally, taking into account that the model will be used for regional scale, which will require the use of big data, the EGB, RF, DNN are designed to efficiently handle large datasets, making them suitable for the regional scale analysis. These last models can be scaled to incorporate more data from different geographical areas without significant adjustments, enhancing their applicability in future research (Krizhevsky et al., 2012).

Comment 8: The authors mainly use MAE and  $R^2$  for model validation. It is recommended to consider additional metrics commonly used in data-driven model evaluation. Relying solely on these two statistics may not comprehensively assess model performance.

Response: Thank you for your valuable suggestion. We have expanded the metrics used for model validation to provide a more comprehensive assessment of the model performance. In addition to Mean Absolute Error (MAE) and  $R^2$ , we have included additional metrics such as Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), and symmetric mean absolute percentage errors (SMAPE). These metrics will offer a broader perspective on the accuracy and reliability of our predictions. The updated section detailing the selected metrics is included in the revised manuscript as,

“The model performance evaluation is a process of quantifying the difference between the observed value not used in the modeling process and the predicted value by the model. Different metrics are applied depending on the type of task, whether it is a classification or a regression problem. Subsequently, the widely used evaluation metrics for regression models, namely,  $R^2$ , MAE, RMSE, MAPE and SMAPE, were utilized to evaluate the model performances. The metric formulae and evaluation criteria are summarized in Table 3.

**Table 3.** Model evaluation metrics.

Metrics	Evaluation	Reference
$RMSE$ $= \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$	<ul style="list-style-type: none"> <li>• Measures the square root of the average squared differences between predicted and actual values.</li> <li>• Lower values indicate better model performance.</li> </ul>	Hyndman and Koehler, 2006.

$MAE = \frac{1}{n} \sum_{i=1}^n  y_i - \hat{y}_i $	<ul style="list-style-type: none"> <li>The average of the absolute differences between predicted and actual values.</li> <li>Lower values indicate better model performance.</li> </ul>	Willmott and Matsuura, 2005.
$MAPE = \frac{100}{n} \sum_{i=1}^n \left  \frac{y_i - \hat{y}_i}{y_i} \right $	<ul style="list-style-type: none"> <li>Measures the accuracy of a model as a percentage, which can be more interpretable.</li> <li>Lower values indicate better model performance.</li> </ul>	Armstrong, 2001.
$SMAPE$ $= \frac{100}{n} \sum_{i=1}^n \frac{ y_i - \hat{y}_i }{ y_i  -  \hat{y}_i }$	<ul style="list-style-type: none"> <li>Unlike MAPE, which can be skewed by very small actual values, SMAPE accounts for both the actual and predicted values, making it symmetric.</li> <li>SMAPE is expressed as a percentage</li> <li>Mitigates the impact of small actual values on the error metric, providing a more balanced assessment.</li> <li>Lower values indicate better model performance.</li> </ul>	Hyndman and Koehler, 2006
$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$	<ul style="list-style-type: none"> <li>Represents the proportion of variance in the dependent variable that can be explained by the independent variables.</li> <li>Values closer to 1 indicate a better fit</li> </ul>	Darlington, 1990; Chicco et al., 2021

\* $y_i$  and  $\hat{y}_i$  representing the actual and predicted value and,  $\bar{y}$  and  $n$  standing for the mean of actual value and number of observations in the dataset, respectively.

Comment 9: The summary of the various data-driven models (Table 3) indicates that the EGB model is the best-performing. However, the variable importance analysis shown in Figure 5 highlights only a subset of predictor variables, raising questions about whether different models utilize different sets of features. Further clarification is needed.

Response: We appreciate your insightful comment regarding the inconsistencies between the summary of the various data-driven models in Table 3 (now Table 4) and the variable importance analysis presented in Figure 5 (now Figure 6). In the earlier version of the manuscript, the variable importance features with a value of gain below 0.01 were removed from the plot. To avoid those inconsistencies and make clear all variables, the figure was updated and reflect all variables used in the updated manuscript depicted below,

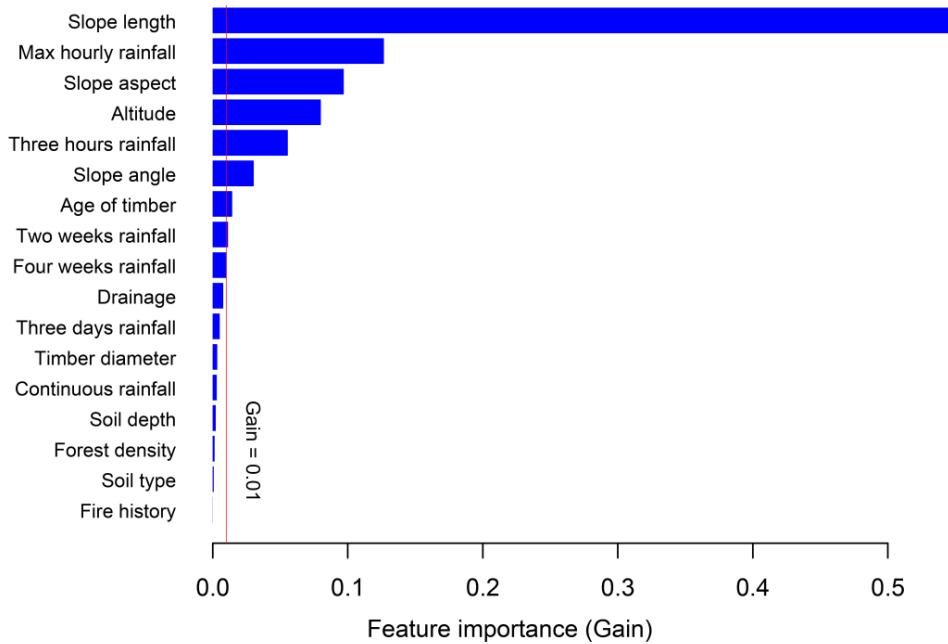


Figure 6. Variable importance for the EGB model.

