A Transferable Machine Learning Approach for Identifying Rainfall-Induced Cliff-Type (Shallow) Landslides in Seismic and Non-Seismic Regions

Sushama De Silva *, Taro Uchimura, Pang-jo Chun

Posted Date: 13 April 2026

doi: 10.20944/preprints202604.0803.v1

Keywords: rainfall-induced shallow landslide; cliff-type landslide; slope stability; landslide susceptibility; landslide conditioning factors; machine learning; random forest; seismic region; hazard zonation; Japan; Sri Lanka

 doi:10.20944/preprints202604.0803.v1

Article

A Transferable Machine Learning Approach for Identifying Rainfall-Induced Cliff-Type (Shallow) Landslides in Seismic and Non-Seismic Regions

Sushama De Silva [1], Taro Uchimura [2] and Pang-jo Chun [3]

[1] Department of Civil Engineering, Saitama University, 255 Shimookubo, Sakura Ward, Saitama 338-8570, Japan

[2] Department of Civil Engineering, Saitama University, 255 Shimookubo, Sakura Ward, Saitama 338-8570, Japan

[3] Institute of Engineering Innovation, School of Engineering, The University of Tokyo, Tokyo, Japan

* Correspondence: de.silva.d.a.d.026@ms.saitama-u.ac.jp

Abstract

Precise classification of landslide types is essential for effective hazard mitigation; however, many existing landslide inventories lack type-specific information, limiting their applicability in risk management. This study presents a transferable machine learning framework to identify rainfall-induced cliff-type landslides—commonly corresponding to shallow landslides in Japan—from unclassified inventories across both seismic and non-seismic environments. Using Forest-based and Boosted Classification and Regression (FBCR) tools in ArcGIS Pro, the model was developed based on 25 landslide conditioning factors using balanced datasets of cliff-type and non-landslide samples derived from Tokushima and Wakayama Prefectures, Japan.The model achieved strong predictive performance in the training regions, with accuracy and sensitivity exceeding 0.84, an F1 score of approximately 0.84–0.85, and a Matthews correlation coefficient (MCC) ranging from 0.68 to 0.71. Transferability was evaluated by applying the trained model to the Kegalle District in Sri Lanka, where it achieved an accuracy of approximately 80% against available inventory data. Variable importance analysis revealed that rainfall consistently ranks among the most influential triggering factors for cliff-type (shallow) landslides, even in earthquake-prone regions, where seismic-related variables exhibited comparatively lower influence. Key controlling factors included rainfall, slope, elevation, proximity to infrastructure, and hydrological indices.These findings highlight that rainfall remains a dominant trigger for shallow landslides across different tectonic settings. The proposed framework provides a practical approach for complementing missing landslide type information in existing inventories, thereby improving hazard zonation and supporting risk-informed planning in diverse environmental conditions.

Keywords: rainfall-induced shallow landslide; cliff-type landslide; slope stability; landslide susceptibility; landslide conditioning factors; machine learning; random forest; seismic region; hazard zonation; Japan; Sri Lanka

1. Introduction

Rainfall-induced shallow landslides are among the most frequent and destructive natural hazards worldwide, posing severe threats to lives, infrastructure, and the environment [1,2]. Globally, more than 3,876 landslides were recorded between 1995 and 2014, resulting in 11,689 injuries and 163,658 deaths [1]. Between 1998 and 2017, landslides affected an estimated 4.8 million people, causing more than 18,000 deaths, and climate change is projected to further amplify their frequency and intensity [3]. Approximately 90% of landslides causing casualties worldwide are

 doi:10.20944/preprints202604.0803.v1

triggered by rainfall [4], making rainfall-induced slope failure a critical focus for hazard assessment and disaster risk reduction.

Shallow landslides—generally defined as slope failures confined to depths of less than 3 m [5]—are particularly rapid in onset and difficult to predict. They occur when rainfall infiltrates into the near-surface soil layers, progressively increasing pore water pressure, reducing effective shear strength, and ultimately triggering sudden failure [6,7]. The intensity, duration, and antecedent conditions of rainfall are therefore fundamental parameters governing shallow landslide initiation, and their relationship has been extensively investigated through both empirical threshold approaches and physically based models [8–10].

In Japan, steep slope failures—referred to as cliff-type landslides—are classified as one of three major sediment disaster types, alongside landslides and debris flows [11]. Cliff-type failures are typically considered shallow landslides, involving the rapid failure of near-surface soil layers on steep slopes under intense or prolonged rainfall [12]. They represent the dominant landslide type in Japanese national inventories and are particularly frequent during typhoon events and the East Asian summer monsoon season. Despite their prevalence, many national-scale inventories across Asia—where approximately 75% of fatal landslides occur—lack explicit type classification, severely limiting the design of type-appropriate countermeasures [1,13].

The spatial distribution of fatal landslides highlights the disproportionate human impact in Asia, with South and East Asian countries recording the highest concentrations of landslide events and fatalities [1]. South Asia lies largely within seismically active belts, where the collision of the Indian Plate with the Eurasian Plate and the Pacific Ring of Fire generate frequent earthquakes that can act as additional triggers or preconditioners for slope failure [3]. In Japan, the Nankai Trough earthquake belt in the Kansai region is particularly active, creating conditions where both rainfall and seismic forces contribute to shallow landslide hazard.

Landslide susceptibility assessment has evolved substantially with the adoption of machine learning approaches, which offer data-driven alternatives to physically based models that require extensive geotechnical parameterisation [14,15]. Among these methods, the Random Forest (RF) model and gradient boosting techniques have demonstrated consistently strong performance across diverse geographical settings [16]. The Forest-based and Boosted Classification and Regression (FBCR) tool in ArcGIS Pro—which combines ensemble RF methods with gradient boosting—has emerged as a powerful framework for integrating multiple conditioning factors in landslide prediction [17,18]. Applied to cliff-type shallow landslide classification, FBCR enables the identification of the most influential conditioning factors, including rainfall, terrain attributes, and anthropogenic variables, providing a quantitative basis for hazard zonation.

Empirical rainfall intensity–duration (I–D) thresholds have long been used to predict shallow landslide triggering, establishing Japan-specific thresholds considerably lower than global equivalents due to the unique susceptibility of steep Japanese slopes to even moderate rainfall events [8]. Recent advances in deep learning have further extended forecasting capability, demonstrating that the location and timing of rainfall-induced shallow landslides over large areas can be anticipated primarily from rainfall data [10]. Early warning systems integrating rainfall thresholds with real-time monitoring networks—such as the NBRO system in Sri Lanka [19]—represent the operational frontier of rainfall-induced shallow landslide risk management.

This study addresses three critical gaps in the existing literature. First, despite the well-established classification of cliff-type landslides as shallow landslides in Japan, comparative machine learning models explicitly framing cliff-type susceptibility as a rainfall-induced shallow landslide problem across seismic and non-seismic regions have not been developed. Second, the transferability of models trained in seismically active regions to non-seismic tropical contexts—such as Sri Lanka—remains insufficiently explored. Third, many existing landslide inventories in Asia lack type-specific classification, preventing systematic assessment of which conditioning factors most strongly govern cliff-type (shallow) landslide occurrence as distinct from other landslide types.

 doi:10.20944/preprints202604.0803.v1

To address these gaps, this study develops and validates a machine learning-based susceptibility model for rainfall-induced cliff-type shallow landslides across seismic (Wakayama and Mie Prefectures, Japan) and non-seismic (Kegalle District, Sri Lanka) regions, using the FBCR tool in ArcGIS Pro with 25 landslide conditioning factors. The specific objectives are:

1. To examine the influence of landslide conditioning factors (LCFs), including rainfall and other hydrological, topographic, and anthropogenic variables, on cliff-type shallow landslide occurrence across seismic and non-seismic regions.

2. To apply the FBCR machine learning framework to model the relationship between LCFs and cliff-type shallow landslide susceptibility.

3. To validate model performance in earthquake-affected regions of Japan and assess model transferability to a non-seismic tropical context.

4. To compare high-influence LCFs between seismic and non-seismic models, with particular attention to the consistent role of rainfall as a primary triggering factor.

The remainder of this paper is organized as follows: Section 2 presents the literature review; Section 3 describes the study areas and methodology; Section 4 presents the results; Section 5 discusses the findings; and Section 6 presents conclusions and recommendations.

2. Literature Review

2.1. Rainfall-Induced Shallow Landslides: Global Significance and Triggering Mechanisms

Landslides are defined as mass movements of soil and rocks along a slope, occurring when the shear strength of hillside material decreases due to an increase in shear stress [6]. Globally, more than 3,876 landslides were recorded between 1995 and 2014, resulting in 11,689 injuries and 163,658 deaths [1]. Between 1998 and 2017, landslides affected an estimated 4.8 million people, causing more than 18,000 deaths worldwide, and climate change is projected to increase landslide frequency as extreme rainfall events intensify [3,4].

Among all triggering factors, rainfall is recognized as the primary cause of shallow landslides globally [4]. Shallow landslides—generally confined to depths of less than 3 m [5]—are particularly frequent in steep, soil-mantled terrains. When rainfall infiltrates slopes, it increases soil saturation, elevates pore water pressure, and progressively reduces shear strength, ultimately leading to sudden failure [6,7]. Rainfall intensity, duration, and antecedent soil moisture conditions are therefore fundamental controlling parameters in shallow landslide initiation [8,9].

In Japan, steep slope failures—referred to as cliff-type landslides (がけ崩れ)—are classified as shallow landslides involving the rapid failure of near-surface soil layers on steep slopes, predominantly triggered by intense or prolonged rainfall. Empirical rainfall intensity–duration (I–D) thresholds for shallow landslides in Japan were established by analysing 1,174 rainfall-induced events using quantile-regression analysis, identifying the threshold relationship $I = 2.18 \times D^{-0.39}$ [8]. These results confirmed that Japanese shallow cliff-type failures are closely governed by rainfall characteristics unique to the East Asian summer monsoon.

At the slope scale, the physical failure mechanism of shallow landslides involves coupling of rainfall infiltration and slope stability. A probabilistic GIS-based approach integrating the infinite slope model with Monte Carlo simulation demonstrated that pore water pressure increase during rainfall is the primary uncertainty driver in shallow landslide susceptibility assessment [20]. The TRIGRS model using a probabilistic framework achieved improvements of 10–16% in predictive power compared to deterministic approaches in shallow landslide forecasting [21].

The role of soil type, slope gradient, and rainfall pattern in shallow landslide initiation was experimentally demonstrated through large-scale rainfall simulation experiments [7]. Their results showed that rainfall intensity dictates the stability of shallow landslides, with pore water pressure changes being the primary physical indicator of imminent failure. Steeper slopes and coarser-grained soils were found particularly prone to failure under high-intensity rainfall—conditions directly applicable to cliff-type failures in Japan.

 doi:10.20944/preprints202604.0803.v1

2.2. Shallow Landslide Mechanisms in Japan: Cliff-Type Failures

In Japan, cliff-type landslides (slope failures) constitute the dominant sediment disaster category and are triggered primarily by typhoon-driven rainfall and intense monsoon precipitation [8,12]. The mechanisms of rainfall-induced shallow landslides in granite and granodiorite areas of the Northern Abukuma Mountains, Japan, reveal that subsurface hydrological structures—particularly the permeability contrast between the soil–saprolite column and underlying bedrock—strongly govern failure timing and location [22]. In granite terrain, water accumulates in a saturated zone at the soil–bedrock interface under extreme rainfall, generating transient pore pressure increases that trigger cliff-type failures.

The failure characteristics of rainfall-induced shallow landslides in granitic terrain on Shikoku Island, Japan, find that failures are most commonly associated with perched groundwater conditions developing at the interface between decomposed saprolite and partially weathered rock during intense rainfall [12]. During Typhoon Tokage (2004), which triggered more than 300 landslides, average rainfall intensities and durations were quantified, establishing important empirical relationships between rainfall and cliff-type shallow failures.

The large-scale rainfall experimental facility at Japan's National Research Institute for Earth Science and Disaster Resilience (NIED) has advanced understanding of rainfall–shallow landslide relationships through controlled slope failure simulations [23]. Experiments revealed that groundwater level rise and surface deformation begin gradually during rainfall but accelerate rapidly near failure, underscoring the importance of groundwater monitoring for early warning of cliff-type shallow landslides.

