

Article

Not peer-reviewed version

A Model for Complementing Landslide Types (Cliff Type) Missing from Areal Disaster Inventories Based on Landslide Conditioning Factors for Earthquake-Proof Regions

[Sushama Malshani De silva](#) * and [Uchimura Taro](#)

Posted Date: 12 August 2025

doi: 10.20944/preprints202507.1169.v2

Keywords: cliff type; landslide (LS); inventory; landslide conditioning factors (LCF); forest-based and boosted classification and regression tool (FBCR); model; tokushima prefecture (TP); kegalle district (KD)



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Article

A Model for Complementing Landslide Types (Cliff Type) Missing from Areal Disaster Inventories Based on Landslide Conditioning Factors for Earthquake-Proof Regions

Sushama De Silva * and Uchimura Taro

Dept. of Civil Eng., Saitama University, 255 Shimookubo, Sakura Ward, 338-8570

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

Abstract

Precise classification of landslide types is critical for targeted hazard mitigation, although the absence of type-specific classifications in many existing inventories limits their utility for effective risk management. This study develops a transferable machine learning approach to identify cliff-type landslides from unclassified records, with a focus on earthquake-prone regions. Using the Forest-Based and Boosted Classification and Regression (FBCR) tools in ArcGIS Pro, a model was trained on 167 landslide points and 167 non-landslide points from Tokushima Prefecture, Japan. The model achieved high predictive performance, with 84% accuracy and sensitivity, an F1 score of 84%, and a Matthews correlation coefficient (MCC) of 0.68. The trained model was applied to the Kegalle District, Sri Lanka, and validated against a recently updated inventory specifying landslide types, resulting in an accuracy of 80.1%. It also enabled retrospective identification of cliff-type landslides in older inventories, providing valuable insights for early hazard assessment. Spatial analysis showed strong correspondence between predicted cliff-type zones and key conditioning factors, including specific elevation ranges, steep slopes, high soil thickness, and proximity to roads and buildings. This study integrates FBCR-based modelling with a cross-regional application framework for cliff-type landslide classification, offering a practical, transferable tool for refining inventories, guiding countermeasures, and improving preparedness in regions with similar geomorphological and seismic settings.

Keywords: cliff type; landslide (LS); inventory; landslide conditioning factors (LCF); forest-based and boosted classification and regression tool (FBCR); model; tokushima prefecture (TP); kegalle district (KD)

1. Introduction

Landslides are defined as the downward movement of soil, rock, and organic materials under the influence of gravity. They typically occur when the shear strength of slope materials is exceeded by shear stress, often due to external triggers such as intense rainfall, seismic activity, or human disturbances [1]. Globally, landslides pose a significant threat to both human life and infrastructure. Between 1995 and 2014, over 3,876 documented landslide events led to approximately 11,689 injuries and 163,658 fatalities. In 2014 alone, at least 174 landslides occurred worldwide, resulting in severe human and environmental losses [2].

Regional statistics also highlight the devastating impacts of landslides. In the United Kingdom, a prolonged period of above-average rainfall from April to December 2012—one of the wettest periods in the country's meteorological history—triggered a marked increase in landslide events, as recorded in the National Landslide Database (NLD) maintained by the British Geological Survey (BGS) [3]. Similarly, across Europe from 1995 to 2014, 476 landslides were reported, causing 1,370

deaths and 784 injuries. The most affected countries during this period included Turkey (335 deaths), Italy (283), Russia (169), and Portugal (91) [4].

Landslide countermeasures are actions taken to prevent slope instability or to reduce its impacts. These measures can be broadly classified into two categories: structural and non-structural. To implement the most appropriate countermeasure, it is crucial to first identify the specific type and mechanism of the landslide [5].

Landslide classification varies from country to country. In Japan, the term "landslide" encompasses three main phenomena: slope failure (also known as cliff failure), landslides, and debris flows. Collectively, these are referred to as sediment disasters [6]. Although rockfall is not classified as a type of landslide, it is included in inventories under a separate category.

In Sri Lanka, landslides are classified into four types: slope failure, slides, debris flows, and rockfalls. The definitions of these classifications are quite similar in both Japan and Sri Lanka. Although landslides have been categorized in Sri Lanka, a more detailed classification system was introduced to the inventory in 2018 with the support of JICA . This updated classification is referred to as the recently modified inventory in this paper. Past disaster records are essential information for any kind of countermeasures. National Building Research Organization in Sri lanka has been collecting and managing a number of landslide records in the past. Since those have been stored in paper basis, without disaster type and in different formats, it is difficult to utilize for risk assessment and designing countermeasures [5].

Figure 1 shows the distribution of landslide and sediment disaster types in Kegalle District, Sri Lanka, and Tokushima Prefecture, Japan, from 2002 to 2022. Slope failures are the predominant type in both regions, comprising 41% and 58% of recorded events, respectively. Slide/landslide events are more common in Kegalle (30%) than in Tokushima (18%), whereas debris flows occur only in Tokushima (13%). Rock falls account for 26% of events in Kegalle but are absent from the Tokushima records. Other types represent less than 5% in both datasets. For Kegalle District, a recently updated inventory was used to create Figure 1.

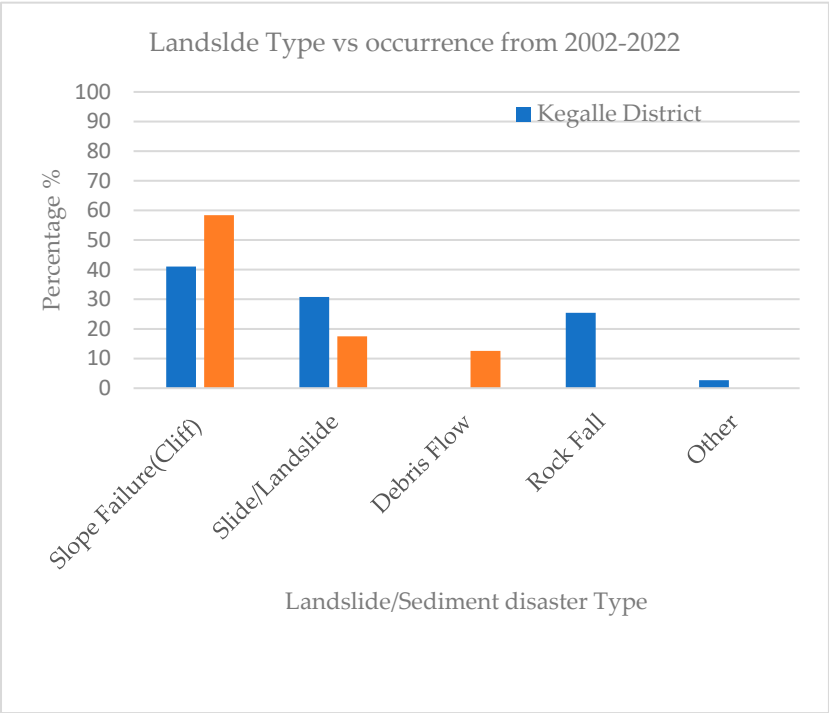


Figure 1. Landslide Type vs occurrence from 2002-2022 in Kegalle District and Tokushima Prefecture.