### Specific Comments

Comment 1: Figure 1(b): The y-axis label is missing.

Response: We appreciate this observation that helped us to improve the clarity of Fig 1(b), which was missing the y-axis. We have corrected this oversight, and the updated Figure 1 now includes the y-axis label. Figure 1(b) illustrates the boxplot of various rainfall features utilized in the model. The revised figure is provided below,

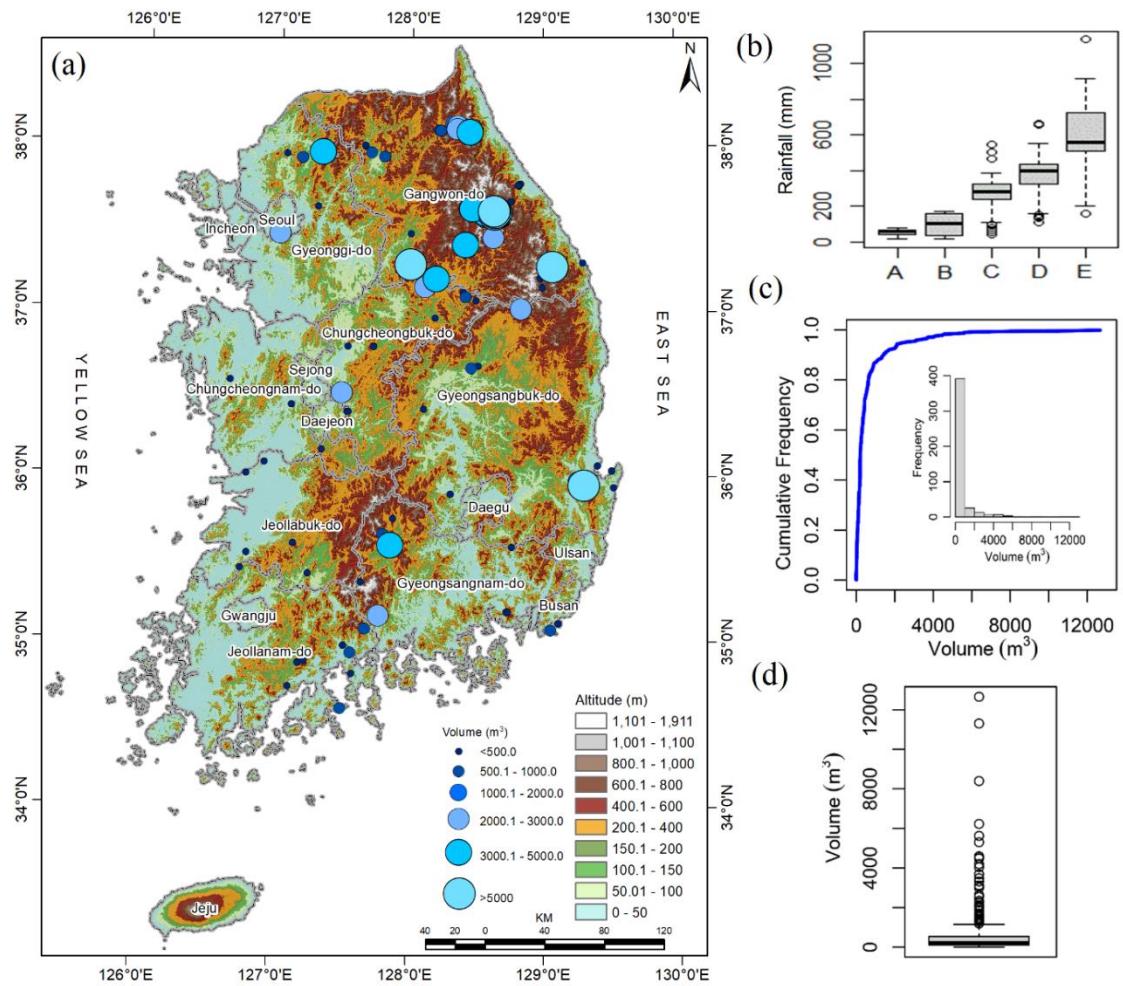


Figure 1. (a) Spatial distribution of landslides in South Korea, (b) temporal variation of rainfall, i.e., A: Maximum hourly rainfall, B: Four weeks rainfall, C: Three hours rainfall, D: Three days rainfall and E: Two weeks rainfall, (c) cumulative frequency distribution of volume of landslides and (d) box plot of volume of landslides.

Comment 2: In Figure 2, it would be better to use the terms ‘Training and Testing Algorithms’ instead of ‘Run and Test Algorithms’. This terminology more accurately reflects the standard processes involved in model development.