2.3. Rainfall-Induced Shallow Landslides in Tropical Regions: Sri Lanka and Asia

Asia accounts for approximately 75% of global fatal landslides, with South and East Asia recording the highest concentrations of events and fatalities [1]. Tropical climate regions are particularly vulnerable because high annual rainfall, high temperatures, and deep weathering produce thick residual soil layers highly susceptible to rainfall-induced saturation and shallow failure [4]. In Sri Lanka, landslides are the leading cause of natural disaster fatalities, accounting for 35% of all disaster deaths over the past decade [24]. The country experiences intense rainfall from two monsoon systems, making rainfall-triggered slope failures a recurring national hazard.

A comprehensive review of rainfall-induced landslide risk assessment in tropical regions analysed over 200 articles [4]. Their review identified that tropical climate conditions—including deep weathering and the presence of organic acids in soils—promote residual soil development, increasing shallow landslide susceptibility. Human activities, particularly rapid urbanization and infrastructure development in hilly terrain, were identified as having a more significant influence on increasing landslide frequency than climate change alone.

The TRIGRS model applied to the Bulathsinhala area, Sri Lanka, demonstrated that rainfall-induced increases in pore water pressure govern shallow landslide initiation consistent with the infinite slope stability model [25]. Shallow landslides in Sri Lanka typically fail at depths of 1–3 m at the boundary between colluvium or thin residual soil overburden and underlying parent rock—consistent with the international shallow landslide classification and analogous to cliff-type failures in Japan.

2.4. Rainfall Thresholds and Early Warning Systems

Empirical rainfall intensity–duration (I–D) thresholds are among the most widely used tools for shallow landslide early warning. Japan-specific thresholds have been shown to be considerably lower than global equivalents, reflecting the high susceptibility of steep weathered slopes to moderate rainfall events [8]. Early warning systems for rainfall-induced shallow landslides in the USA have also been extensively reviewed, with threshold-based warning systems implemented along the

 doi:10.20944/preprints202604.0803.v1

Pacific Coast, highlighting that antecedent soil moisture significantly influences the rainfall amount required to trigger failure [9].

Recent studies have demonstrated that deep learning can effectively forecast the location and timing of rainfall-induced shallow landslides across large areas using rainfall data, where antecedent rainfall (Ra) and triggering rainfall (Rt) jointly control landslide occurrence probability [10]. In Sri Lanka, the national landslide early warning system utilizes regional rainfall thresholds at multiple levels (watch, alert, evacuation) through a dense network of automated rain gauges [19]. The Aranayake disaster (2016), triggered by accumulated rainfall of 446.5 mm over three days in the Kegalle District and resulting in 127 fatalities, highlighted both the dominant role of extreme rainfall and the urgent need for effective early warning systems in tropical landslide-prone regions [19,26].

Further studies have advanced rainfall-triggered landslide risk assessment through integrated modelling approaches, demonstrating that rainfall not only initiates slope failure but also reduces shear strength, thereby increasing susceptibility to subsequent post-rainfall, earthquake-induced failures [26]. This interaction is particularly significant in seismically active regions, where rainfall and seismic forces act as compound hazard triggers.

2.5. Machine Learning Approaches for Rainfall-Induced Shallow Landslide Susceptibility

Machine learning methods have been increasingly applied to rainfall-induced shallow landslide susceptibility mapping, offering data-driven alternatives to physically based models requiring extensive geotechnical parameterisation [14–16]. Comparison of Random Forest, logistic model tree, and classification and regression tree models for landslide susceptibility prediction demonstrated that Random Forest provided superior accuracy and best handled complex, non-linear interactions among conditioning factors [16]. The study confirmed rainfall, slope, and proximity to roads and streams as consistently important conditioning factors—findings directly applicable to cliff-type shallow landslide research in Japan and Sri Lanka.

Probabilistic physically based models have been shown to improve predictive performance by accounting for spatial variability in soil properties [21], while deep learning approaches can effectively forecast shallow landslide occurrence primarily from rainfall data [10]. The combination of conditioning factor analysis and machine learning classification, as employed in this study, integrates both spatial susceptibility assessment and rainfall-triggering processes into a unified framework.

In tropical contexts, quantitative risk assessment is complicated by limited long-term landslide data, complex geological conditions, and rapid land use changes [4]. Machine learning approaches trained on available inventory data therefore offer a practical path to improving shallow landslide susceptibility mapping in data-scarce environments. The FBCR tool in ArcGIS Pro, combining ensemble Random Forest with gradient boosting, is well-suited to such contexts, offering robust performance across diverse geological and climatic settings.

2.6. Research Gaps Addressed by This Study

Despite significant advances in rainfall-induced shallow landslide research, several critical gaps remain. First, most studies separately address either physically based triggering models or spatial susceptibility mapping, without integrating conditioning factor analysis specifically for cliff-type (shallow) landslide classification across seismic and non-seismic regions. Second, the role of rainfall as a consistent high-influence conditioning factor across seismic and non-seismic contexts has not been systematically compared using the same machine learning framework. Third, national-scale landslide inventories in Asia—including Japan and Sri Lanka—often lack explicit type classification, preventing targeted assessment of cliff-type shallow landslide conditioning factors [1,4,13].

This study addresses these gaps by developing a machine learning model that identifies rainfall-induced cliff-type (shallow) landslides from unclassified inventories, demonstrating that rainfall consistently ranks as a primary conditioning factor across both seismic and non-seismic regions. The comparative framework across four study regions (Wakayama Prefecture, Mie Prefecture,

 doi:10.20944/preprints202604.0803.v1

Tokushima Prefecture, and Kegalle District) provides new evidence for rainfall-driven shallow landslide susceptibility across diverse geological, climatic, and tectonic settings, contributing to improved hazard zonation and type-specific countermeasure planning for rainfall-induced shallow landslide risk reduction.

3. Materials and Methods

3.1. Study Areas

This study focuses on four study areas across two countries—Japan and Sri Lanka—representing both seismic and non-seismic geological contexts (Figure 1). The selection of study areas was guided by three criteria: (1) availability of a reliable landslide inventory with type-specific classification, (2) comparable elevation and annual average rainfall ranges between study and reference areas to ensure meaningful LCF transferability, and (3) representation of both seismically active and non-seismic environments to enable transferability assessment.

Figure 1. Study area maps: (a) Wakayama, Mie, and Tokushima Prefectures, Japan (seismic regions); (b) Kegalle District, Sri Lanka (non-seismic region). Source: Adapted from De Silva and Uchimura [29,34].

3.1.1. Japan: Wakayama, Mie, and Tokushima Prefectures (Seismic Region)

Wakayama and Mie Prefectures are located in the Kansai region of southwestern Japan, forming the Kii Peninsula (Figure 1a). Both prefectures lie within the Nankai Trough earthquake belt, where frequent seismic activity and historically documented earthquake-induced landslides highlight their tectonic vulnerability [3]. The Kii Peninsula is characterised by steep mountainous terrain, with elevations ranging from −5 m to 1,869 m (mean: 265 m), and receives among the highest annual average rainfall in Japan, averaging approximately 3,800 mm/yr—a value driven by orographic enhancement of East Asian monsoon and typhoon precipitation on the southern Pacific-facing slopes [8,29]. This combination of steep terrain, intense rainfall, and seismic activity creates a high-risk environment for cliff-type shallow landslide occurrence, making Wakayama Prefecture the primary training area for the seismic model and Mie Prefecture the seismic validation area due to its comparable geomorphological and hydrological characteristics.

Tokushima Prefecture, located on Shikoku Island to the southwest and also within the Nankai Trough seismic belt, served as the training area for the non-seismic model (Figure 1b). Tokushima shares comparable topographic and hydrological characteristics with the Kii Peninsula, with elevations ranging from −3 m to 1,936 m (mean: 230 m) and annual average rainfall of approximately 2,128 mm/yr [34]. The similarity in elevation ranges and rainfall regimes between Tokushima and the Kii Peninsula supports meaningful cross-regional comparison of LCF influence, as outlined in Table 1. Despite being within the Nankai Trough seismic belt, Tokushima was selected as the non-seismic training region because the earthquake-distance factor was simulated to test model transferability to non-seismic contexts (Section 3.5).

 doi:10.20944/preprints202604.0803.v1

Table 1. Elevation, annual rainfall, and population density characteristics of study and reference areas.

Study / Reference Area	Role	Elevation Range (m)	Mean Elevation (m)	Annual Avg. Rainfall (mm/yr)	Population Density (p/km²)
Wakayama & Mie Prefectures, Japan (Kii Peninsula)	Seismic — model training & validation	−5 to 1,869	265	~3,800	187–392
Tokushima Prefecture, Japan	Non-seismic — model training	−3 to 1,936	230	~2,128	173.5
Kegalle District, Sri Lanka	Non-seismic — transferability validation	5 to 1,938	309	~2,306	530.4

Note: Kii Peninsula refers to Wakayama and Mie Prefectures combined. Annual rainfall values represent 20-year mean (2002–2021). Data sources: De Silva and Uchimura [29,34].

3.1.2. Sri Lanka: Kegalle District (Non-Seismic Region)

Kegalle District is located in the central highlands of Sri Lanka (Figure 1c) and is one of the most landslide-prone districts in the country [24]. The district is characterised by rugged terrain with elevations ranging from 5 m to 1,938 m (mean: 309 m)—comparable to the Japanese study regions—and a high population density of 530.4 persons/km², approximately three times that of the Japanese study areas (173.5–392 p/km²). This high population density in steep terrain significantly increases landslide exposure and risk.

Although Sri Lanka does not experience earthquake-induced landslides, Kegalle receives high annual average rainfall of approximately 2,306 mm/yr from two monsoon systems—the South West monsoon (May–September) and North East monsoon (December–February)—which are the primary triggers of shallow landslide events [19]. This rainfall regime, while lower than the Kii Peninsula (3,800 mm/yr), is comparable to Tokushima Prefecture (2,128 mm/yr), supporting the transferability of rainfall-based LCF relationships across the two non-seismic study contexts. Historical records consistently rank Kegalle among the districts with the greatest number of landslide events and fatalities in Sri Lanka [24], making it an appropriate transferability test area for assessing model robustness under contrasting geological, climatic, and tectonic conditions relative to the Japanese training regions.

The comparability of elevation ranges and annual rainfall across all four study areas — as summarised in Table 1 — was a key criterion in study area selection, ensuring that the 25 LCFs, particularly rainfall and topographic factors, operate within comparable physical parameter ranges across seismic and non-seismic contexts.

3.2. Landslide Inventories

Landslide inventories were obtained from the following sources: Wakayama and Mie Prefectural divisional disaster management centres (Japan); Tokushima Prefecture disaster management records (Japan); and the National Building Research Organisation (NBRO) of Sri Lanka for Kegalle District. For Kegalle District, a recently updated inventory incorporating explicit landslide type classification—introduced with JICA support in 2018—was used to enable type-specific validation [19].

In both Japan and Sri Lanka, cliff-type (slope failure) landslides are classified as shallow slope failures involving the rapid mobilisation of near-surface soil layers on steep slopes. As shown in Figure 2, slope failures are the predominant landslide type across all study regions: comprising approximately 85% of recorded events in Wakayama Prefecture, 82% in Mie Prefecture, and 58% in Tokushima Prefecture based on prefectural disaster management inventories, and 41% in Kegalle

 doi:10.20944/preprints202604.0803.v1

District based on the updated NBRO inventory (2002–2022) [29,34]. Debris flows represent the second most frequent type in Japanese prefectures, accounting for 10–14% of recorded events, while rock falls constitute approximately 25% of events in Kegalle District—reflecting the contrasting geological and geomorphological conditions between the steep granitic hillslopes of Japan and the central highland terrain of Sri Lanka. The consistently high proportion of cliff-type slope failures across all four study areas confirms the suitability of this landslide type as the primary focus of the susceptibility models developed in this study. Table 2 summarises the inventory statistics used for model training and spatial validation.