Numerous studies have explored the factors contributing to landslide occurrence, commonly referred to as landslide conditioning factors. These include topographic elements (e.g., slope,

elevation, aspect), geological conditions (e.g., soil type, lithology), land use, and proximity to natural and built features such as water bodies, streams, and buildings. Various modeling approaches have been applied to evaluate the relative importance of these factors. A summary of recent research is presented in Table 1, highlighting the numbers of conditioning factors and incidents considered, along with the associated conclusions.

Table 1. Current research status of landslide triggering factors.

No of LS considered		1460	30	1624	0.37M	198	1748	212	141	91	2000
LCF considered		24	12	16	6	10	9	12	9	21	16
1	Land use Land cover										
2	Climate (Rainfall)										
3	Elevation /Altitude										
4	Aspect/Exposure										
5	Slope										
6	Profile Curvature										
7	TPI										
8	TRI										
9	TWI										
10	STI										
11	SPI										
12	DR										
13	DDR										
14	NDVI/ vegetation										
15	Geology										
16	Soil Type										
17	Distance to faults										
18	Distance to roads										
19	Distance to epicenter										
20	Plane curvature										
21	Distance from structures										
22	Distance from streams										
23	Thickness										
24	Flow accumulation										
Reference		[7]	[8]	[9]	[10]	[11]	[12]	[13]	[14]	[15]	[16]
Legend											
		Highly affected		Least affected							
		Moderately affected		Not considered							

The matrix presented summarizes the consideration of various Landslide Conditioning Factors (LCFs) across multiple studies, categorized by the number of landslides (LS) examined in each case. The horizontal axis represents individual studies or datasets, with the respective number of LS considered indicated at the top of each column. The vertical axis lists the 24 LCFs that are commonly applied in landslide susceptibility modelling, including topographic, geological, hydrological, and anthropogenic parameters.

The analysis reveals that Land Use/Land Cover, Climate (Rainfall), Elevation/Altitude, Slope, and NDVI/Vegetation are the most consistently used LCFs, often classified as highly influential across datasets. In contrast, variables such as Thickness, Plane Curvature, Distance from Structures, and Flow Accumulation are less frequently considered. Some parameters, such as Soil Type and Distance to Faults, are included selectively and often marked as least influential, reflecting their varying importance depending on the local geological context.

The variation in LCF usage across studies reflects both the differences in data availability and the geomorphological context of the study areas. For example, hydrological factors such as TWI and SPI are predominantly considered in studies focusing on rainfall-induced landslides, whereas Distance to Epicenter is only relevant in seismically triggered events.

Overall, the comparative assessment highlights a trend toward prioritizing topographic and vegetation-related factors as primary determinants of landslide susceptibility, while certain localized parameters are incorporated on a case-by-case basis. Further information regarding LCF will be discussed under the first objective, which is to investigate the factors that affect landslide occurrence in the Materials and Methods section.

In Sri Lanka, landslides pose a serious threat, particularly within the central highlands, which encompass 12 districts recognized for their high susceptibility to slope instability [17]. While some landslides occur due to natural factors such as intense rainfall or geological conditions, anthropogenic activities—such as deforestation, unregulated construction, and poor land-use practices—have significantly amplified the risk. Alarming, nearly 30% of the national population resides within these mountainous regions, increasing their exposure to such hazards. Major landslide events recorded in 2003, 2007, 2010, 2011, 2012, 2014, 2015, and 2016 collectively resulted in nearly 1,000 fatalities. Furthermore, approximately 20,000 km²—equivalent to 30.7% of Sri Lanka's total land area—is classified as highly landslide-prone [18].

Given this alarming trend, the identification and mapping of landslide-prone zones are essential for disaster risk reduction and sustainable land-use planning. Landslide assessments are typically conducted through a tiered approach that includes susceptibility mapping (identifying areas likely to experience landslides), hazard mapping (considering frequency and magnitude), and risk mapping (accounting for both hazard and vulnerability) [19].

Significance of the Study:

Historical disaster records are essential for developing effective and targeted countermeasures against landslides. In Sri Lanka, the National Building Research Organisation (NBRO) has been systematically collecting landslide-related data for many years. However, much of this information has traditionally been stored in paper-based formats or across disparate systems, making it difficult to access and analyze for risk assessment and mitigation planning. While NBRO currently maintains a disaster inventory that includes basic attributes such as the date of occurrence, location, scale, and rainfall data, there remains a critical need to organize and categorize this information in a structured digital format. To address this gap, NBRO has initiated the development of an Excel-based database that compiles key parameters, including disaster location, occurrence date, rainfall conditions, landslide type, event scale, and the resulting damage [20]. Such a system would greatly enhance the ability to design context-specific structural and non-structural interventions.

Although landslide susceptibility models, such as logistic regression, have been widely developed, type of landslide is not considered in countries including Bangladesh, India, Indonesia, and Nepal, the significance and behavior of contributing variables are known to vary considerably between regions [21]. Moreover, many existing models are applied at national or regional scales and often overlook the fact that landslides are influenced by distinct sets of conditioning factors [22]. Failing to account for these differences may lead to generalized or ineffective mitigation measures.

Therefore, this study aims to enhance landslide risk management by predicting the type of landslide reported in the National Inventory, enabling the implementation of more appropriate and effective countermeasures. Given the severe consequences of landslides on human life, infrastructure, and the environment, improving predictive capability is a critical step toward reducing future impacts and enhancing community resilience.

Aim and Objectives:

Aim of this project is to develop a model for earthquake-unaffected regions to identify cliff-type landslides from a landslide inventory where the type is not specified. This will reference an area with a suitable inventory and a similar range of elevation and annual average rainfall.

Objectives are,

1. To investigate the factors that influence the occurrence of landslides.
2. To identify the most suitable techniques and tools for determining the relationship between Landslide Causative Factors (LCF) and landslide types.

3. To select an area that has an appropriate inventory and a comparable range of elevation and annual average rainfall.

4. To develop and train a model that finds the relationship between LCF and landslide type, including triggering LCF, and to validate the model.

5. To predict and validate cliff-type landslides in the inventory of the study area.

Study area:

Sri Lanka is an island in South Asia, and the Kegalle District, indicated in Figure 2 below, serves as the study area. The map delineates the administrative extent of Kegalle District (highlighted in light yellow with an orange boundary) within the Sabaragamuwa Province of Sri Lanka. The district boundary is clearly demarcated to differentiate it from surrounding administrative units.