Response: Thank you for your insightful observation. Figure 2 (now Figure 3) has been updated in the revised manuscript as follows,

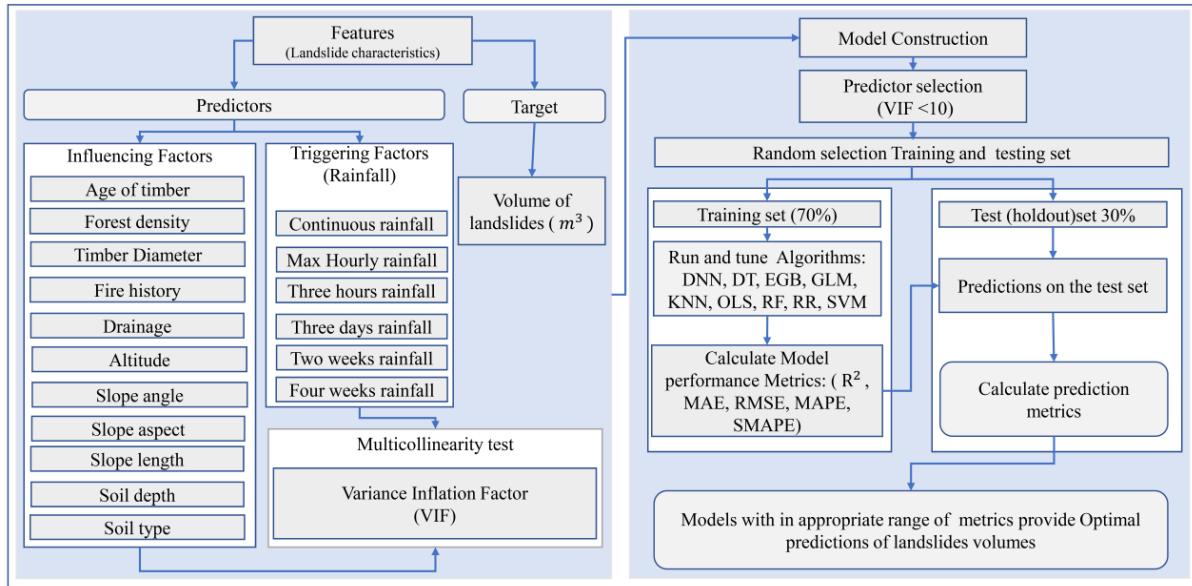


Figure 3. Workflow for the prediction of volume of landslides due to rainfall.

Comment 3: Line No. 104-107: I recommend that the authors use the acronyms for the different data-driven models here, as they have already been defined earlier in the manuscript. Consistent use of these acronyms throughout the manuscript will improve clarity and readability.

Response: We appreciate the comment regarding the consistency in the use of acronyms in the manuscript to keep the clarity. The suggested consistency in the use of acronyms was adopted in the updated manuscript.

Comment 4: Line No. 150-152: ‘Thus, planting vegetation is recommended as a better practice to improve soil cohesion and prevent potential landslides due to soil root interaction (Gong et al., 2017; Phillips et al., 2021)’. This is a recommendation, not a description. Please provide appropriate descriptions.

Response: Thank you for your comment. We understand the misuse of English connectives. The sentence was modified in the revised manuscript as follows,

“The absence of vegetation allows rainwater to seep away fine topsoil, causing shallow landslides (Gonzalez-Ollauri and Mickovski, 2017). On the contrary, vegetation improves soil cohesion and prevents potential shallow landslides due to soil-root interaction (Gong et al., 2021; Phillips et al., 2021).”

Comment 5: Why did the authors use 70% of the inventory and 30% for the validation? Why not 50% for each? The authors should state in the methods section why they used these percentages. Further, the authors need to clarify whether training and testing data were chosen randomly or if any specific criteria were used for the analysis.

Response: Thank you for your insightful suggestion. The training and testing were split randomly to train the model. The 10-fold cross-validation was performed to obtain an optimal model. This division aligns with common practices in machine learning, where a 70 (training): 30 (test) ratio is frequently used to ensure adequate training data while reserving a sufficient amount for model testing. In addition, Nguyen et al. (2021) also highlighted that 0.7/0.3 is the optimal split for the data.

Comment 6: Please review the references cited in the text, as there are frequent errors with the use of commas and semicolons between references. This issue occurs multiple times throughout the manuscript and needs correction for proper citation formatting.

Response: We thank you for pointing out the mis-references in the manuscript. We understand the concern regarding improper citations, and we carefully read and corrected all misreferencing.

Comment 7: Chen et al. (2015) is not cited correctly in the text. There are two different articles by Chen et al. (2015) listed in the references section. The authors need to distinguish between these references by specifying them as Chen et al. (2015a) and Chen et al. (2015b). The reference section and citation in the text should be updated accordingly to reflect these distinctions.

Response: We appreciate the observation regarding the two authors with the same name, and the correction has been done accordingly, as follows,

Chen, Z., Luo, R., Huang, Z., Tu, W., Chen, J., Li, W., ... and Ai, Y. (2015a). Effects of different backfill soils on artificial soil quality for cut slope revegetation: Soil structure, soil erosion, moisture retention and soil C stock. *Ecological engineering*, 83, 5-12.

Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., ... and Zhou, T. (2015b). Xgboost: extreme gradient boosting. *R package version 0.4-2*, 1(4), 1-4.

Comment 8: Line 133 cites (Kafle, 2022), but this article is either missing from the reference

section or is not cited correctly. Please verify and ensure that this reference is properly included and formatted in the reference section.

Response: The citation was mis-referenced. The citation has been corrected to "Kafle et al. (2022)" and is now properly included and formatted in the reference section of the revised manuscript.

Kafle, L., Xu, W. J., Zeng, S. Y., & Nagel, T. (2022). A numerical investigation of slope stability influenced by the combined effects of reservoir water level fluctuations and precipitation: A case study of the Bianjiazhai landslide in China. *Engineering Geology*, 297, 106508.

Comment 9: Line no 215: Chowdhury (2023)- article not present in the reference section or not mentioned in the correct form.

Response: Thank you for your comment. The reference has been updated to "Chowdhury et al. (2023)" in the revised manuscript, and it is now correctly included in the reference section as follows,

Chowdhury, M. Z. I., Leung, A. A., Walker, R. L., Sikdar, K. C., O'Beirne, M., Quan, H., and Turin, T. C. (2023). A comparison of machine learning algorithms and traditional regression-based statistical modeling for predicting hypertension incidence in a Canadian population. *Scientific Reports*, 13(1), 13.