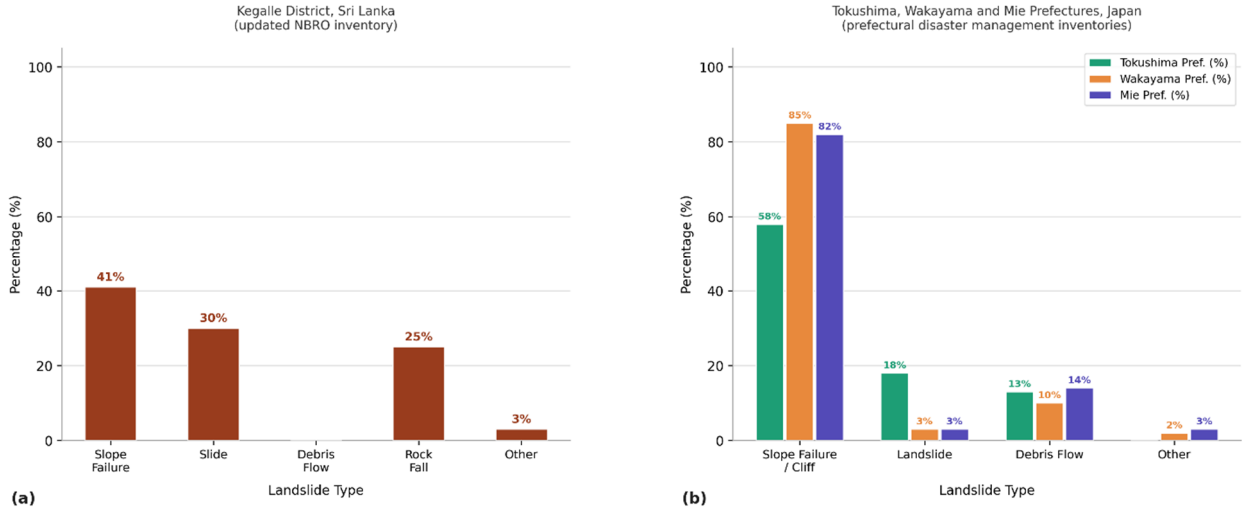


Figure 2. Comparison of landslide type distributions across study regions: (a) Kegalle District, Sri Lanka, based on the updated NBRO inventory; (b) Tokushima, Wakayama, and Mie Prefectures, Japan, based on prefectural disaster management inventories (2002–2022).

Table 2. Landslide inventory information across study and reference areas.

Area	Role	Total LS	Cliff/Slope Failure	Purpose
Wakayama (WP)	Seismic training	641	535	Model training (seismic)
Mie (MP)	Seismic validation	429	193	Spatial validation (seismic)
Tokushima (TP)	Non-seismic training	334	167	Model training (non-seismic)
Kegalle (KD)	Transferability test	221	89	Spatial validation (non-seismic)

Table 2 shows that the largest cliff-type landslide dataset was available for Wakayama Prefecture (535 points), which was therefore selected as the primary seismic training area. Mie Prefecture (193 cliff-type points) and Kegalle District (89 points) served as independent validation areas, ensuring that model performance was assessed on data entirely separate from the training dataset. The relatively smaller inventory for Tokushima Prefecture (167 cliff-type points) reflects the more limited historical records available for non-seismic regions, while still providing a sufficient dataset for reliable gradient-boosted model training with a balanced 70/30 split [5].

 doi:10.20944/preprints202604.0803.v1

3.3. Selection of Landslide Conditioning Factors (LCFs)

A total of 59 conditioning factors were initially identified through a systematic review of 31 peer-reviewed landslide susceptibility mapping studies from diverse geographical contexts. Root cohesion was excluded as its contribution can only be reliably assessed where detailed tree plantation and vegetation structure records are available. Therefore, 58 factors were retained for initial screening, subsequently reduced to 25 LCFs for model development based on literature frequency and data availability.

The 25 selected LCFs encompass five categories: (1) topographic factors, (2) hydrological indices, (3) soil and geological factors, (4) anthropogenic factors, and (5) seismic factors (Table 3). Among these, 12 factors—namely rainfall, land use, elevation (DEM), aspect, slope, profile curvature, TWI, NDVI, distance from faults, distance from roads, plane curvature, and distance from water bodies—were selected as each was referenced in more than 25% of the reviewed literature (Figure 3). Rainfall was consistently identified as a highly influential factor in 69% of reviewed studies despite relatively fewer references, reflecting its context-dependent dominance as a triggering factor for shallow landslides in monsoon-driven environments [4,8].

Table 3. The 25 landslide conditioning factors (LCFs) used in this study, by category.

Category	Landslide Conditioning Factors
Topographic	Elevation (DEM), Slope, Aspect, Profile Curvature, Plan Curvature, Topographic Position Index (TPI), Topographic Roughness Index (TRI), Direct Radiation, Direct Duration of Radiation
Hydrological	Topographic Wetness Index (TWI), Stream Power Index (SPI), Sediment Transportation Index (STI), Flow Accumulation, Flow Direction, Distance from Streams, Distance from Water Bodies
Soil & Geological	Soil Type, Soil Thickness, Geology, NDVI
Anthropogenic	Land Use, Distance from Buildings (Structures), Distance from Roads (Transportation Network)
Seismic	Distance from Earthquake Epicentres, Distance from Faults

Figure 3 illustrates that among all 58 LCFs reviewed, rainfall (Factor 2) was referenced in 65% of studies and concluded as highly influential in 58%—the highest proportion among all hydrological factors. Notably, elevation (Factor 3) and slope (Factor 5) also show high reference percentages of 68% and 87% respectively, confirming that topographic factors are universally acknowledged drivers of shallow landslide susceptibility. In contrast, factors such as sediment transportation index (STI), topographic position index (TPI), and stream power index (SPI) show consistently low influence ratings across reviewed literature, which is consistent with their low variable importance rankings observed in the models developed in this study. This analysis provides the scientific basis for the 25 LCFs selected, with particular emphasis on rainfall as the primary hydrological triggering factor for cliff-type shallow landslide occurrence across both seismic and non-seismic study regions.

 doi:10.20944/preprints202604.0803.v1

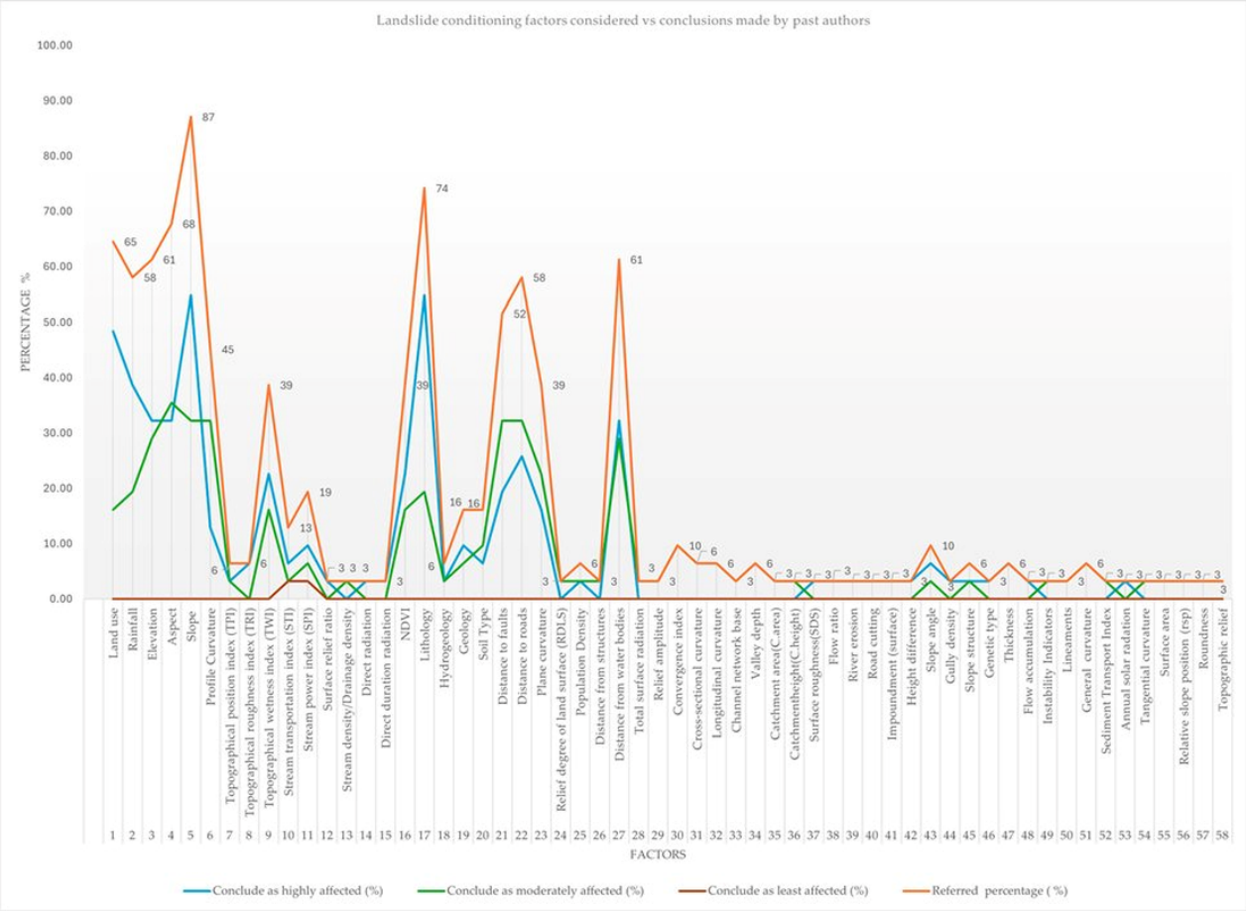


Figure 3. LCFs considered versus conclusions made by past authors, showing the percentage of studies referencing each factor and the proportion concluding high, moderate, or low influence. Source: Adapted from De Silva and Uchimura [29].

Table 4. Data types and collection sources for all study areas.

LCF	Wakayama, Mie & Tokushima Pref.	Kegalle District
Land use, Road, Structures, Water bodies	MLIT website [30]	Survey Department, Sri Lanka
Geology, Fault	MLIT website [30]	Geological Survey & Mines Bureau, Sri Lanka
Rainfall	AMeDAS website [31]	Meteorological Department, Sri Lanka
Soil type	MLIT website [30]	ISRIC World Soil Information [32]
Soil thickness	ISRIC World Soil Information [32]	ISRIC World Soil Information [32]
DEM, NDVI	USGS platform (satellite imagery)	USGS platform (satellite imagery)
Earthquake epicentres	USGS Earthquake Catalog	Simulated (see Section 3.5)

Annual rainfall data spanning 20 years were obtained from the AMeDAS network (Japan Meteorological Agency) for Japanese study areas [31] and from the Sri Lanka Meteorological Department for Kegalle District. Rainfall data were interpolated using the Inverse Distance Weighting (IDW) method and aggregated using the cell statistics tool in ArcGIS Pro to produce the mean annual average rainfall layer (Figures 4 and 5).