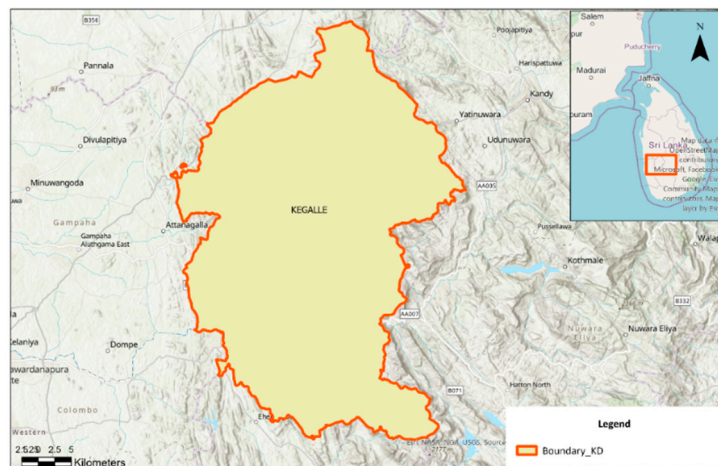


Figure 2. Administrative boundary of Kegalle District (KD) - Study Area (SA).

Figure 3 below illustrates the Digital Elevation Model (DEM) of the Kegalle District, Sri Lanka, highlighting the spatial variation in elevation across the region. The DEM values range from 10 m to 1,934 m above mean sea level, with lower elevations (depicted in green) concentrated in the western and central portions of the district, and higher elevations (depicted in red) predominantly located along the eastern boundary adjacent to the Central Province highlands. The district boundary is demarcated in orange for spatial reference. The inset map in the upper left corner shows the location of Kegalle District within Sri Lanka, providing geographical context for the study area. This elevation distribution reflects the district's varied topography, which plays a critical role in influencing geomorphological processes and potential landslide susceptibility.

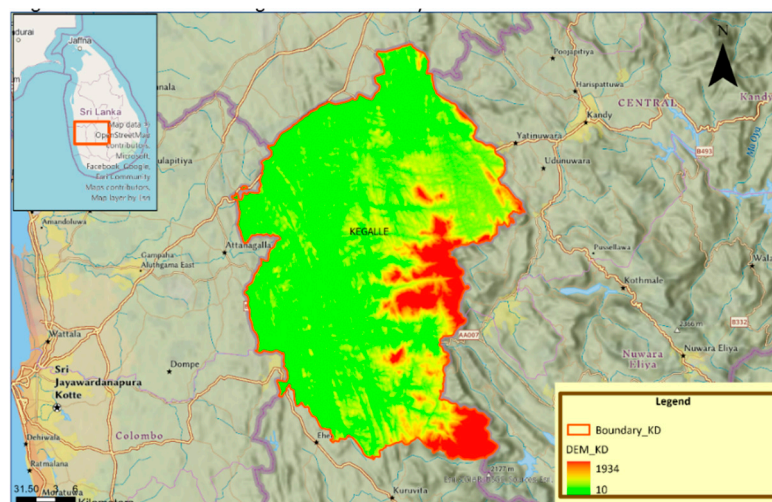


Figure 3. Elevation range in Kegalle District (KD).

The Figure 4 presents the spatial distribution of annual average rainfall within the defined study boundary, estimated using Inverse Distance Weighting. (IDW) interpolation. Rainfall intensity is depicted using a blue-scale gradient, where lighter tones (166.988 mm) indicate lower rainfall and darker tones (451.768 mm) represent higher rainfall concentrations.

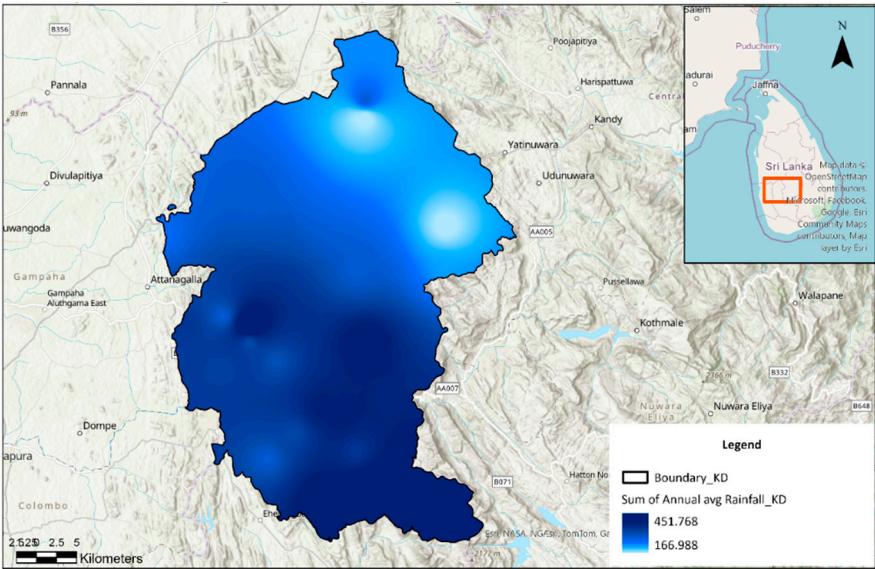


Figure 4. Annual average rainfall in Kegalle District (KD).

The Figure 5 illustrates the structural geology and lithological composition of the study area within Kegalle District. Geological contacts and boundaries are shown using standardized symbols, including inferred and approximate geological boundaries, axial traces of folds, faults, shear zones, and thrust lines. Lithological units are color-coded, encompassing rock types such as biotite-hornblende gneiss, calc-gneiss, granite gneiss, quartzite, marble, charnockite, and garnet granulite. Structural features such as antiformal and synformal folds, fractures, and overturned structures are indicated with distinct symbology. The inset map (upper right) locates the study area within Sri Lanka, providing broader geographic context. The base map includes topographic relief to aid interpretation of structural patterns in relation to terrain.

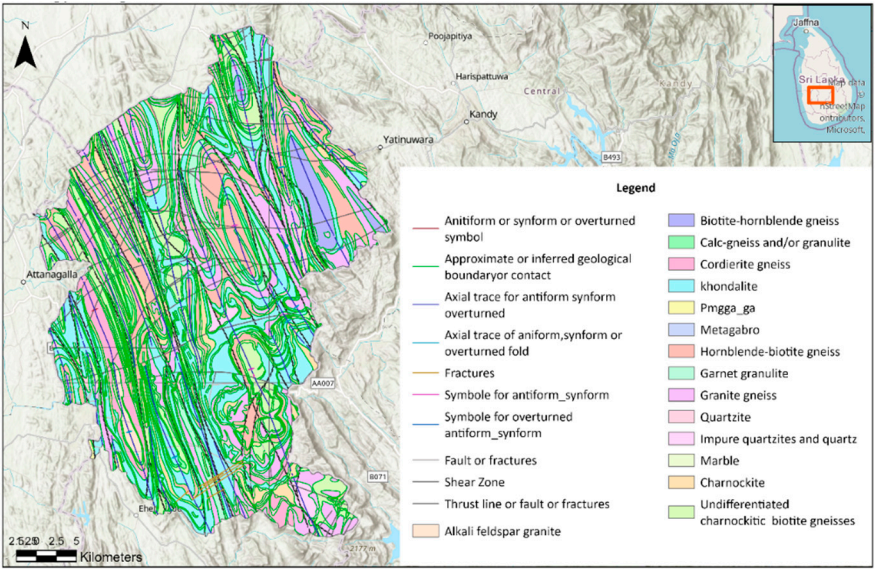


Figure 5. Geology of Kegalle District (KD).