Comment 10: Line no 236: (Team, 2022)- article is not present in the reference section or not mentioned in the correct form.

Response: Thank you for pointing out the citation issue. The reference has been corrected to "R Core Team (2022)" in the revised manuscript. The updated reference is provided below:

"R core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: <<https://www.R-project.org/>>."

Comment 11: Line no 239: Jerome et al. (2012)- article not present in the reference section or not mentioned in the correct form.

Response: We appreciate your observation regarding the mis-reference in line 239 of the

manuscript concerning Jerome et al. (2012). The reference was updated as “Jerome et al. (2010)” in the updated manuscript. The updated citation is as follows,

Friedman, J. H., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of statistical software*, 33, 1-22. [URL:<https://www.jstatsoft.org/v33/i01/>.](https://www.jstatsoft.org/v33/i01/)”

Comment 12: Please verify the unit of soil depth. Most landslide inventories presented in Table 2 and Figure 1(d) suggest that the landslides are shallow-seated based on their volume distributions; the unit of soil depth does not seem to align with this observation.

Response: We thank the reviewer for highlighting the mistake in the soil depth unit. The correct unit for soil depth is centimeters (cm). However, for consistency and clarity, we have transformed the soil depth data into meters (m) and updated the manuscript accordingly.

Comment 13: I suggest adding a few lines in the discussion section to highlight the practical applicability of the proposed model. This would provide insight into how the model can be used in real-world scenarios and its potential impact on practice or policy.

Response: Thank you for your suggestion regarding the practical applicability of our proposed model. In the revised discussion section, we included a few lines highlighting how the model can be utilized in real-world scenarios.

We emphasized that the model’s ability to accurately predict landslide volumes can aid in disaster risk management by providing timely information for early warning systems. Additionally, the insights gained from the model can inform land-use planning and policy decisions, allowing stakeholders to identify high-risk areas and implement mitigation strategies effectively. By integrating the model into existing monitoring frameworks, agencies can enhance their response capabilities and better allocate resources during heavy rainfall events.

## References

Alcantara, A. L., and Ahn, K. H. (2020). Probability distribution and characterization of daily precipitation related to tropical cyclones over the Korean Peninsula. *Water*, 12(4), 1214.

Alcántara-Ayala, I. (2021). Integrated landslide disaster risk management (ILDRiM): the challenge to avoid the construction of new disaster risk. *Environmental Hazards*, 20(3), 323-344.

Alcántara-Ayala, I., & Sassa, K. (2023). Landslide risk management: from hazard to disaster risk reduction. *Landslides*, 20(10), 2031-2037.

Amatya, S. C. (2016). Landslide disaster management in Nepal: a near-future perspective. Nepal-Japan Friendship Association of Water Induced Disaster (NFAD), Japan Department of Water Induced Disaster Management (DWIDM).

Armstrong, J. S. (2001). Combining forecasts (pp. 417-439). Springer US.

Asada, H., & Minagawa, T. (2023). Impact of vegetation differences on shallow landslides: a case study in Aso, Japan. *Water*, 15(18), 3193.

Barik, M. G., Adam, J. C., Barber, M. E., & Muhunthan, B. (2017). Improved landslide susceptibility prediction for sustainable forest management in an altered climate. *Engineering geology*, 230, 104-117.

Bernardie, S., Desramaut, N., Malet, J.-P., Gourlay, M., and Grandjean, G. (2014). Prediction of changes in landslide rates induced by rainfall. *Landslides*, 12(3), 481–494. doi:10.1007/s10346-014-0495-8

Breiman, L. (2017). Classification and regression trees. Routledge.

Cellek, S. (2021). The effect of aspect on landslide and its relationship with other parameters. In *Landslides*. IntechOpen.

Chang, K. T., and Chiang, S. H. (2009). An integrated model for predicting rainfall-induced landslides. *Geomorphology*, 105(3-4), 366-373.

Chatra, A. S., Dodagoudar, G. R., & Maji, V. B. (2019). Numerical modelling of rainfall effects on the stability of soil slopes. *International Journal of Geotechnical Engineering*.

Chen, C. W., Oguchi, T., Hayakawa, Y. S., Saito, H., and Chen, H. (2017). Relationship between landslide size and rainfall conditions in Taiwan. *Landslides*, 14, 1235-1240.

Chen, L., Guo, Z., Yin, K., Shrestha, D. P., & Jin, S. (2019). The influence of land use and land cover change on landslide susceptibility: a case study in Zhushan Town, Xuan'en County (Hubei, China). *Natural hazards and earth system sciences*, 19(10), 2207-2228.

Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 785-794).

Chen, X., Zhang, L., Zhang, L., Zhou, Y., Ye, G., & Guo, N. (2021). Modelling rainfall-induced landslides from initiation of instability to post-failure. *Computers and geotechnics*, 129, 103877.

Chen, Z., Luo, R., Huang, Z., Tu, W., Chen, J., Li, W., ... and Ai, Y. (2015a). Effects of different backfill soils on artificial soil quality for cut slope revegetation: Soil structure, soil erosion, moisture retention and soil C stock. *Ecological engineering*, 83, 5-12.

Cheung, R. W. (2021). Landslide risk management in Hong Kong. *Landslides*, 18(10), 3457-3473.

Chicco, D., Warrens, M. J., & Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ computer science*, 7, e623.

Cohen, D., & Schwarz, M. (2017). Tree-root control of shallow landslides. *Earth Surface Dynamics*, 5(3), 451-477.