 doi:10.20944/preprints202604.0803.v1

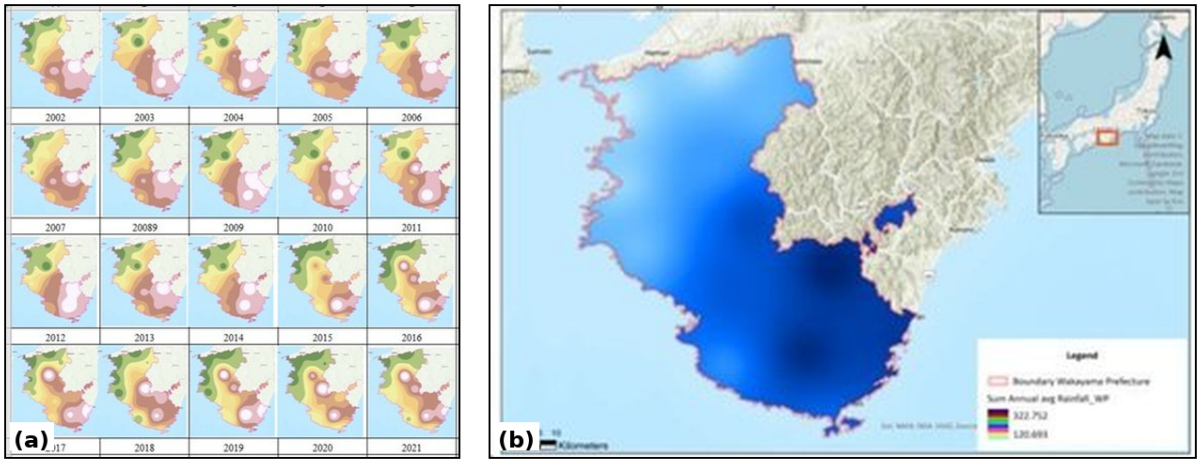
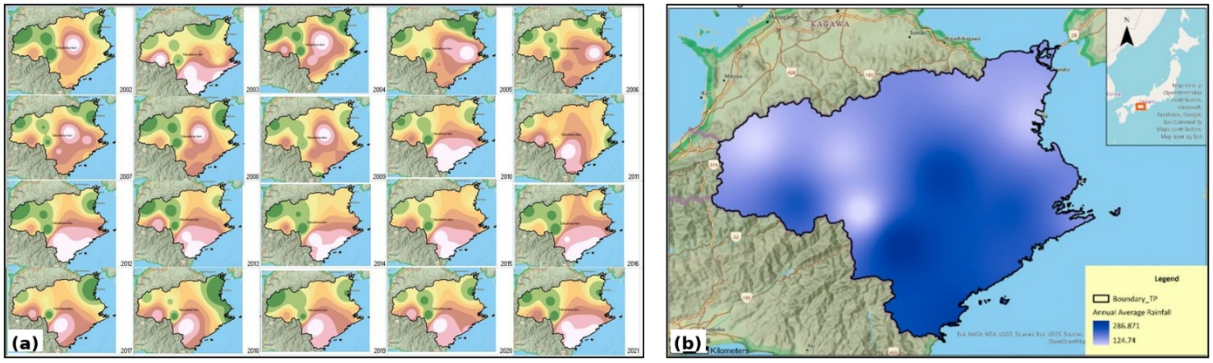


Figure 4. Annual average rainfall layers: (a) year-by-year annual average rainfall (2002–2021) and (b) 20-year mean annual average rainfall for Wakayama Prefecture. Source: Adapted from De Silva and Uchimura [29].

Figure 4 illustrates the spatial and temporal distribution of annual average rainfall across Wakayama Prefecture over the 20-year study period (2002–2021). The year-by-year maps (Figure 4a) reveal considerable inter-annual variability in rainfall distribution, with consistently higher precipitation concentrated in the steep mountainous interior of the prefecture — particularly in the Kii Peninsula, which is one of the wettest regions in Japan due to orographic enhancement of monsoon and typhoon rainfall. Notably, extreme rainfall years such as 2011 — associated with Typhoon Talas, which caused catastrophic landslide damage across the Kii Peninsula — are clearly reflected in the spatial extent of high-rainfall zones.

The 20-year mean annual average rainfall map (Figure 4b) confirms that Wakayama Prefecture receives between 120.89 and 322.75 mm of mean annual rainfall, with the highest values concentrated along the southern and inland mountainous terrain. This persistent spatial pattern of high rainfall in areas of steep topography is a fundamental predisposing condition for cliff-type shallow landslide occurrence, directly supporting the high variable importance ranking of rainfall in the seismic model developed in this study.

Figure 5 illustrates the spatial distribution of annual average rainfall across Tokushima Prefecture over the 20-year study period (2002–2021). The year-by-year maps (Figure 5a) reveal notable inter-annual variability in rainfall distribution, with consistently higher precipitation observed in the mountainous interior. The 20-year mean annual average (Figure 5b) confirms that Tokushima Prefecture receives between 124.74 and 286.87 mm of mean annual rainfall, with the highest values concentrated in steep inland areas. Together with the Wakayama rainfall data (Figure 4), these results confirm that annual average rainfall is a critical conditioning factor for cliff-type shallow landslide occurrence across both seismic study regions.



 doi:10.20944/preprints202604.0803.v1

Figure 5. Annual average rainfall layers for Tokushima Prefecture: (a) year-by-year annual average rainfall (2002–2021); (b) 20-year mean annual average rainfall. Source: Adapted from De Silva and Uchimura [34].

3.4. Machine Learning Methodology: Forest-Based and Boosted Classification and Regression (FBCR)

3.4.1. Model Selection

The Forest-based and Boosted Classification and Regression (FBCR) tool in ArcGIS Pro was employed for model development [17]. Among machine learning techniques, the Random Forest (RF) model has consistently demonstrated superior performance in landslide susceptibility prediction, with 11 out of 31 reviewed studies (35%) applying RF and achieving the highest accuracy among tested methods [15,16]. The gradient-boosted model type within the FBCR tool was specifically selected due to its sequential decision tree construction approach, in which each subsequent tree corrects the errors of the previous trees, allowing the model to combine multiple weak learners into a strong predictive model. Gradient boosting incorporates regularisation and early stopping to prevent overfitting and provides greater control over hyperparameters [17,33].

3.4.2. Training Dataset Preparation

Landslide susceptibility modelling was approached as a binary classification problem [5]. Accordingly, an equal number of non-landslide points were generated from landslide-free areas, and a 70/30 ratio was maintained for training and validation splits. Landslide points were coded as 1 (cliff-type landslide occurrence) and non-landslide points as 0. Non-landslide points were created using buffer, erase, and random point generation tools in ArcGIS Pro to ensure spatial separation from recorded landslide locations (Figure 6).

For the seismic model (Wakayama Prefecture): 535 cliff-type landslide points and 535 non-landslide points (1,070 total) [29]. For the non-seismic model (Tokushima Prefecture): 167 cliff-type landslide points and 167 non-landslide points (334 total) [34].

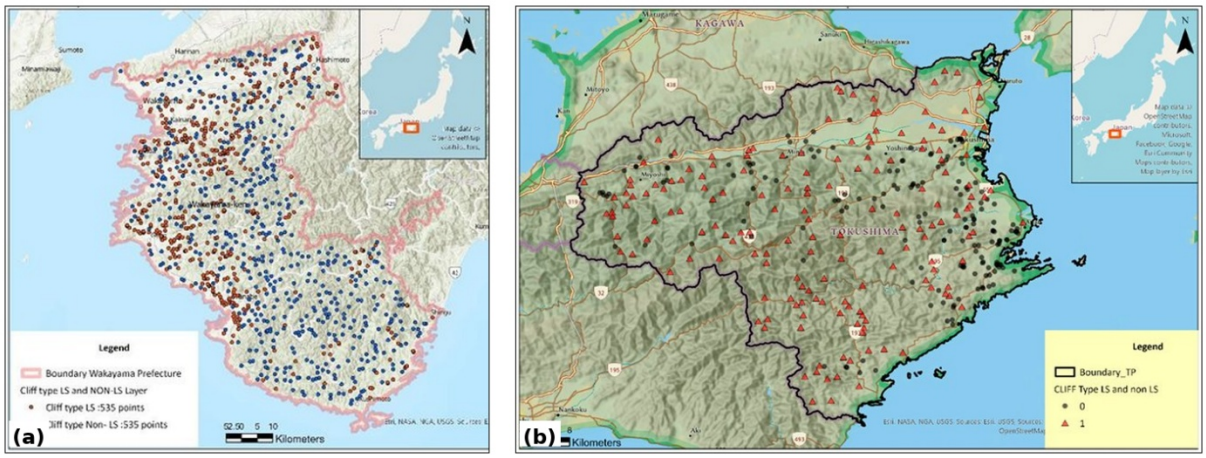


Figure 6. Cliff-type landslide occurrence points and randomly generated non-occurrence points used for model training: (a) Wakayama Prefecture (seismic model, 535 + 535 points); (b) Tokushima Prefecture (non-seismic model, 167 + 167 points). Source: Adapted from De Silva and Uchimura [29,34].

Figure 6 illustrates the spatial distribution of cliff-type landslide occurrence points (coded as 1) and randomly generated non-occurrence points (coded as 0) used for model training. In Wakayama Prefecture (Figure 6a), the 535 cliff-type landslide points show a strong spatial association with the steep inland mountainous terrain, with clusters concentrated along the southern Kii Peninsula where annual rainfall and slope gradients are highest. The 535 randomly generated non-landslide points are distributed across relatively stable lowland and coastal areas, ensuring spatial separation from

 doi:10.20944/preprints202604.0803.v1

recorded landslide zones. In Tokushima Prefecture (Figure 6b), the 167 cliff-type landslide points (red triangles) are concentrated in the mountainous interior, while the 167 non-landslide points (black circles) are distributed across flatter coastal and lowland terrain. The balanced spatial distribution of occurrence and non-occurrence points in both regions ensures that the FBCR model learns to differentiate landslide-prone from stable terrain without spatial bias [5,29,34].

All 25 LCF layers were standardised to a uniform spatial extent, raster format, pixel type, pixel depth, cell size of 33.95 m (Japan) and equivalent resolution (Sri Lanka), and the WGS 1984 Web Mercator (Auxiliary Sphere) coordinate system. Standard geoprocessing tools (copy raster, clip raster, float, define projection, and project raster) were applied during preprocessing.

3.4.3. Model Parameters

Training parameters were iteratively adjusted to optimise model performance. Table 5 presents the final optimised parameter values used for both models.

Table 5. FBCR model parameters for seismic and non-seismic models [17,33].

No.	Parameter	Seismic Model (WP)	Non-Seismic Model (TP)
1	Number of Trees	500	500
2	Data Available per Tree (%)	100	100
3	Lambda	1	0.01
4	Gamma	0	0
5	Eta	0.03	0.01
6	Maximum Number of Bins	0	154
7	Number of LCFs considered	25	24 (without seismic layer); 21*

The FBCR model parameters for seismic and non-seismic cases are summarized in Table 5. A total of 25 LCFs were initially considered for model development in the seismic regions. However, for the non-seismic model, the number of input variables was reduced to 21 based on variable importance analysis. Specifically, land use, soil type, and geology consistently exhibited very low importance scores during initial training and were therefore excluded from the final model to improve efficiency and reduce redundancy as shown in Figure 7 below.

 doi:10.20944/preprints202604.0803.v1

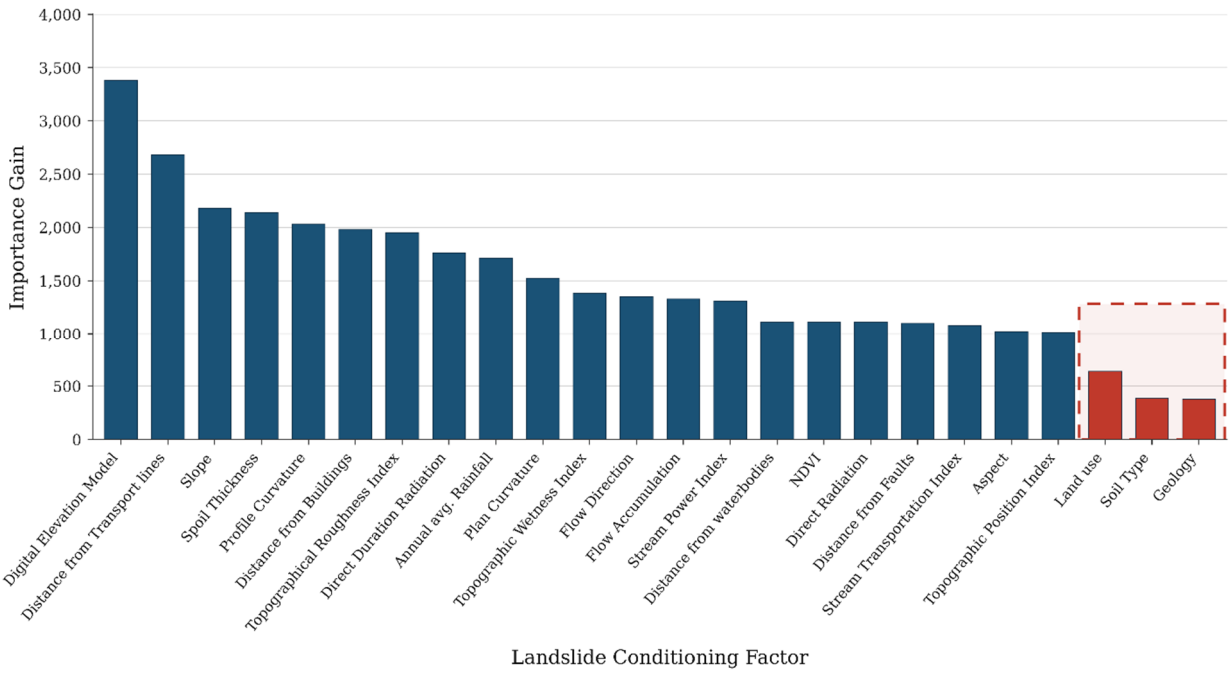


Figure 7. Distribution of variable importance gain during model training for 24 landslide conditioning factors (LCFs). Variables with very low importance (Land use, Soil type, Geology) are indicated by the dashed red boundary and were subsequently removed, reducing the number of LCFs from 24 to 21.