The Figure 6 depicts the spatial distribution of land use categories within the Kegalle District boundary. Land use types are color-coded to represent barren land, coconut plantations, dense and open forests, forest plantations, homesteads/gardens, water bodies, other cultivation areas, paddy fields, rubber plantations, rock outcrops, scrub lands, sparsely used croplands, tea plantations, and other miscellaneous uses. The predominant land use types in the Kegalle District are homesteads/gardens (39.8%), paddy fields (25.6%), and rubber plantations (18.6%), reflecting a mixed agro-residential landscape.

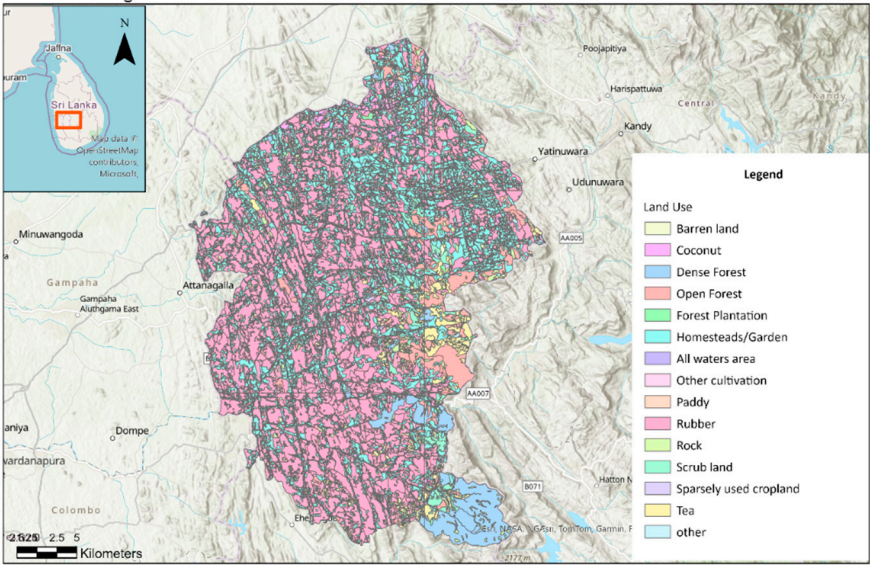


Figure 6. Land use in Kegalle District (KD).

In Sri Lanka, out of 25 districts, 10—specifically Badulla, Nuwara-Eliya, Kegalle, Kandy, Ratnapura, Matale, Kalutara, Matara, Galle, and Hambantota—are highly susceptible to landslides. According to statistics from 1974 to 2020, Kegalle District ranks fifth in the number of landslide incidents and second in fatalities associated with landslides [23].

2. Materials and Methods

2.1. Investigation of Factors Affecting Landslide Occurrence

A total of 52 conditioning factors were identified through a literature review. As shown in Figures 7–9, thirty-one research papers from different countries that considered LCF for susceptibility mapping were used to identify 52 LCF for this study. Root cohesion was excluded from this study, as it requires specific knowledge of the tree plantation area. Therefore, only 51 factors were included for initial screening. Further screening occurred during model training, which will be discussed in the results and discussion section.

The pie chart illustrates the proportion of landslide susceptibility mapping referred for each year from 2017 to 2023. The largest contribution occurred in 2023 (32.3%), followed by 2021 (19.4%), 2020 (16.1%), and 2019 (9.7%).

The pie chart (Figure 8) displays the referred global distribution of landslide susceptibility mapping efforts by country. China accounts for the largest share (32.3%), followed by Sri Lanka (12.9%) and Africa (9.7%). Moderate contributions are recorded for India, Iran, Austria, and Japan (each 6.5%). Smaller proportions (3.2%) are reported for America, Nepal, Pakistan, Slovakia, Turkey, and the global “World” category.



Figure 7. Referred research papers on landslide susceptibility mapping to identify LCF : Year wise.

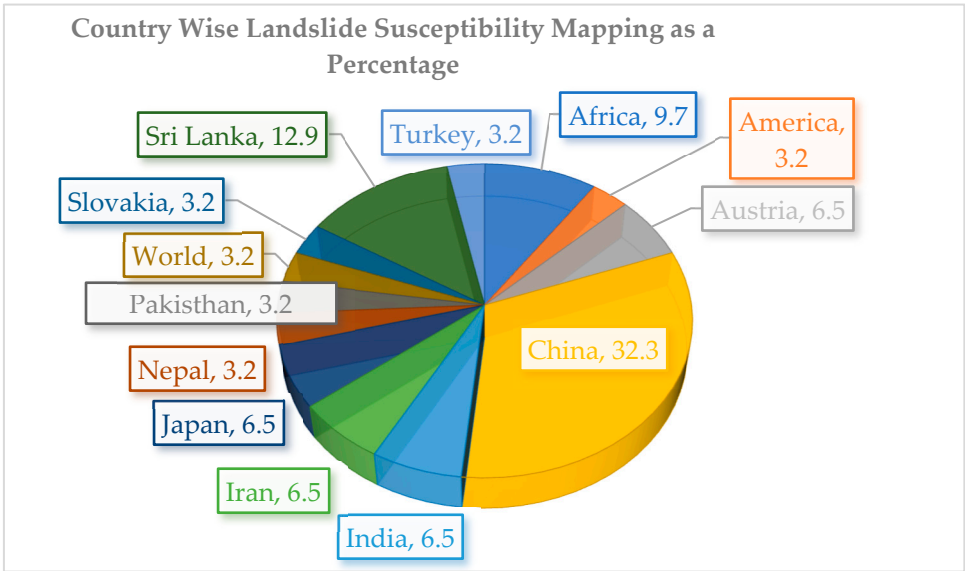


Figure 8. Referred research papers on landslide susceptibility mapping to identify LCF: country wise.

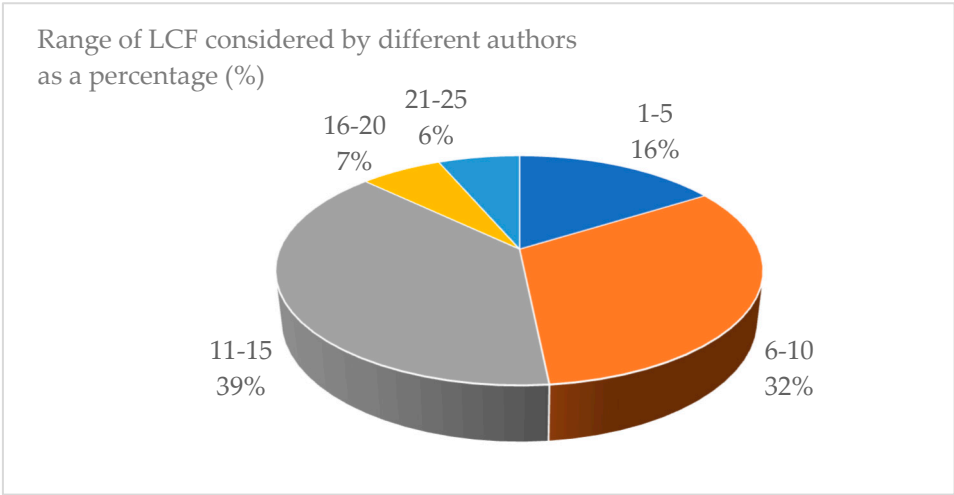


Figure 9. Range of LCF considered by different authors as a percentage (%).