Conte, E., Pugliese, L., and Troncone, A. (2022). A simple method for predicting rainfall-induced shallow landslides. *Journal of Geotechnical and Geoenvironmental Engineering*, 148(10), 04022079

Culler, E. S., Livneh, B., Rajagopalan, B., & Tiampo, K. F. (2021). A data-driven evaluation of post-fire landslide susceptibility. *Natural Hazards and Earth System Sciences Discussions*, 2021, 1-24. <https://doi.org/10.5194/nhess-23-1631-2023>

Dai, F. C., & Lee, C. F. (2001). Frequency–volume relation and prediction of rainfall-induced landslides. *Engineering geology*, 59(3-4), 253-266.

Dai, F. C., Lee, C. F., & Ngai, Y. Y. (2002). Landslide risk assessment and management: an overview. *Engineering geology*, 64(1), 65-87.

Dai, K., Xu, Q., Li, Z., Tomás, R., Fan, X., Dong, X., ... and Ran, P. (2019). Post-disaster assessment of 2017 catastrophic Xinmo landslide (China) by spaceborne SAR interferometry. *Landslides*, 16, 1189-1199.

Darlington, R. B. (1990). Regression and linear models. McGraw-Hill College.

Dikshit, A., Satyam, N., & Pradhan, B., 2019. Estimation of rainfall-induced landslides using the TRIGRS model. *Earth Systems and Environment*, 3, 575-584.

Donnarumma, A., Revellino, P., Grelle, G., and Guadagno, F. M. (2013). Slope angle as indicator parameter of landslide susceptibility in a geologically complex area. *Landslide Science and Practice: Volume 1: Landslide Inventory and Susceptibility and Hazard Zoning*, 425-433.

Duc, D. M. (2013). Rainfall-triggered large landslides on 15 December 2005 in Van Canh district, Binh Dinh province, Vietnam. *Landslides*, 10(2), 219-230.

Evans, S. G., Mugnozza, G. S., Strom, A., and Hermanns, R. L. (Eds.). (2007). *Landslides from massive rock slope failure (Vol. 49)*. Springer Science and Business Media.

Fan, X., Xu, Q., Liu, J., Subramanian, S. S., He, C., Zhu, X., & Zhou, L. (2019). Successful early warning and emergency response of a disastrous rockslide in Guizhou province, China. *Landslides*, 16, 2445-2457.

Fan, X., Xu, Q., Scaringi, G., Dai, L., Li, W., Dong, X., ... & Havenith, H. B. (2017). Failure mechanism and kinematics of the deadly June 24th 2017 Xinmo landslide, Maoxian, Sichuan, China. *Landslides*, 14, 2129-2146. <https://doi.org/10.1007/s10346-017-0907-7>

Gariano, S. L., Rianna, G., Petrucci, O., and Guzzetti, F. (2017). Assessing future changes in the occurrence of rainfall-induced landslides at a regional scale. *Science of the total environment*, 596, 417-426.

Gelman, A. (2007). Data analysis using regression and multilevel/hierarchical models. Cambridge University Press.

Gong, Q., Wang, J., Zhou, P., and Guo, M. (2021). A regional landslide stability analysis method under the combined impact of rainfall and vegetation roots in south China. *Advances in Civil Engineering*, 2021, 1-12.

Gonzalez-Ollauri, A., and Mickovski, S. B. (2017). Hydrological effect of vegetation against rainfall-induced landslides. *Journal of Hydrology*, 549, 374-387.

Greenwood, J. R., Norris, J. E., & Wint, J. (2004). Assessing the contribution of vegetation to slope stability. *Proceedings of the Institution of Civil Engineers-Geotechnical*

Engineering, 157(4), 199-207.

Gutierrez-Martin, A. (2020). A GIS-physically-based emergency methodology for predicting rainfall-induced shallow landslide zonation. *Geomorphology*, 359, 107121.

Guzzetti, F., Peruccacci, S., Rossi, M., & Stark, C. P. (2008). The rainfall intensity-duration control of shallow landslides and debris flows: an update. *Landslides*, 5, 3-17.

Hastie, T. (2009). The elements of statistical learning: data mining, inference, and prediction.

Highland, L. and Bobrowsky, P. (2008). The Landslide Handbook: A Guide to Understanding Landslides, United States Geological Survey, Reston, VA, Circular 1325, <https://pubs.usgs.gov/circ/1325/> (last access: 6 March 2023), 2008.

Holcombe, E. A., Beesley, M. E., Vardanega, P. J., & Sorbie, R. (2016, March). Urbanisation and landslides: hazard drivers and better practices. In *Proceedings of the Institution of Civil Engineers-Civil Engineering* (Vol. 169, No. 3, pp. 137-144). Thomas Telford Ltd.

Hovius, N., Stark, C. P., & Allen, P. A. (1997). Sediment flux from a mountain belt derived by landslide mapping. *Geology*, 25(3), 231-234.

Hyde, K. D., Riley, K., & Stoof, C. (2016). Uncertainties in predicting debris flow hazards following wildfire. *Natural hazards*

Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International journal of forecasting*, 22(4), 679-688.

Hyun, Y. K., Kar, S. K., Ha, K. J., & Lee, J. H. (2010). Diurnal and spatial variabilities of monsoonal CG lightning and precipitation and their association with the synoptic weather conditions over South Korea. *Theoretical and applied climatology*, 102, 43-60.

Intrieri, E., Carlà, T., and Gigli, G. (2019). Forecasting the time of failure of landslides at slope-scale: A literature review. *Earth-science reviews*, 193, 333-349.

Islam, M. A., Islam, M. S., & Islam, T. (2017, September). Landslides in Chittagong hill tracts and possible measures. In *Proceedings of the international conference on disaster risk mitigation*, Dhaka, Bangladesh (Vol. 23).