This reduction indicates that these categorical variables contributed minimally to the prediction of cliff-type (shallow) landslide occurrence in the non-seismic context, where topographic and rainfall-related factors were found to be more dominant. In contrast, all 25 conditioning factors were retained in the seismic model to preserve potential interactions between seismic and environmental variables. The results highlight the importance of variable selection in improving model performance and interpretability, particularly when transferring models across different geological and hazard settings. The categorical option was enabled for land use, soil type, and geology. Model outputs included the trained model, trained feature layer, variable importance table (VIT), confusion matrix (CM), and validation table, all stored in a geodatabase.

3.4.4. Model Performance Evaluation

Model performance was evaluated using the following metrics: (1) Accuracy—overall proportion of correctly classified points; (2) Sensitivity—proportion of true cliff-type landslide points correctly identified (true positive rate); (3) F1-Score—harmonic mean of precision and sensitivity; (4) Matthews Correlation Coefficient (MCC)—balanced performance metric accounting for all four confusion matrix cells, ranging from −1 (total disagreement) to +1 (perfect prediction); and (5) Validation accuracy distribution—mean, median, and standard deviation of accuracy across multiple validation runs. These metrics were chosen to provide a comprehensive and unbiased assessment of model reliability, particularly for the imbalanced prediction of cliff-type landslide occurrences in diverse landscapes.

3.5. Seismic Classification of Study Regions and Simulation of Earthquake Distance Factor

Although Japan is widely recognised as one of the world's most seismically active countries, the designation of study regions as seismic or non-seismic in this study is based not on the general occurrence of earthquakes but specifically on the recorded occurrence of earthquake-induced cliff-type landslides within each study region. This distinction is critical: the presence of seismic activity does not necessarily imply that earthquakes have directly triggered cliff-type slope failures in a given area.

 doi:10.20944/preprints202604.0803.v1

To determine the seismic classification of each candidate region, earthquake incident records from 2002 to 2022 were examined for Wakayama Prefecture (WP), Tokushima Prefecture (TP), and Mie Prefecture (MP) using USGS Earthquake Catalog data. Events occurring within a 90 km radius of the centre of each prefecture were considered. Earthquake-induced landslide likelihood was assessed by classifying seismic events according to established magnitude thresholds: magnitudes below 4.0 were classified as producing rare or nonexistent landslide occurrence; magnitudes between 4.0 and 5.5 as low to moderate likelihood; and magnitudes above 5.5 as high likelihood [35,36]. Full details of the seismic activity distribution and earthquake likelihood maps for each study region are presented in De Silva and Uchimura [29,34].

The analysis revealed that no earthquake-induced cliff-type landslides occurred in Tokushima Prefecture, despite its location within the Nankai Trough seismic belt. By contrast, 12 earthquake-induced cliff-type landslides were identified in Wakayama Prefecture and 7 in Mie Prefecture. On this empirical basis, Tokushima Prefecture was classified as the non-seismic training region, Wakayama Prefecture as the seismic training region, and Mie Prefecture as the seismic validation region. The selection of study regions was further supported by the comparable elevation ranges and annual average rainfall between all candidate regions (Table 1), ensuring meaningful cross-regional LCF transferability.

To evaluate model applicability beyond seismically active regions, the earthquake-distance factor was simulated for Kegalle District, Sri Lanka, where earthquake-induced landslides are absent. The simulation was performed by assigning a constant raster value equal to twice the maximum earthquake-epicentre distance observed in the Wakayama training dataset, ensuring the model structure was preserved while effectively neutralising the seismic factor's influence in the non-seismic context [29]. This approach allowed the seismic model to be applied to a contrasting geological and geomorphological environment without structural modification, enabling direct comparison of conditioning factor importance across seismic and non-seismic settings.

3.6. Spatial Validation Approach

Spatial validation was performed by overlaying the predicted cliff-type landslide areas (output raster) with actual cliff-type landslide inventory points in designated subareas of each validation region. Validation accuracy was calculated as the percentage of recorded cliff-type landslide inventory points correctly overlaid within the predicted cliff-type landslide area:

Spatial Validation Accuracy (%) = (Inventory points correctly overlaid / Total inventory points) × 100

This metric reflects the model's ability to spatially delineate cliff-type shallow landslide-prone areas relative to actual recorded events, providing a direct and interpretable measure of predictive reliability for hazard zonation applications.

3.7. Methodology Summary

Figure 8 presents the unified methodology framework integrating both the seismic and non-seismic models as a single comparative research design. The framework proceeds through five sequential phases: (1) shared data preparation, in which 25 LCFs were compiled and study regions were classified as seismic or non seismic based on recorded earthquake-induced cliff-type landslide occurrence [35,36]; (2) parallel model development, in which independent FBCR gradient-boosted models were trained on Wakayama Prefecture (seismic) and Tokushima Prefecture (non-seismic) datasets; (3) prediction and spatial validation across training and independent validation regions; (4) transferability assessment, in which both models were applied to Kegalle District, Sri Lanka, with the earthquake-distance factor simulated for the non-seismic context [29]; and (5) comparative LCF analysis, which identified rainfall as a consistently high-influence conditioning factor across both models — confirming its dominant role in cliff-type shallow landslide susceptibility regardless of seismic setting.

 doi:10.20944/preprints202604.0803.v1

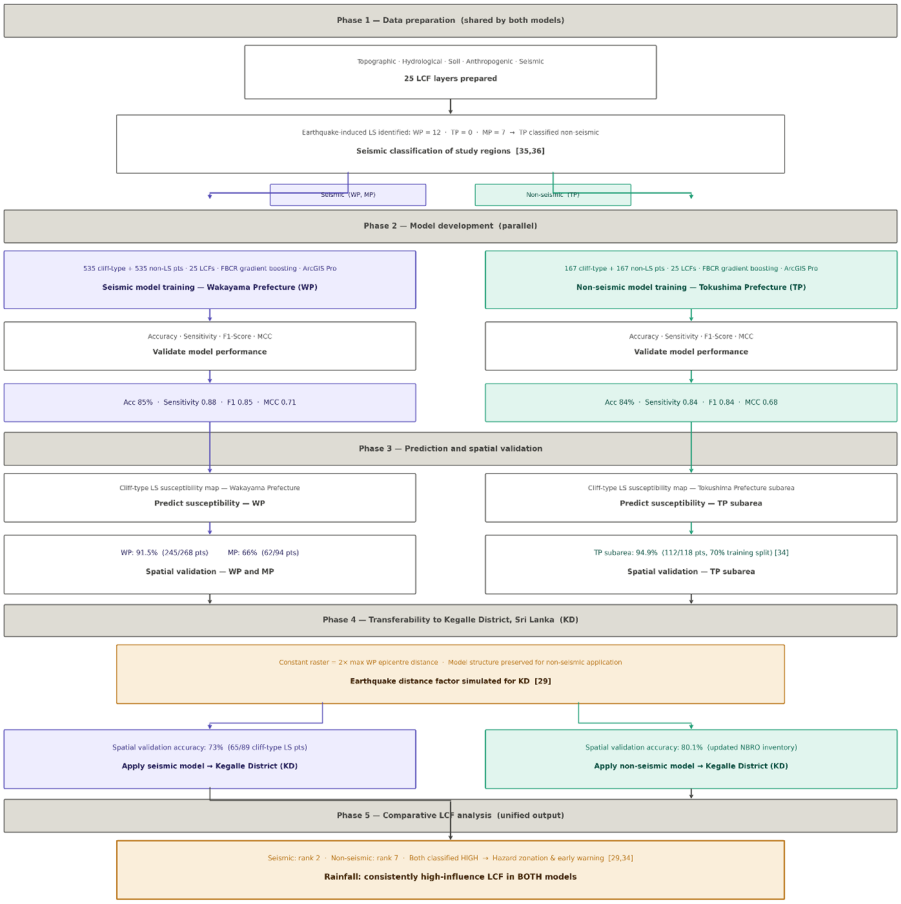


Figure 8. Unified methodology flowchart for seismic and non-seismic cliff-type shallow landslide susceptibility models.

4. Results

4.1. Model Performance — Seismic Model (Wakayama Prefecture)

4.1.1. Variable Importance Analysis

Variable importance analysis from the FBCR tool identified the top high-influence LCFs for the seismic model trained on Wakayama Prefecture data. As presented in Table 6, distance from buildings ranked first, followed by rainfall, elevation (DEM), distance from transportation networks, soil thickness, slope, and distance from earthquake epicentres — all classified as high-influence factors. These findings confirm that anthropogenic proximity, rainfall, and terrain attributes collectively govern cliff-type shallow landslide susceptibility in seismically active regions. Notably, rainfall ranked second overall, demonstrating its dominant role as a triggering LCF consistent with the East Asian monsoon climate of the Kii Peninsula [29].

Geology ranked 14th in variable importance, indicating that while geological substrate provides the underlying susceptibility context, it is not the primary driver of cliff-type shallow landslide occurrence at the spatial scale of this study. Soil type, topographic position index (TPI), and sediment transportation index (STI) were consistently classified as least significant, confirming that these factors have limited discriminatory power for cliff-type landslide classification in the Wakayama training context.

 doi:10.20944/preprints202604.0803.v1

Table 6. Variable importance ranking for the seismic model (Wakayama Prefecture training) and non-seismic model (Tokushima Prefecture training).

Rank	Seismic model — LCF (WP)	Influence	Non-seismic model — LCF (TP)	Influence
1	Distance from Buildings	High	Elevation (DEM)	High
2	Rainfall (Annual Average)	High	Distance from Roads (Transportation)	High
3	Distance from Water Bodies	High	Soil Thickness	High
4	Elevation (DEM)	High	Slope	High
5	Distance from Roads (Transportation)	High	Distance from Buildings	High
6	Distance from Earthquake Epicentres	High	Profile Curvature	High
7	Slope	Moderate	Topographic Roughness Index (TRI)	Moderate
8	Soil Thickness	Moderate	Rainfall (Annual Average)	Moderate
9	Profile Curvature	Moderate	Direct Duration Radiation (DDR)	Moderate
10	Distance from Faults	Moderate	Topographic Wetness Index (TWI)	Moderate
11	Plan Curvature	Moderate	Plan Curvature	Moderate
12	Direct Duration Radiation (DDR)	Moderate	Flow Accumulation	Moderate
13	Direct Radiation (DR)	Moderate	Direct Radiation (DR)	Moderate
14	Geology	Moderate	Flow Direction	Moderate
15	Topographic Roughness Index (TRI)	Moderate	Stream Power Index (SPI)	Moderate
16	Flow Accumulation	Moderate	Sediment Transportation Index (STI)	Moderate
17	Flow Direction	Moderate	Topographic Position Index (TPI)	Moderate
18	Topographic Wetness Index (TWI)	Moderate	Distance from Streams	Moderate
19	Land Use	Moderate	Distance from Faults	Low
20	NDVI	Moderate	Aspect	Low
21	Aspect	Moderate	NDVI	Low
22	Distance from Streams	Moderate	Distance from Water Bodies	Low
23	Soil Type	Low	Land Use	Low
24	Topographic Position Index (TPI)	Low	Soil Type	Low

 doi:10.20944/preprints202604.0803.v1

25	Sediment Transportation Index (STI)	Low	Geology	Low

4.1.2. Training Model Performance

The seismic model trained on Wakayama Prefecture data achieved strong predictive performance. As summarised in Table 7, training accuracy was 85%, with sensitivity of 0.88 for cliff-type landslide prediction, an F1-Score of 0.85, and a Matthews Correlation Coefficient (MCC) of 0.71. The high sensitivity value confirms that the model correctly identified 88% of actual cliff-type landslide occurrences, which is particularly important for hazard zonation applications where missed predictions carry significant risk implications. The MCC of 0.71 — which accounts for class imbalance across all four confusion matrix cells — indicates robust and balanced model performance across both cliff-type and non-cliff-type classifications.