The above Figure 9, pie chart illustrates the distribution of Landslide Conditioning Factors (LCFs) considered by various authors in landslide susceptibility studies. The number of LCFs used varies notably, reflecting differences in methodological approaches, data availability, and study objectives. The largest proportion of studies (39%) incorporated 11–15 LCFs, suggesting a preference for a moderately comprehensive factor set that balances analytical robustness with data manageability. The second most common range is 6–10 LCFs (32%), which may be adopted in studies with data limitations or those focusing on specific regional conditions.

A smaller share of studies (16%) used 1–5 LCFs, likely representing preliminary assessments, rapid hazard mapping, or research emphasizing a limited set of dominant conditioning variables (e.g., slope, lithology, and rainfall). Only 7% of studies applied 16–20 LCFs, and 6% considered 21–25 LCFs—these larger ranges are typically found in highly detailed, data-rich investigations aiming for maximum predictive accuracy. This distribution reflects a methodological trade-off: including more LCFs may capture complex interactions influencing landslide susceptibility but also increases data demands, processing time, and the risk of multicollinearity in statistical models. Conversely, fewer LCFs simplify analysis but may omit relevant influencing factors, potentially reducing model reliability.

Figure 10 below illustrates the Landslide Conditioning Factors (LCFs) considered, along with conclusions drawn by previous authors. Dominant Factors Frequently Considered and Judged Highly Influential: Slope (87%), land use (65%), elevation (66%), and aspect (61%) are among the most frequently cited LCFs, with a high proportion of studies concluding they are highly influential in landslide occurrence. Lithology (74%), distance to roads (58%), and distance from rivers/streams (61%) also exhibit high reference and high impact ratings, suggesting their strong geotechnical and geomorphological relevance.

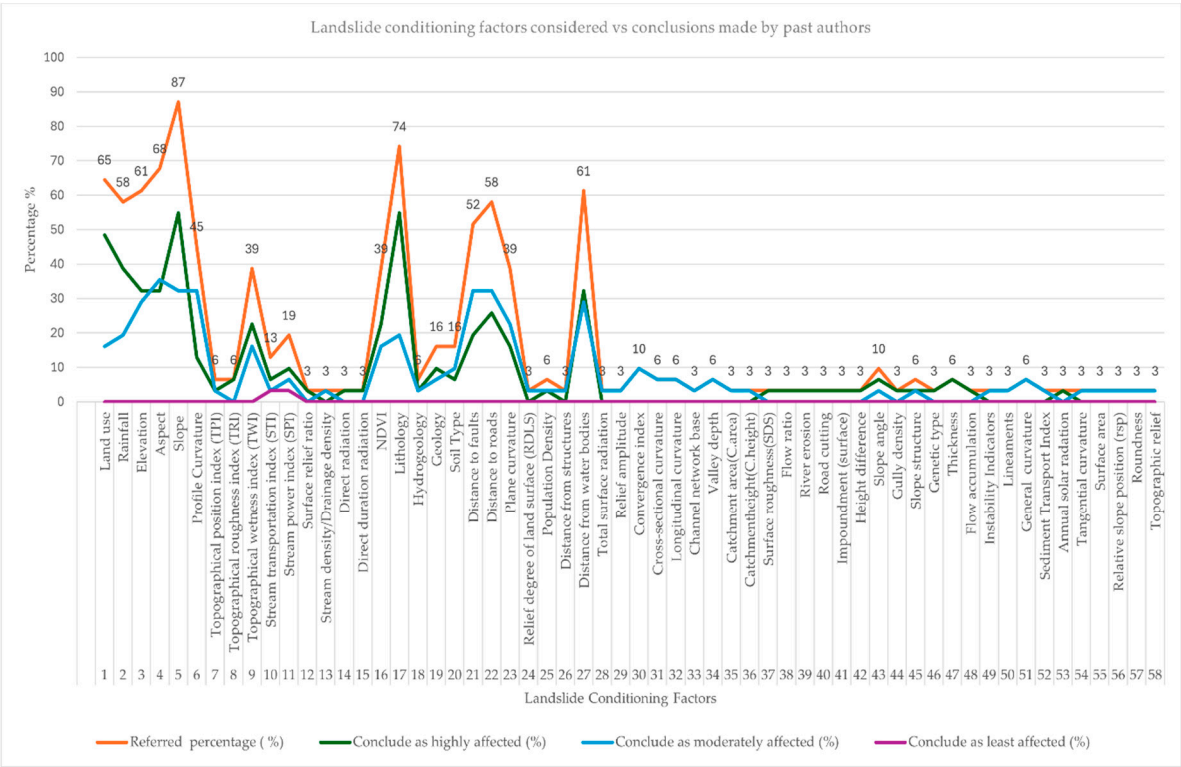


Figure 10. LCF considered vs conclusions made by past authors. [7–16,24–35].

Moderately Considered Factors: Factors such as topographical position index (39%), geology (38%), drainage density (52%), and soil type (16%) are moderately referenced but still considered significant by many authors. This indicates that while they are not universally included, they are often judged important when data is available.

Factors with Low Reference but High Impact in Specific Contexts: Rainfall (69% considered highly influential despite fewer references), stream power index (19%), and road curvature (10%) show cases where local or regional conditions may elevate their importance. These are often context-dependent, where climatic or anthropogenic pressures dominate landslide triggers.

Least Referenced and Least Impactful Factors: Several LCFs (e.g., river proximity, sediment load of river, instability indications, topographic wetness index) are rarely referenced (3–6%) and are generally concluded as having minimal influence. These may be either redundant with other factors or less relevant in most study areas.

Trend in Author Preferences: The high peaks for slope, lithology, and elevation confirm that terrain morphology and geological structure are universally acknowledged as the most critical drivers of landslide susceptibility. The clustering of low-reference factors suggests that while numerous potential LCFs exist, researchers tend to prioritize a core set of well-established variables.

2.2. Identification of the Most Suitable Techniques to Determine the Relationship Between LCFs and Landslide Types

The Forest-Based and Boosted Classification and Regression (FBCR) tool, implemented in GIS Pro software, was used to analyze the relationship between LCFs and various types of landslides. The random forest model demonstrated superior performance in landslide prediction [36]. The Random Forest Classifier (RFC) provided the best results for susceptibility assessment [2], and the developed Random Forest Machine (RFM) is a promising tool for assisting local authorities in mitigating shallow landslide hazards [37]. One-third of the reviewed literature employed the Random Forest method and reported the highest accuracy, while other studies applied different machine learning techniques.