Jin, H. G., Lee, H., & Baik, J. J. (2022). Characteristics and possible mechanisms of diurnal variation of summertime precipitation in South Korea. *Theoretical and Applied Climatology*, 148(1), 551-568.

Ju, L. Y., Zhang, L. M., and Xiao, T. (2023). Power laws for accurate determination of landslide volume based on high-resolution LiDAR data. *Engineering Geology*, 312, 106935.

Jung, M. J., Jeong, Y. J., Shin, W. J., & Cheong, A. C. S. (2024). Isotopic distribution of bioavailable Sr, Nd, and Pb in Chungcheongbuk-do Province, Korea. *Journal of Analytical Science and Technology*, 15(1), 46.

Jung, Y., Shin, J. Y., Ahn, H., and Heo, J. H. (2017). The spatial and temporal structure of extreme rainfall trends in South Korea. *Water*, 9(10), 809.

Kafle, L., Xu, W. J., Zeng, S. Y., & Nagel, T. (2022). A numerical investigation of slope stability influenced by the combined effects of reservoir water level fluctuations and precipitation: A case study of the Bianjiazhai landslide in China. *Engineering Geology*, 297, 106508.

Kang, M. W., Yibeltal, M., Kim, Y. H., Oh, S. J., Lee, J. C., Kwon, E. E., & Lee, S. S. (2022).

Enhancement of soil physical properties and soil water retention with biochar-based soil amendments. *Science of the total environment*, 836, 155746.

Keefer, R. F. (2000). *Handbook of soils for landscape architects*. Oxford University Press.

Khan, M. A., Basharat, M., Riaz, M. T., Sarfraz, Y., Farooq, M., Khan, A. Y., ... and Shahzad, A. (2021). An integrated geotechnical and geophysical investigation of a catastrophic landslide in the Northeast Himalayas of Pakistan. *Geological Journal*, 56(9), 4760-4778

Khan, Y. A., Lateh, H., Baten, M. A., and Kamil, A. A. (2012). Critical antecedent rainfall conditions for shallow landslides in Chittagong City of Bangladesh. *Environmental Earth Sciences*, 67, 97-106.

Kim, D. E., Seong, Y. B., Weber, J., & Yu, B. Y. (2020). Unsteady migration of Taebaek Mountain drainage divide, Cenozoic extensional basin margin, Korean Peninsula. *Geomorphology*, 352, 107012.

Kim, H. G., & Park, C. Y. (2021). Landslide susceptibility analysis of photovoltaic power stations in Gangwon-do, Republic of Korea. *Geomatics, Natural Hazards and Risk*, 12(1), 2328-2351.

Kim, J., Lee, K., Jeong, S., & Kim, G. (2014). GIS-based prediction method of landslide susceptibility using a rainfall infiltration-groundwater flow model. *Engineering geology*, 182, 63-78.

Kim, M. S., Onda, Y., Kim, J. K., & Kim, S. W. (2015). Effect of topography and soil parameterisation representing soil thicknesses on shallow landslide modelling. *Quaternary International*, 384, 91-106.

Kim, S. W., Chun, K. W., Kim, M., Catani, F., Choi, B., and Seo, J. I. (2021). Effect of antecedent rainfall conditions and their variations on shallow landslide-triggering rainfall thresholds in South Korea. *Landslides*, 18, 569-582.

Kitutu, M. G., Muwanga, A., Poesen, J., and Deckers, J. A. (2009). Influence of soil properties on landslide occurrences in Bududa district, Eastern Uganda. *African journal of agricultural research*, 4(7), 611-620.

Klimeš, J., Stemberk, J., Blahut, J., Krejčí, V., Krejčí, O., Hartvich, F., & Kycl, P. (2017). Challenges for landslide hazard and risk management in 'low-risk' regions, Czech Republic—landslide occurrences and related costs (IPL project no. 197). *Landslides*, 14, 771-780

Korup, O., Clague, J. J., Hermanns, R. L., Hewitt, K., Strom, A. L., and Weidinger, J. T. (2007). Giant landslides, topography, and erosion. *Earth and Planetary Science Letters*, 261(3-4), 578-589.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25.

Kunz, M., & Kottmeier, C. (2006). Orographic enhancement of precipitation over low mountain ranges. Part II: Simulations of heavy precipitation events over southwest Germany. *Journal of applied meteorology and climatology*, 45(8), 1041-1055.

Lacerda, W. A., Palmeira, E. M., Netto, A. L. C., & Ehrlich, M. (Eds.). (2014). *Extreme rainfall induced landslides: an international perspective*. Oficina de Textos.

Lann, T., Bao, H., Lan, H., Zheng, H., & Yan, C. (2024). Hydro-mechanical effects of vegetation on slope stability: A review. *Science of the Total Environment*, 171691.

Lee, D. B., Kim, Y. N., Sonn, Y. K., & Kim, K. H. (2023). Comparison of Soil Taxonomy (2022) and WRB (2022) Systems for classifying Paddy Soils with different drainage grades in South Korea. *Land*, 12(6), 1204.

Lee, D. H., Cheon, E., Lim, H. H., Choi, S. K., Kim, Y. T., & Lee, S. R. (2021). An artificial neural network model to predict debris-flow volumes caused by extreme rainfall in the central region of South Korea. *Engineering Geology*, 281, 105979.

Lee, D. H., Cheon, E., Lim, H. H., Choi, S. K., Kim, Y. T., and Lee, S. R. (2021). An artificial neural network model to predict debris-flow volumes caused by extreme rainfall in the central region of South Korea. *Engineering Geology*, 281, 105979.

Lee, D. H., Kim, Y. T., and Lee, S. R. (2020). Shallow landslide susceptibility models based on artificial neural networks considering the factor selection method and various non-linear activation functions. *Remote Sensing*, 12(7), 1194.