Table 7. FBCR model training performance metrics for seismic and non-seismic models.

Performance Metric	**Seismic Model (WP training)**	**Non-Seismic Model (TP training)**
Training Accuracy (%)	85%	84%
F1-Score	0.85	0.84
Matthews Correlation Coefficient (MCC)	0.71	0.68
Sensitivity (True Positive Rate)	0.88	0.84

4.1.3. Spatial Validation — Seismic Regions (Wakayama and Mie Prefectures)

Spatial validation of the seismic model was conducted by overlaying the predicted cliff-type landslide susceptibility map with inventory points in designated subareas of Wakayama and Mie Prefectures. In Wakayama Prefecture (the training region), spatial validation accuracy was 91.5%, with 245 out of 268 cliff-type landslide inventory points correctly overlaid within predicted cliff-type susceptibility zones. In Mie Prefecture (the independent seismic validation region), spatial validation accuracy was 66%, with 62 out of 94 cliff-type inventory points correctly overlaid. The higher accuracy in Wakayama compared to Mie is expected, as the model was trained on Wakayama data; the 66% accuracy in Mie nonetheless demonstrates that the seismic model captures the conditioning factor relationships governing cliff-type failures across the broader Nankai Trough seismic region.

Spatial validation results are summarised in Table 8. The predicted cliff-type landslide susceptibility maps and corresponding spatial validation overlays for Wakayama, Mie, and Kegalle District (seismic model) are presented in Figure 9.

 doi:10.20944/preprints202604.0803.v1

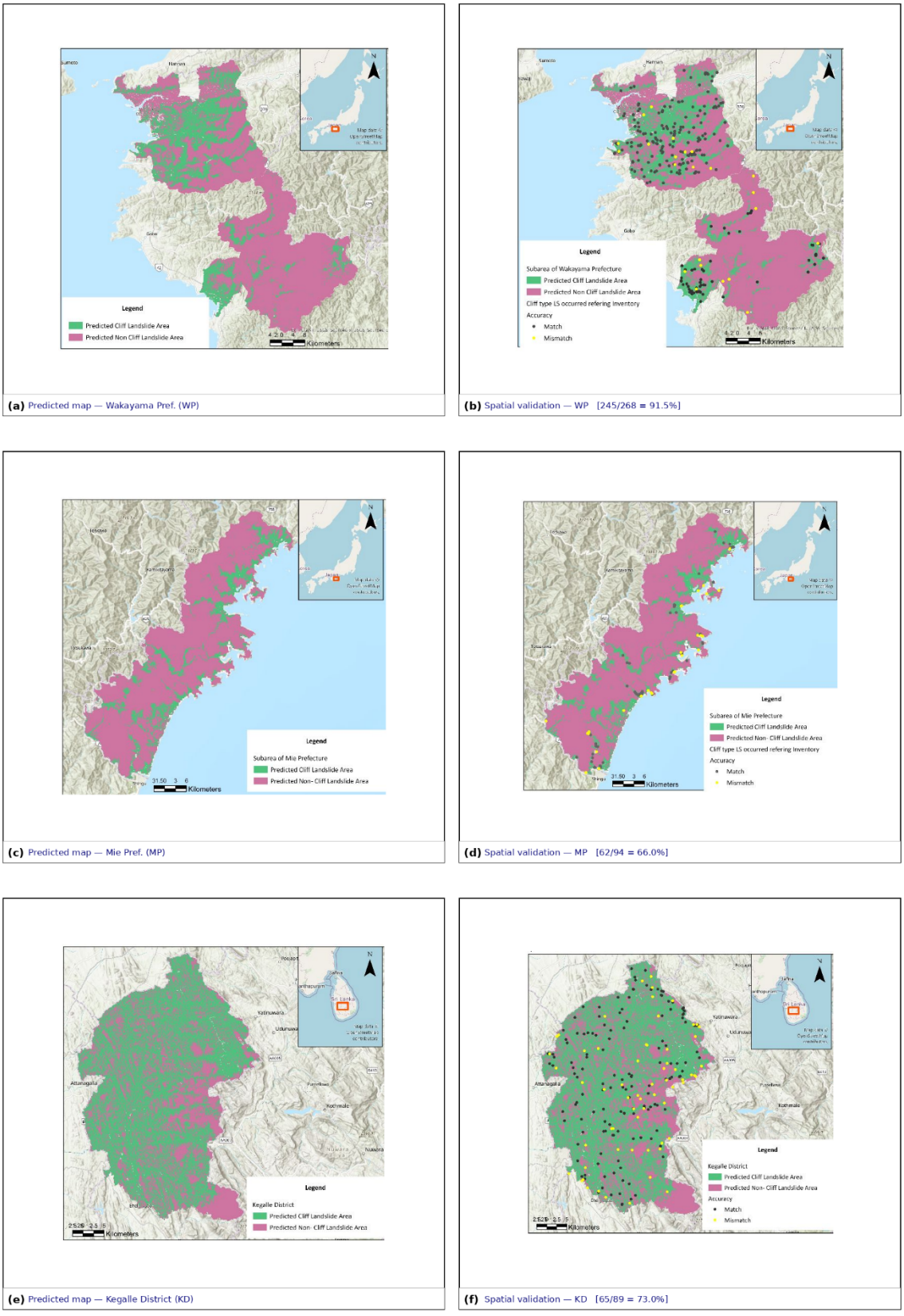


Figure 9. Cliff-type shallow landslide susceptibility maps — seismic model (Wakayama Prefecture, WP): (a) predicted susceptibility map, WP subarea; (b) spatial validation overlay, WP (245/268 = 91.5%); (c) predicted susceptibility map, Mie Prefecture (MP); (d) spatial validation overlay, MP (62/94 = 66.0%); (e) predicted susceptibility map, Kegalle District (KD); (f) spatial validation overlay, KD — seismic model (65/89 = 73.0%).

 doi:10.20944/preprints202604.0803.v1

Pink = cliff-type LS susceptibility zone; green = non-cliff LS zone; black dots = correctly predicted inventory points; yellow dots = incorrectly predicted points. Source: Adapted from De Silva and Uchimura [29].

The lower accuracy in Mie Prefecture may reflect differences in geomorphological characteristics, slope aspect distributions, and the spatial configuration of anthropogenic infrastructure compared to Wakayama. In particular, the Mie Peninsula shows more complex coastal terrain with steeper and more irregular slope profiles, which may increase the frequency of cliff-type failures in terrain zones not well represented in the Wakayama training dataset.

Table 8. Spatial validation results across all study and reference areas.

Validation area	Model applied	Total inventory points	Correctly overlaid points	Spatial validation accuracy (%)
Tokushima Pref. subarea (TP)[1]	Non-seismic model (TP trained)	118	112	94.9%
Wakayama Pref. (WP)	Seismic model (WP trained)	268	245	91.5%
Mie Pref. (MP)	Seismic model (WP trained)	94	62	66.0%
Kegalle District (KD)[2]	Seismic model (WP trained)	89	65	73.0%
Kegalle District (KD)[2]	Non-seismic model (TP trained)	89	72	80.1%

4.1.4. Model Transferability to Non-Seismic Context — Kegalle District (Seismic Model)

To assess transferability of the seismic model to a non-seismic tropical context, the trained Wakayama model was applied to Kegalle District, Sri Lanka, with the earthquake-distance factor simulated as a constant raster equal to twice the maximum epicentre distance observed in the Wakayama training dataset. Spatial validation against the available Kegalle cliff-type inventory yielded an accuracy of 73%, with 65 out of 89 inventory points correctly overlaid. This result confirms that the seismic model, trained on Japanese cliff-type shallow landslide conditioning factors, retains meaningful predictive capability in a non seismic tropical context, despite differences in geological setting, rainfall regime, and soil characteristics. The 73% validation accuracy demonstrates that the core conditioning factors — rainfall, elevation, slope, soil thickness, and distance from infrastructure — govern cliff-type shallow landslide occurrence across both seismic and non-seismic environments.

4.2. Model Performance — Non-Seismic Model (Tokushima Prefecture)

4.2.1. Variable Importance Analysis

The non-seismic model trained on Tokushima Prefecture data identified elevation (DEM) as the most influential LCF, followed by distance from transportation networks, soil thickness, slope, distance from buildings, profile curvature, and rainfall—all classified as high-influence factors (Table 6). This pattern differs from the seismic model, where distance from buildings ranked first and rainfall ranked second. The elevated importance of elevation in the non-seismic model reflects the strong correlation between terrain altitude and cliff-type shallow landslide occurrence in Tokushima's mountainous interior, where the steep granitic hillslopes are particularly susceptible to rainfall-induced shallow failures above approximately 200 m elevation.

Notably, rainfall ranked seventh in the non-seismic model, compared to second in the seismic model. While this represents a relative downranking, rainfall was still classified as a high-influence

 doi:10.20944/preprints202604.0803.v1

LCF in both models, confirming its consistent role as a primary conditioning factor for cliff-type shallow landslide susceptibility regardless of tectonic setting. The lower relative rank of rainfall in the non-seismic model may reflect the stronger topographic gradient in Tokushima, where elevation and slope together create a dominant terrain control on landslide occurrence that partially masks the within-region variability attributable to rainfall distribution.

4.2.2. Training Model Performance

The non-seismic model achieved training accuracy of 84%, with sensitivity of 84%, an F1-Score of 0.84, and an MCC of 0.68 (Table 7). These values are closely comparable to the seismic model (accuracy: 85%, sensitivity: 88%, F1: 0.85, MCC: 0.71), demonstrating consistent model performance across both seismic and non-seismic training contexts. The slightly lower MCC of 0.68 compared to 0.71 may reflect the smaller training dataset (334 points for the non-seismic model versus 1,070 for the seismic model), which reduces the diversity of conditioning factor combinations captured during training.

4.2.3. Model Transferability to Non-Seismic Context — Kegalle District (Non-Seismic Model)

The non-seismic model trained on Tokushima Prefecture was applied to Kegalle District, Sri Lanka, and validated against the recently updated NBRO cliff-type landslide inventory. Spatial validation accuracy was 80.1%, representing an improvement of approximately 7 percentage points compared to the seismic model's transferability accuracy of 73% in the same validation area. This improved accuracy may reflect the greater similarity between Tokushima's non-seismic conditioning factor environment and Kegalle's non-seismic tropical context, particularly with respect to the absence of earthquake-distance influence and the comparable roles of elevation, slope, and soil thickness in both regions. The 80.1% transferability accuracy confirms that the non-seismic model provides a robust basis for cliff-type shallow landslide susceptibility mapping in tropical non-seismic highland environments such as Kegalle District.

A notable feature of the predicted susceptibility maps for Kegalle District (KD) — observed consistently in both the seismic and non-seismic model outputs — is the accurate delineation of legally protected forest areas as non-cliff-type landslide zones (green). In particular, the Peak Wilderness Sanctuary (Sri Pada/Adam's Peak; 22,379 ha, declared 1940, UNESCO World Heritage wilderness area since 2010), located in the southern part of Kegalle District, and the Kitulgala Forest Reserve (Makandawa Forest Reserve; 1,155 ha), situated in the right-central portion of the district along the Kelani River, are both consistently classified as non-cliff LS areas in the predicted outputs of both models. This spatial correspondence is geophysically meaningful: these protected forest zones are characterised by dense intact tropical rainforest canopy, minimal anthropogenic disturbance, and negligible recorded cliff-type landslide history in the NBRO inventory — with zero cliff-type landslide incidents recorded within these boundaries. The model's ability to identify these ecologically stable zones as non-susceptible areas — without being explicitly trained on land-use or vegetation cover data — demonstrates that the conditioning factors incorporated in both models, particularly elevation, slope configuration, soil thickness, and distance from infrastructure, collectively capture the geomorphological stability associated with intact highland forest terrain. This finding reinforces the reliability and ecological coherence of both the seismic and non-seismic models in Kegalle District, and supports their potential application in conservation planning and sustainable land-use management in Sri Lanka's central highland region.