The FBCR tool utilizes two supervised machine learning methods: an adaptation of the random forest algorithm developed by Leo Breiman and Adele Cutler, and the Extreme Gradient Boosting (XGBoost) algorithm created by Tianqi Chen and Carlos Guestrin. It enables predictions for both categorical (classification) and continuous (regression) variables. Explanatory variables can include fields in the attribute table of the training features, raster datasets, and distance features used to calculate proximity values. In addition to validating model performance based on training data, predictions can be made for either features or a prediction raster [38,39].

The gradient-boosted model was chosen for its methodological approach, which builds a model through a boosting technique where each decision tree is created sequentially using the original training data. Each subsequent tree corrects the errors of the previous trees, allowing the model to combine multiple weak learners to produce a strong predictive model. This technique incorporates regularization and early stopping, which helps prevent overfitting and provides greater control over hyperparameters, though it is more complex [38,39].

2.3. Selection of a Study Area with an Appropriate Inventory and Comparable Elevation and Annual Average Rainfall Ranges

Sri Lanka has similar topographical and geological conditions to Japan [40]. All information related to the 47 prefectures was thoroughly analyzed, leading to the selection of Wakayama Prefecture (WP) and Tokushima Prefecture (TP) as reference areas. The availability of adequate inventory data, including types of landslides, elevation ranges, and annual average rainfall ranges, was considered, as outlined in Table 2.

The elevation range was determined to be 14, based on factors such as elevation, aspect, slope, profile curvature, plane curvature, TWI, STI, SPI, TRI, TPI, direct radiation, duration of direct radiation, flow accumulation, and flow direction, derived from the Digital Elevation Model (DEM). Population density was also compared in the chosen prefectures. Factors such as land use, NDVI, distance to roads, and distance from structures correlate with population density. Environmental LCFs, including geology, soil type, soil thickness, and distance from water bodies, posed challenges in comparison.

The distance from the earthquake epicenter (LCF) was analyzed separately for Japan, an earthquake-prone country, and Sri Lanka, which is not. Landslide and earthquake inventories [41] for the Target Province (TP) and Western Province (WP) were examined to identify instances of earthquake-induced landslides. Earthquake incidents from 2002 to 2022 that occurred within a 90 km radius of the centers of the TP and WP were considered (see Figure 11). The analysis revealed that no earthquake-induced landslides occurred in the TP, whereas 12 cliff-type landslides were identified in the WP. An earthquake-induced landslide likelihood map was generated by reviewing existing literature[42,43], which indicated the following likelihood of landslide occurrence based on earthquake magnitude: for magnitudes less than 4, landslides are rare or nonexistent; for magnitudes between 4 and 5.5, the likelihood is low to moderate; and for magnitudes greater than 5.5, the likelihood is high.

Table 2. Elevation and Annual average rainfall ranges in Study Area (SA) and Reference Area (RA).

Study/Reference Area	Elevation range	Rainfall range
Kegalle District (KD)	10-1929 m	167-452 mm/yr
Wakayama Prefecture	0-1369 m	120-333 mm/yr
Tokushima Prefecture	0-1949 m	125-287 mm/yr

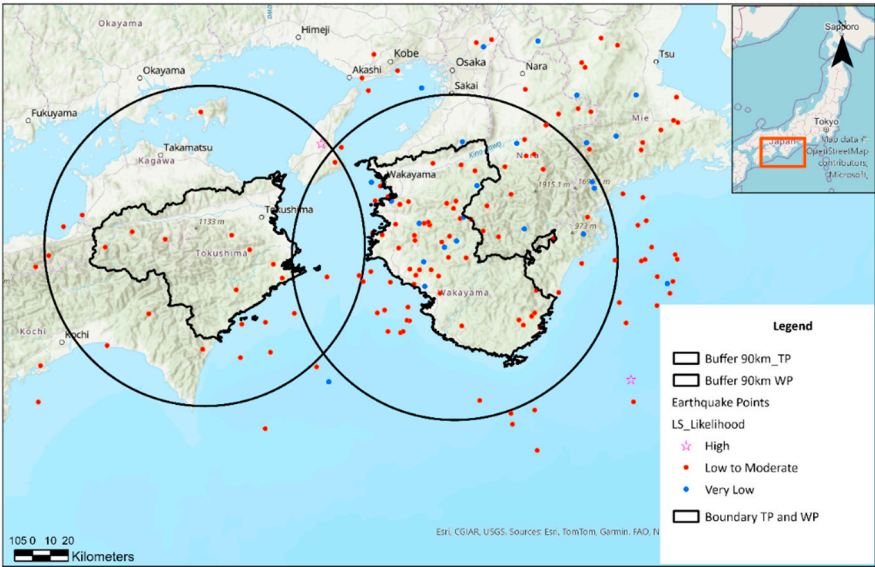


Figure 11. Earthquakes occurred in WP and TP.

2.4. Development, Training, and Validation of a Model to Analyse the Relationship Between LCFs, Landslide Types, and Triggering Factor Occurrence

A total of 24 layers were created by collecting data from relevant authorities and processing it with GIS Pro software. NDVI (Normalized Difference Vegetation Index) and DEM (Digital Elevation Model) layers were generated using satellite images downloaded from the USGS website[44]. Additional layers, including Aspect, Slope, Profile Curvature, Plane Curvature, Topographic Wetness Index (TWI), Stream Transportation Index (STI), Stream Power Index (SPI), Topographic Roughness Index (TRI), Topographic Position Index (TPI), Direct Radiation, Duration of Direct Radiation, Flow Accumulation, and Flow Direction, were developed based on the DEM layer. The soil thickness layer was created using information from the ISRIC website [45] . Additional details regarding the remaining layers can be found in Table 3 below.

Table 3. Data type and data collection sources.

Landslide Conditioning Factor	Kegalle District	Tokushima and Wakayama Pref.
Land use, Road, Structures, waterbodies	Survey Dept.	MLIT website [46]
Geology, Fault	Geological Survey and Mines Bureau	MLIT website [46]
Rainfall	Meteorological department	AMeDAS website [47]
Soil type	ISRIC World soil information website [48]	MLIT website [46]

Aspect, also referred to as exposure, indicates the compass direction a terrain surface faces. Slope represents the rise or fall of the land surface. Profile Curvature runs parallel to the maximum slope, while Plane Curvature is perpendicular to it. The Topographic Wetness Index (TWI), also known as the Compound Topographic Index (CTI), quantifies the topographic control over hydrological processes. The Stream Transportation Index (STI) describes erosion and deposition processes, while the Stream Power Index (SPI) indicates potential flow erosion at specific topographic points.

The Topographic Roughness Index (TRI) measures elevation differences between adjacent DEM cells, calculating the difference in elevation values from a center cell and the eight surrounding cells. The Topographic Position Index (TPI) categorizes topographic positions into upper, middle, and lower landscape segments. The Direct Duration Radiation layer indicates the length of incoming solar radiation at each location, while the NDVI measures the greenness and density of vegetation captured in satellite images.