Lee, J. U., Cho, Y. C., Kim, M., Jang, S. J., Lee, J., & Kim, S. (2022). The effects of different geological conditions on landslide-triggering rainfall conditions in South Korea. *Water*, 14(13), 2051.

Lee, M. J. (2016). Rainfall and landslide correlation analysis and prediction of future rainfall base on climate change. In *Geohazards Caused by Human Activity*. IntechOpen.

Lee, S. G. (2009). The Effects of Landslide in South Korea and Some Issues for Successful Management and Mitigation. *한국토양비료학회 학술발표회 초록집*, 181-191.

Lee, S. W., Kim, G., Yune, C. Y., and Ryu, H. J. (2013). Development of landslide-risk assessment model for mountainous regions in eastern Korea. *Disaster advances*, 6(6), 70-79.

Li, B. V., Jenkins, C. N., & Xu, W. (2022a). Strategic protection of landslide vulnerable mountains for biodiversity conservation under land-cover and climate change impacts. *Proceedings of the National Academy of Sciences*, 119(2), e2113416118.

Li, C. J., Guo, C. X., Yang, X. G., Li, H. B., & Zhou, J. W. (2022b). A GIS-based probabilistic analysis model for rainfall-induced shallow landslides in mountainous areas. *Environmental Earth Sciences*, 81(17), 432.

Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R news*.

Liu, Y., Deng, Z., & Wang, X. (2021). The effects of rainfall, soil type and slope on the processes and mechanisms of rainfall-induced shallow landslides. *Applied Sciences*, 11(24), 11652.

Luino, F., De Graff, J., Biddoccu, M., Faccini, F., Freppaz, M., Roccati, A., ... & Turconi, L. (2022). The Role of soil type in triggering shallow landslides in the alps (Lombardy, Northern Italy). *Land*, 11(8), 1125.

Lusiana, N., Shinohara, Y., & Imaizumi, F. (2024). Quantifying effects of changes in forest age distribution on the landslide frequency in Japan. *Natural Hazards*, 1-20.

Martinović, K., Gavin, K., Reale, C., and Mangan, C. (2018). Rainfall thresholds as a landslide indicator for engineered slopes on the Irish Rail network. *Geomorphology*, 306, 40-50.

Mateos, R. M., López-Vinielles, J., Poyiadji, E., Tsagkas, D., Sheehy, M., Hadjicharalambous, K., ... & Herrera, G. (2020). Integration of landslide hazard into urban planning across Europe. *Landscape and urban planning*, 196, 103740.

McKenna, J. P., Santi, P. M., Amblard, X., and Negri, J. (2012). Effects of soil-engineering properties on the failure mode of shallow landslides. *Landslides*, 9, 215-228.

Nguyen, Q. H., Ly, H. B., Ho, L. S., Al-Ansari, N., Le, H. V., Tran, V. Q., ... & Pham, B. T. (2021). Influence of data splitting on performance of machine learning models in prediction of shear strength of soil. *Mathematical Problems in Engineering*, 2021(1), 4832864.

Omwega, A. K. (1989). Crop cover, rainfall energy and soil erosion in Githunguri (Kiambu District), Kenya. The University of Manchester (United Kingdom).

Panday, S., & Dong, J. J. (2021). Topographical features of rainfall-triggered landslides in Mon State, Myanmar, August 2019: Spatial distribution heterogeneity and uncommon large relative heights. *Landslides*, 18(12), 3875-3889.

Park, C. Y. (2015). The classification of extreme climate events in the Republic of Korea. *Journal of the Korean association of regional geographers*, 21(2), 394-410.

Park, S. J., & Lee, D. K. (2021). Predicting susceptibility to landslides under climate change impacts in metropolitan areas of South Korea using machine learning. *Geomatics, Natural Hazards and Risk*, 12(1), 2462-2476.

Park, S.J. (2022). Assessment of disaster risks induced by climate change, using machine learning techniques (Doctoral dissertation, 서울대학교 대학원).

Paudel, P. P., Omura, H., Kubota, T., and Morita, K. (2003). Landslide damage and disaster management system in Nepal. *Disaster Prevention and Management: An International Journal*, 12(5), 413-419.

Phillips, C., Hales, T., Smith, H., and Basher, L. (2021). Shallow landslides and vegetation at the catchment scale: A perspective. *Ecological Engineering*, 173, 106436.

Pradhan, S., Toll, D. G., Rosser, N. J., & Brain, M. J. (2022). An investigation of the combined effect of rainfall and road cut on landsliding. *Engineering Geology*, 307, 106787. <https://doi.org/10.1016/j.enggeo.2022.106787>

Qiu, H., Regmi, A. D., Cui, P., Cao, M., Lee, J., and Zhu, X. (2016). Size distribution of loess slides in relation to local slope height within different slope morphologies. *Catena*, 145, 155-163.

Rahman, M. S., Ahmed, B., & Di, L. (2017). Landslide initiation and runout susceptibility modeling in the context of hill cutting and rapid urbanization: a combined approach of weights of evidence and spatial multi-criteria. *Journal of Mountain Science*, 14(10), 1919-1937.

Ran, Q., Wang, J., Chen, X., Liu, L., Li, J., & Ye, S. (2022). The relative importance of antecedent soil moisture and precipitation in flood generation in the middle and lower Yangtze River basin. *Hydrology and Earth System Sciences*, 26(19), 4919-4931.

Razakova, M., Kuzmin, A., Fedorov, I., Yergaliev, R., and Ainakulov, Z. (2020). Methods of calculating landslide volume using remote sensing data. In *E3S Web of Conferences* (Vol. 149, p. 02009). EDP Sciences.