The predicted susceptibility maps and spatial validation overlays for Tokushima Prefecture and Kegalle District (non-seismic model) are presented in Figure 10.

 doi:10.20944/preprints202604.0803.v1

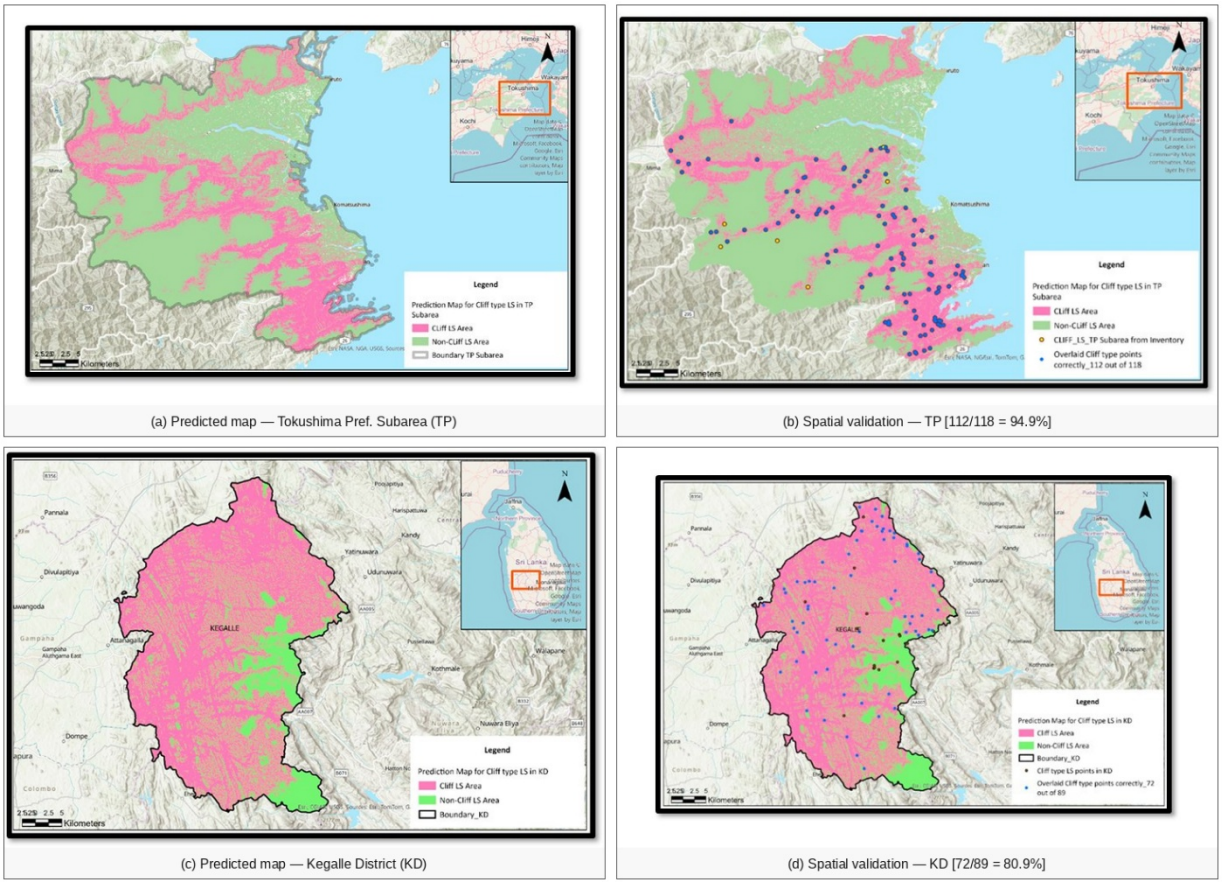


Figure 10. Cliff-type shallow landslide susceptibility maps — non-seismic model (Tokushima Prefecture, TP): (a) predicted susceptibility map, TP subarea; (b) spatial validation overlay, TP subarea (112/118 = 94.9%); (c) predicted susceptibility map, Kegalle District (KD); (d) spatial validation overlay, KD — non-seismic model (72/89 = 80.1%). Pink = cliff-type LS susceptibility zone; green = non-cliff LS zone; black dots = correctly predicted inventory points; yellow dots = incorrectly predicted points. Source: Adapted from De Silva and Uchimura [34].

4.3. Comparative Analysis: Seismic Versus Non-Seismic Models

Table 9 summarises the comparative characteristics of the seismic and non-seismic models across all performance metrics, variable importance rankings, and spatial validation results. Both models achieved comparable training performance (84–85% accuracy, MCC 0.68–0.71), confirming the robustness of the FBCR gradient-boosting framework across different tectonic and climatic contexts. The key distinction between models lies in the relative importance of the earthquake-distance factor (high influence in the seismic model, neutralised in the non-seismic model) and the primary terrain variable (distance from buildings dominant in the seismic model; elevation dominant in the non-seismic model). Critically, rainfall was classified as a high-influence LCF in both models, ranking second in the seismic model and seventh in the non-seismic model. This consistent high-influence classification across two independently trained models — applied to four geographically distinct study regions — provides strong empirical evidence that rainfall is a primary conditioning factor for cliff-type shallow landslide occurrence regardless of seismic setting, consistent with the findings of Saito et al. [8] and Amarasinghe et al. [4].

Table 9. Comparative summary of seismic and non-seismic model characteristics.

Characteristic	Seismic Model (WP)	Non-Seismic Model (TP)
Training region	Wakayama Prefecture, Japan	Tokushima Prefecture, Japan

 doi:10.20944/preprints202604.0803.v1

Training accuracy	85%	84%
Sensitivity	0.88	0.84
F1-Score	0.85	0.84
MCC	0.71	0.68
Top LCF (rank 1)	Distance from Buildings	Elevation (DEM)
Rainfall rank	2nd (High influence)	7th (High influence)
Seismic factor rank	7th (Distance from epicentres — High)	Simulated (neutralised)
Spatial val. — training region	91.5% (Wakayama)	—
Spatial val. — seismic region	66% (Mie Prefecture)	N/A
Spatial val. — non-seismic transfer	73% (Kegalle District)	80.1% (Kegalle District)

5. Discussion

5.1. Role of Rainfall as a Primary LCF Across Seismic and Non-Seismic Regions

The consistent classification of rainfall as a high-influence LCF in both the seismic and non-seismic models represents the central finding of this study. The annual average rainfall layer — derived from 20 years of AMeDAS data (2002–2021) for Japanese study areas and equivalent records for Kegalle District — captures the long-term spatial pattern of moisture availability that predisposes slopes to shallow failure. In monsoon-driven environments such as the Kii Peninsula and Tokushima Prefecture, inter-annual variability in rainfall distribution is substantial (Figures 3 and 4), and the spatial concentration of high rainfall in inland mountainous terrain strongly corresponds to the areas of highest cliff-type landslide density. This spatial coincidence confirms that rainfall operates not only as a short-term triggering mechanism — as quantified by empirical intensity–duration (I–D) thresholds [8] — but also as a long-term conditioning factor that determines the chronic susceptibility of specific slope zones to shallow failure.

The finding that rainfall ranked second in the seismic model (behind distance from buildings) and seventh in the non-seismic model (behind elevation, transport distance, soil thickness, slope, building distance, and profile curvature) does not diminish its significance. In both cases, rainfall was classified as high-influence — the highest category in the FBCR variable importance framework — alongside topographic and anthropogenic factors. This is consistent with experimental findings demonstrating that rainfall intensity is the primary physical determinant of shallow landslide failure timing [7], as well as statistical evidence indicating its dominant role in landslide occurrence [10], who showed that rainfall data alone can forecast the location and timing of shallow landslides over large areas. The present study extends these findings to a spatial conditioning factor context, demonstrating that the mean annual rainfall distribution is a robust predictor of cliff-type shallow landslide susceptibility at the regional scale.

5.2. Seismic Model: Anthropogenic and Terrain Factors

The dominance of distance from buildings in the seismic model (Wakayama Prefecture) is consistent with the high population density and distributed settlement pattern of the Kii Peninsula, where residential structures are frequently located at the foot of steep slopes — zones of maximum cliff-type landslide hazard. This anthropogenic proximity effect has been documented in previous susceptibility studies in Japan [29] and reflects the historical pattern of rural settlement in valley bottoms and lower hillslopes where terrain gradients transition sharply. The high variable importance of rainfall (rank 2) in this seismic context suggests that, once proximity to human

 doi:10.20944/preprints202604.0803.v1

infrastructure is accounted for, the spatial distribution of rainfall is the dominant natural determinant of cliff-type failure zones.

The 7th-ranked importance of earthquake epicentre distance in the seismic model confirms that seismic activity provides a secondary conditioning influence on cliff-type failure susceptibility in the Nankai Trough region. While the model does not simulate dynamic seismic triggering, the spatial distribution of historical epicentre locations serves as a proxy for cumulative seismic stress and slope fatigue — factors known to reduce the rainfall threshold required to trigger failure in previously damaged slopes [26]. The successful transferability of the seismic model to non-seismic Kegalle District (73% accuracy), achieved by neutralising the earthquake-distance factor, confirms that the core rainfall, terrain, and anthropogenic factor relationships are sufficiently general to operate across tectonic contexts.

5.3. Non-Seismic Model: Terrain Dominance and Transferability

The dominance of elevation in the non-seismic model (Tokushima Prefecture) reflects the strong vertical zonation of cliff-type shallow landslide occurrence in Tokushima's granitic mountainous interior. Failure at the saprolite–bedrock interface in granitic terrain is strongly elevation-dependent, with perched groundwater conditions most likely to develop at intermediate elevations where the soil–rock transition is sharpest and rainfall infiltration paths are concentrated [12]. The second-ranked importance of distance from transportation networks in the non-seismic model is consistent with the finding that road construction in steep terrain disrupts natural drainage patterns, creates undercutting of unstable slopes, and introduces concentrated runoff that increases pore water pressure along slope-parallel road cuts — all factors that predispose adjacent slopes to cliff-type failure.

The superior transferability of the non-seismic model to Kegalle District (80.1%) compared to the seismic model (73%) supports the hypothesis that non-seismic conditioning factor relationships are more directly portable to non-seismic tropical contexts. Kegalle's central highland terrain, thick weathered saprolite profiles, and monsoon-driven rainfall regime share fundamental physical similarities with Tokushima's granitic mountainous setting, facilitating the transfer of elevation-, slope-, and soil-thickness-based susceptibility patterns. The 80.1% validation accuracy is notably higher than reported in several comparable machine learning transferability studies in tropical Asia [3,4], confirming the robustness of the FBCR gradient-boosting framework for cross-regional cliff-type shallow landslide susceptibility mapping.

5.4. Implications for Hazard Zonation and Countermeasure Planning

The rainfall-dominated susceptibility patterns identified in this study have direct implications for hazard zonation and type-specific countermeasure planning. The consistent high influence of rainfall across both models confirms that cliff-type shallow landslide susceptibility is strongly governed by climatic and hydrological conditioning factors regardless of tectonic setting. This finding supports the use of FBCR-derived susceptibility maps to identify and prioritise high-risk slope zones for structural and non-structural countermeasures, including slope reinforcement, drainage improvement, and land-use restrictions in cliff-type landslide-prone areas.

The spatial validation accuracies reported in this study (66–91.5% in seismic regions; 73–80.1% in non-seismic regions) are comparable with or superior to those reported in recent machine learning susceptibility studies in similar contexts [15,16,20]. The variation in accuracy across validation regions reflects the inherent challenge of transferring susceptibility models to areas with partially different geomorphological characteristics — particularly the lower accuracy in Mie Prefecture (66%) compared to Wakayama (91.5%), which underscores the importance of region-specific inventory data for model training. Future studies should explore the integration of transfer learning approaches or multi-region ensemble models to improve cross-regional generalisability.

 doi:10.20944/preprints202604.0803.v1

5.5. Limitations

Several limitations of this study should be acknowledged. First, the use of mean annual average rainfall as the primary rainfall LCF captures long-term spatial patterns of moisture availability but does not account for short-duration extreme rainfall events, which are the proximate triggers of individual cliff-type failures. Incorporating event-based rainfall intensity data or antecedent soil moisture indices in future model iterations would strengthen the mechanistic link between rainfall conditioning and triggering. Second, the seismic model's earthquake-distance factor is based on historical epicentre locations and does not simulate real-time seismic hazard, limiting its utility for immediate post-earthquake landslide hazard assessment. Third, the relatively small training dataset for the non-seismic model (334 points) constrains the diversity of conditioning factor combinations captured during training, potentially limiting performance in terrain zones underrepresented in the Tokushima inventory. Future work should prioritise the expansion of cliff type landslide inventories in non-seismic regions to support more robust model development.