Layers related to soil type, land use, and geology were processed using tools such as polygon-to-raster, copy raster, and float. The distances to structures, roads, streams, and faults were calculated using the distance accumulation tool. The 20-year average annual rainfall data was processed year by year using the Inverse Distance Weighting (IDW) tool in GIS and then summarized using the cell statistics tool.

Data on landslide incidents from 2002 to 2022 were collected from the National Building Research Organization (NBRO) in Sri Lanka for the KD area, while data for TP and WP were obtained from the SABO Prefectural Department. Since 2018, with support from JICA, the NBRO has worked to modify and maintain landslide inventories more effectively, including categorizing different landslide types. For validation, landslide incident data, including types, were gathered for KD from the recently updated NBRO inventory. The number of landslide points is presented in Table 4 below.

Table 4. Number of landslides in each inventory.

SA /RA	All types of LS points	Cliff/Slope failure LS points
KD	214	Not specified before modifying the inventory
WP	641	535
TP	260	167 – To create and train the Model

Landslide susceptibility mapping using data mining methods can be considered a binary classification task. Therefore, the same number of non-landslide points were randomly selected from landslide-free areas and divided using a 70/30 ratio [49]. Landslide points were labeled as 1, and non-landslide points as 0. A layer was created based on 167 cliff-type landslide points and 167 non-landslide points to run the model. Non-landslide points were generated using tools such as buffer, erase, and create random points in GIS Pro.

All layers were processed to have the same extent, raster format, pixel type, pixel depth, and cell size of 33.952976 m, and were set to use the WGS 1984 Web Mercator (auxiliary sphere) coordinate system. Tools used for processing included copy raster, clip raster, float, define projection, and project raster.

The model was developed using the Forest-based and Boosted Classification and Regression (FBCR) tool. This process involved inputting 24 explanatory variable layers along with a point layer representing cliff types, categorized as landslide (LS) and non-landslide (non-LS), for a total of 334 points. The prediction type selected was "train only," and the model utilized was a gradient boosted model, which constructs a series of sequential decision trees. Each subsequent decision tree is designed to minimize the error (bias) of the previous tree. As a result, the gradient boosted model effectively combines several weak learners to create a robust predictive model.

For training, the input feature consisted of the cliff type point layer, while the variable to be predicted was the occurrence of LS in cliffs. The explanatory training datasets were loaded, and for land use, soil type, and geology layers, the categorical box was checked. Output files—including the trained model, trained features, variable importance table (VIT), confusion matrix (CM), and validation table—were saved in a geodatabase. The training data exclusion rate for validation was increased from 10% to 30%. Environment settings were configured, and the model was trained by adjusting various parameters as shown table 5 below. Throughout the training process, key metrics such as accuracy, sensitivity, Matthews correlation coefficient (MCC), F1 score, mean, median, standard deviation, and the shape of the histogram were monitored.

Table 5. Advanced Model option parameters [38,39].

No	Parameter	Valid Values	Default	Recommendation	Value used
1	Number of Trees	An integer greater than 1.	Based on the number of features and variables. For regression, it is one-third of the total explanatory variables; for classification, it is the square root of the total variables.	Use as many trees as your machine allows for better accuracy.	500
2	Minimum Leaf Size	An integer greater than 1.	5 for regression and 1 for classification.	Increase for large datasets to reduce runtime.	5
3	Data Available per Tree (%)	An integer between 1 and 100.	100% of the data, but each tree uses approximately two-thirds of the training data randomly.	Lower percentages can decrease runtime for large datasets.	12
4	Number of Randomly Sampled Variables	An integer less than or equal to the total explanatory variables.	For regression, one-third of explanatory variables; for classification, the square root of explanatory variables.	Recommendation: Increasing this number may lead to overfitting, especially if dominant variables exist.	5
5	Lambda	A double greater than or equal to 0.	NA	Controls regularization strength; higher values reduce overfitting.	5
6	Gamma	A double greater than or equal to 0.	NA	Higher values make the model more conservative in splitting.	0.01
7	Eta	A double between 0 and 1.	NA	Lower values (eg, 0.1) improve accuracy but increase runtime.	0
8	Maximum Number of Bins	An integer greater than 1 or 0	NA	Higher values allow finer splits but increase runtime.	0.01

Explanatory variables identified as unimportant were found to potentially impact the model's accuracy and other parameters. Consequently, geology, soil type, and land use conditioning factors (LCF) were removed due to their low importance, as illustrated in Figure 12. The model was finalized once all output parameters met established criteria.

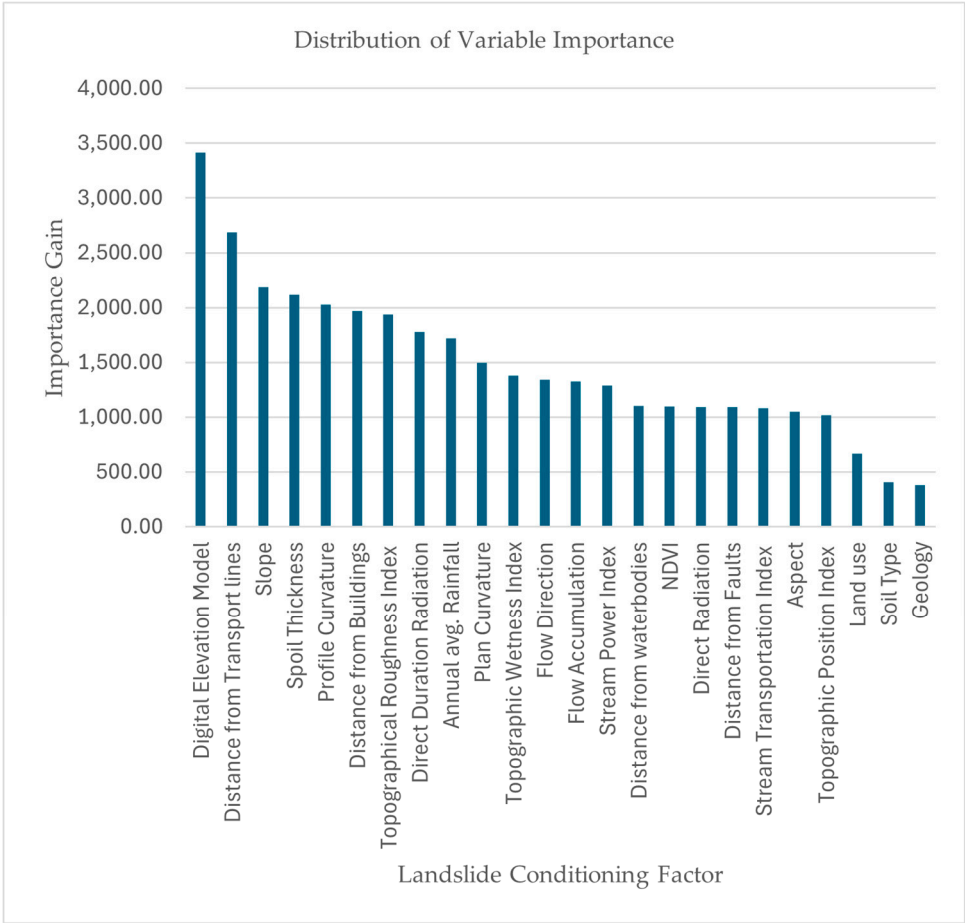


Figure 12. Distribution of Variable Importance when Model is under training.