Rosi, A., Peternel, T., Jemec-Auflič, M., Komac, M., Segoni, S., and Casagli, N. (2016). Rainfall thresholds for rainfall-induced landslides in Slovenia. *Landslides*, 13, 1571-1577.

Rotaru, A., Oajdea, D., and Răileanu, P. (2007). Analysis of the landslide movements. *International journal of geology*, 1(3), 70-79.

Saito, H., Korup, O., Uchida, T., Hayashi, S., & Oguchi, T. (2014). Rainfall conditions, typhoon frequency, and contemporary landslide erosion in Japan. *Geology*, 42(11), 999-1002.

Sato, T., Katsuki, Y., & Shuin, Y. (2023). Evaluation of influences of forest cover change on landslides by comparing rainfall-induced landslides in Japanese artificial forests with different ages. *Scientific reports*, 13(1), 14258.

Scheidl, C., Heiser, M., Kamper, S., Thaler, T., Klebinder, K., Nagl, F., ... and Seidl, R. (2020). The influence of climate change and canopy disturbances on landslide susceptibility in headwater catchments. *Science of the total environment*, 742, 140588.

Smith, H. G., Neverman, A. J., Betts, H., & Spiekermann, R. (2023). The influence of spatial patterns in rainfall on shallow landslides. *Geomorphology*, 437, 108795.

Spiker, E. C., & Gori, P. (2003). National landslide hazards mitigation strategy, a framework for loss reduction (No. 1244). US Geological Survey.

Stoof, C. R., Vervoort, R. W., Iwema, J., Van Den Elsen, E., Ferreira, A. J. D., & Ritsema, C. J. (2012). Hydrological response of a small catchment burned by experimental fire. *Hydrology and Earth System Sciences*, 16(2), 267-285.

Sun, H. Y., Wong, L. N. Y., Shang, Y. Q., Shen, Y. J., and Lü, Q. (2010). Evaluation of drainage tunnel effectiveness in landslide control. *Landslides*, 7, 445-454.

Tacconi Stefanelli, C., Casagli, N., & Catani, F. (2020). Landslide damming hazard susceptibility maps: a new GIS-based procedure for risk management. *Landslides*, 17, 1635-1648.

Tsai, T. L., & Chen, H. F. (2010). Effects of degree of saturation on shallow landslides triggered by rainfall. *Environmental Earth Sciences*, 59, 1285-1295.

Turner, T. R., Duke, S. D., Fransen, B. R., Reiter, M. L., Kroll, A. J., Ward, J. W., ... & Bilby, R. E. (2010). Landslide densities associated with rainfall, stand age, and topography on forested landscapes, southwestern Washington, USA. *Forest Ecology and Management*, 259(12), 2233-2247.

Um, M. J., Yun, H., Cho, W., & Heo, J. H. (2010). Analysis of orographic precipitation on Jeju-Island using regional frequency analysis and regression. *Water resources management*, 24, 1461-1487.

Van Tien, P., Luong, L. H., Duc, D. M., Trinh, P. T., Quynh, D. T., Lan, N. C., ... & Loi, D. H. (2021). Rainfall-induced catastrophic landslide in Quang Tri Province: the deadliest single landslide event in Vietnam in 2020.

Wang, D., Hollaus, M., Schmaltz, E., Wieser, M., Reifeltshammer, D., & Pfeifer, N. (2016). Tree stem shapes derived from TLS data as an indicator for shallow landslides. *Procedia Earth and Planetary Science*, 16, 185-194.

Wei, Z. L., Shang, Y. Q., Sun, H. Y., Xu, H. D., and Wang, D. F. (2019). The effectiveness of a drainage tunnel in increasing the rainfall threshold of a deep-seated landslide. *Landslides*, 16, 1731-1744.

Wieczorek, G. (1987). In central Santa Cruz Mountains, California. Debris flows/avalanches: process, recognition, and mitigation, 7, 93.

Willmott, C. J., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate research*, 30(1), 79-82.

Winter, M. G., & Bromhead, E. N. (2012). Landslide risk: some issues that determine societal acceptance. *Natural Hazards*, 62, 169-187.

Yan, L., Xu, W., Wang, H., Wang, R., Meng, Q., Yu, J., and Xie, W. C. (2019). Drainage controls on the Donglingxing landslide (China) induced by rainfall and fluctuation in reservoir water levels. *Landslides*, 16, 1583-1593.

Yang, H., & Adler, R. F. (2008). Predicting global landslide spatiotemporal distribution: integrating landslide susceptibility zoning techniques and real-time satellite rainfall estimates. *International Journal of Sediment Research*, 23(3), 249-257.

Yoon, S. S., & Bae, D. H. (2013). Optimal rainfall estimation by considering elevation in the Han River Basin, South Korea. *Journal of Applied Meteorology and Climatology*, 52(4), 802-818.

Yun, H. S., Um, M. J., Cho, W. C., & Heo, J. H. (2009). Orographic precipitation analysis with regional frequency analysis and multiple linear regression. *Journal of Korea Water Resources Association*, 42(6), 465-480.

Yune, C. Y., Jun, K. J., Kim, K. S., Kim, G. H., and Lee, S. W. (2010). Analysis of slope hazard-triggering rainfall characteristics in Gangwon Province by database construction. *Journal of the Korean Geotechnical Society*, 26(10), 27-38.

Zachar, D. (2011). Soil erosion. Elsevier.

Zaruba, Q., and Mencl, V. (2014). Landslides and their control. Elsevier.

Zhang, K., Wang, S., Bao, H., & Zhao, X. (2019). Characteristics and influencing factors of rainfall-induced landslide and debris flow hazards in Shaanxi Province, China. *Natural Hazards and Earth System Sciences*, 19(1), 93-105.