6. Conclusions

This study developed and validated a machine learning-based susceptibility framework for rainfall-induced cliff-type shallow landslides (がけ崩れ) across seismic (Wakayama and Mie Prefectures, Japan) and non seismic (Kegalle District, Sri Lanka) regions using the Forest-based and Boosted Classification and Regression (FBCR) tool in ArcGIS Pro with 25 landslide conditioning factors (LCFs). The principal conclusions are as follows:

1. Rainfall was consistently classified as a high-influence LCF in both the seismic and non-seismic models, ranking second in the seismic model and seventh in the non-seismic model. This consistent finding across four geographically distinct study regions — spanning two countries, two tectonic settings, and two monsoon regimes — provides strong empirical evidence that the mean annual average rainfall distribution is a primary conditioning factor for cliff-type shallow landslide susceptibility, independent of seismic context.

2. Both models achieved comparable training performance (accuracy: 84–85%, sensitivity: 84–88%, F1: 0.84–0.85, MCC: 0.68–0.71), demonstrating the robustness of the FBCR gradient-boosting framework for cliff-type landslide classification across seismic and non-seismic training contexts.

3. Spatial validation confirmed model transferability across seismic and non-seismic regions: the seismic model achieved 91.5% accuracy in Wakayama Prefecture, 66% in Mie Prefecture, and 73% in non-seismic Kegalle District; the non-seismic model achieved 80.1% transferability accuracy in Kegalle District. These results confirm that the core rainfall, terrain, and anthropogenic LCF relationships are sufficiently generalisable to support cross-regional susceptibility mapping.

4. Region-dependent susceptibility patterns were identified: distance from buildings and rainfall dominated in the seismic model, reflecting the anthropogenic proximity effects and monsoon rainfall gradient of the Kii Peninsula; elevation and distance from transportation networks dominated in the non-seismic model, reflecting the vertical zonation of cliff-type failures in Tokushima's granitic mountainous terrain.

5. The findings support the use of FBCR-derived susceptibility maps to inform hazard zonation and type-specific countermeasure planning for rainfall-triggered shallow slope failures in both seismic and non-seismic contexts. Future research should address the current limitations by incorporating event-based rainfall intensity data, expanding cliff-type landslide inventories in non-seismic tropical regions, and exploring transfer learning approaches to improve cross-regional model generalisation.

Author Contributions: Conceptualization, S.D.S. and T.U.; methodology, S.D.S.; software, S.D.S.; validation, S.D.S. and T.U.; formal analysis, S.D.S.; investigation, S.D.S.; resources, T.U. and P.-j.C.; data curation, S.D.S.; writing — original draft preparation, S.D.S.; writing — review and editing, T.U. and P.-j.C.; visualization, S.D.S.; supervision, T.U.; project administration, T.U. All authors have read and agreed to the published version of the manuscript.

 doi:10.20944/preprints202604.0803.v1

Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Data Availability Statement: The landslide inventory data for Wakayama and Mie Prefectures were obtained from Prefectural disaster management centres (Japan). Tokushima Prefecture inventory data were obtained from Tokushima Prefecture disaster management records. Kegalle District inventory data were obtained from the National Building Research Organisation (NBRO), Sri Lanka. Rainfall data for Japan were obtained from the AMeDAS network (Japan Meteorological Agency). Soil thickness data were obtained from ISRIC World Soil Information. Additional data sources are listed in Table 4.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Gómez, D.; García, E.F.; Aristizábal, E. Spatial and temporal landslide distributions using global and open landslide databases. Nat. Hazards 2023, 117, 25–55. https://doi.org/10.1007/s11069-023-05848-8
2. Ali, S.A.; Parvin, F.; Vojteková, J.; Costache, R.; Linh, N.T.T.; Pham, Q.B.; Vojtek, M.; Gigović, L.; Ahmad, A.; Ghorbani, M.A. GIS-based landslide susceptibility modeling: A comparison between fuzzy multi-criteria and machine learning algorithms. Geosci. Front. 2021, 12, 857–876.
3. Ibrahim, M.; Al-Bander, B. An integrated approach for understanding global earthquake patterns and enhancing seismic risk assessment. Int. J. Inf. Technol. 2024, 16, 2001–2014.
4. Amarasinghe, M.P.; Kulathilaka, S.A.S.; Robert, D.J.; Zhou, A.; Jayathissa, H.A.G. Risk assessment and management of rainfall-induced landslides in tropical regions: a review. Nat. Hazards 2024, 120, 2179–2231.
5. Bennett, G.L.; Miller, S.R.; Roering, J.J.; Schmidt, D.A. Landslides, threshold slopes, and the survival of relict terrain in the wake of the Mendocino Triple Junction. Geology 2016, 44, 363–366.
6. Moresi, F.V.; Maesano, M.; Collalti, A.; Sidle, R.C.; Matteucci, G.; Mugnozza, G.S. Mapping landslide prediction through a GIS-based model: A case study in a catchment in southern Italy. Geosciences 2020, 10, 309.
7. Liu, Y.; Deng, Z.; Wang, X. The effects of rainfall, soil type and slope on the processes and mechanisms of rainfall-induced shallow landslides. Appl. Sci. 2021, 11, 11652.
8. Saito, H.; Nakayama, D.; Matsuyama, H. Relationship between the initiation of a shallow landslide and rainfall intensity–duration thresholds in Japan. Geomorphology 2010, 118, 167–175.
9. Baum, R.L.; Godt, J.W. Early warning of rainfall-induced shallow landslides and debris flows in the USA. Landslides 2010, 7, 259–272.
10. Mondini, A.C.; Guzzetti, F.; Melillo, M. Deep learning forecast of rainfall-induced shallow landslides. Nat. Commun. 2023, 14, 2466.
11. Junichi, K.; Naoki, I. Outline of measures for sediment disaster by the Sabo department of MLIT, Japan. Landslides 2020, 17, 2503–2513.
12. Dahal, R.K.; Hasegawa, S.; Nonomura, A.; Yamanaka, M.; Masuda, T.; Nishino, K. Failure characteristics of rainfall-induced shallow landslides in granitic terrains of Shikoku Island of Japan. Environ. Geol. 2009, 56, 1295–1310.
13. Modugno, S.; Johnson, S.C.M.; Borrelli, P.; Alam, E.; Bezak, N.; Balzter, H. Analysis of human exposure to landslides with a GIS multiscale approach. Nat. Hazards 2022, 112, 387–412.
14. He, Q.; Shahabi, H.; Shirzadi, A.; et al. Landslide spatial modelling using novel bivariate statistical based Naïve Bayes, RBF Classifier, and RBF Network machine learning algorithms. Sci. Total Environ. 2019, 663, 1–15.
15. Kumar, C.; Walton, G.; Santi, P.; Luza, C. An ensemble approach of feature selection and machine learning models for regional landslide susceptibility mapping in the arid mountainous terrain of Southern Peru. Remote Sens. 2023, 15, 1376.
16. Chen, W.; Xie, X.; Wang, J.; Pradhan, B.; Hong, H.; Bui, D.T.; Duan, Z.; Ma, J. A comparative study of logistic model tree, random forest, and classification and regression tree models for spatial prediction of landslide susceptibility. Catena 2017, 151, 147–160.

 doi:10.20944/preprints202604.0803.v1

17. Esri. Forest-based and boosted classification and regression (Spatial Statistics). ArcGIS Pro Documentation. Available online: https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-statistics/forestbasedclassificationregression.htm (accessed on 10 July 2025).
18. Zhang, W.; He, Y.; Wang, L.; Liu, S.; Meng, X. Landslide susceptibility mapping using random forest and extreme gradient boosting: A case study of Fengjie, Chongqing. Geol. J. 2023, 58, 2372–2387.
19. Konagai, K.; Karunawardena, A.; Bandara, K.N.; et al. Early warning system against rainfall-induced landslide in Sri Lanka. In Progress in Landslide Research and Technology; Springer: Berlin/Heidelberg, Germany, 2023; Volume 1, pp. 217–228.
20. Park, H.J.; Lee, J.H.; Woo, I. Assessment of rainfall-induced shallow landslide susceptibility using a GIS-based probabilistic approach. Eng. Geol. 2013, 161, 1–15.
21. Raia, S.; Alvioli, M.; Rossi, M.; Baum, R.L.; Godt, J.W.; Guzzetti, F. Improving predictive power of physically based rainfall-induced shallow landslide models: a probabilistic approach. Geosci. Model Dev. 2014, 7, 495–514.
22. Kondo, A.; Matsushi, Y. Mechanisms of rainfall-induced shallow landslides regulated by hydrological subsurface structures: Cases in granite and granodiorite areas in Northern Abukuma Mountains, Japan. Eng. Geol. 2026, 366, 108692.
23. Sakai, N.; Ishizawa, T.; Danjo, T. Experimental research on rain-induced landslide mechanism using large-scale rainfall experimental facility. In Progress in Landslide Research and Technology; Springer: Berlin/Heidelberg, Germany, 2024; Volume 3, pp. 173–185.
24. National Building Research Organisation (NBRO). Annual Report 2022; NBRO: Colombo, Sri Lanka, 2022.
25. Jayasekara, E.I.; Weerasekara, N.K.; Jayathissa, H.A.G.; Gunatilake, A.A.J.K. Evaluation of shallow landslide-triggering scenarios through a physically based approach: A case study from Bulathsinhala area, Sri Lanka. In Proceedings of the 7th International Conference on Debris-Flow Hazards Mitigation, Golden, CO, USA, 10–13 June 2019.
26. Sassa, K.; Doan, L.; Matsunami, K.; et al. Risk identification of large-scale landslides triggered by rainfalls and post-rainfall earthquakes in Sri Lanka. In Progress in Landslide Research and Technology; Springer: Berlin/Heidelberg, Germany, 2025; Volume 3, pp. 207–229.
27. Chowdhury, M.S. A review on landslide susceptibility mapping research in Bangladesh. Heliyon 2023, 9, e17972.
28. Batar, A.K.; Watanabe, T. Landslide susceptibility mapping and assessment using geospatial platforms and weights of evidence (WoE) method in the Indian Himalayan region. ISPRS Int. J. Geo-Inf. 2021, 10, 155.
29. De Silva, S.; Uchimura, T. A model for complementing landslide types (cliff type) missing from areal disaster inventories based on landslide conditioning factors in earthquake prone areas. Intell. Inform. Infrastruct. 2025, JSCEIIAI-2025-0020.R1. (In press)
30. Ministry of Land, Infrastructure, Transport and Tourism (MLIT). Download site for digital land information. Available online: https://nlftp.mlit.go.jp/ksj/index.html (accessed on 10 July 2025).
31. Japan Meteorological Agency (JMA). AMeDAS. Available online: https://www.jma.go.jp/jma/en/Activities/amedas/amedas.html (accessed on 10 July 2025).
32. ISRIC—World Soil Information. SoilGrids FAQ. Available online: https://www.isric.org/explore/soilgrids/faq-soilgrids (accessed on 10 July 2025).
33. Esri Academy. Available online: https://www.esri.com/training/ (accessed on 10 July 2025).
34. De Silva, S.; Uchimura, T. A model for complementing landslide types (cliff type) missing from areal disaster inventories based on landslide conditioning factors for earthquake-proof regions. ResearchGate 2025. Available online: https://www.researchgate.net/publication/394922308 (accessed on 6 April 2026).
35. Keefer, D.K. Landslides caused by earthquakes. Bull. Geol. Soc. Am. 1984, 95, 406–421.
36. Papadopoulos, G.A.; Plessa, A. Magnitude–distance relations for earthquake-induced landslides in Greece. Eng. Geol. 2000, 58, 377–386.

 doi:10.20944/preprints202604.0803.v1

disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.