2.5. Prediction and Validation of Cliff-Type Landslides in the Study Area Inventory

To predict and validate cliff-type landslides in the study area, the satisfactorily trained model was utilized to generate a raster prediction, which was then saved in the geodatabase. This prediction focused on the target area (TP) for validation, as only 70% of the cliff LS points were used for training. The column containing explanatory raster predictions was updated using KD layers, and the model was executed for KD prediction. A cliff-type LS point layer was created with GIS tools, referencing the TP inventory. This layer was overlaid with the predicted output layer of the TP subarea for model validation. Similarly, a cliff-type LS point layer was created using the recently modified KD inventory and overlaid for validation. The model was also employed to predict cliff-type LS within the KD inventory prior to its modification.

Figure 13 below presents the methodological flowchart developed for identifying cliff-type landslide susceptibility. The process begins with the selection and preparation of input datasets, which include 21 landslide conditioning factor layers and landslide inventory points. The model is trained using machine learning classifiers—specifically, the Forest-based and Boosted Tree algorithms—within a defined training area (TP). Validation is conducted using standard performance metrics such as accuracy and sensitivity. The validated model is then applied to the target area (KD) using the same set of conditioning factors. Finally, predictions are compared with a

recently updated landslide inventory to evaluate the model's reliability and accuracy. This structured approach ensures the systematic generalization of the model across different spatial domains.

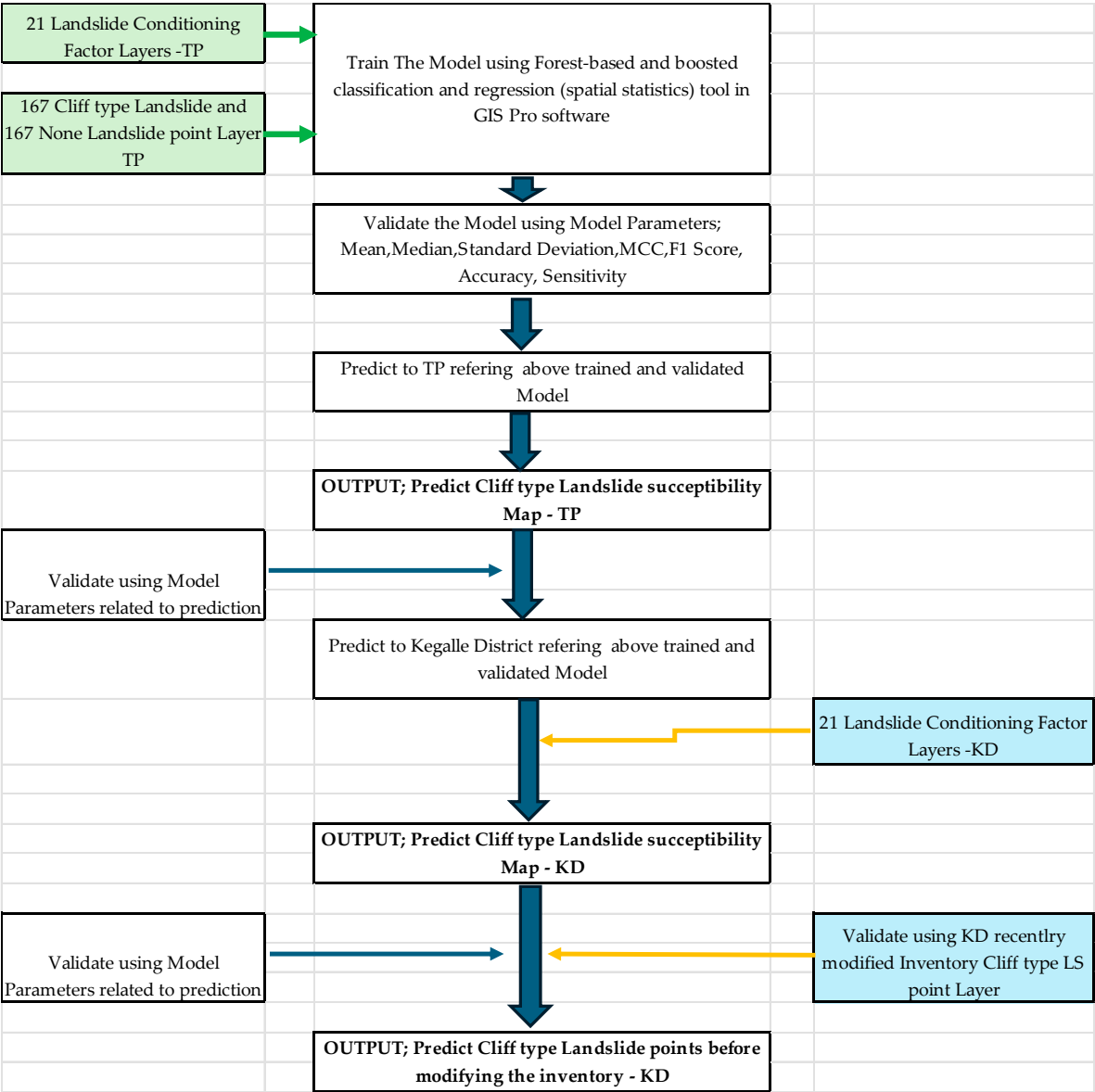


Figure 13. Methodology Flow Chart (TP; Tokushima Prefecture, KD; Kegalle District).

3. Results

3.1. Selected Landslide Conditioning Factors

Based on the comparative analysis of landslide conditioning factors considered by past authors (Figure 10), a total of 25 factors were selected for the present study. Twelve of these were chosen from variables that had been considered by more than 25% of previous studies, reflecting their frequent use and recognized relevance in landslide susceptibility assessment. These include land use, average annual rainfall, elevation, aspect, slope, profile curvature, topographical wetness index (TWI), sediment transportation index (STI), stream power index (SPI), normalized difference vegetation index (NDVI), geology, and distance to faults. In addition, another 12 factors with lower author consideration (< 25%) were selected based on their perceived relevance to the geomorphological and environmental setting of the study areas, as informed by field knowledge and expert judgment. These include distance to roads, plane curvature, distance from water bodies, topographical position index

(TPI), topographical roughness index (TRI), direct radiation, direct duration radiation, distance from structures, flow accumulation, flow direction, soil type, and soil thickness. A further variable, distance from earthquake-induced landslide epicentres, was included to account for seismically triggered slope instabilities. The final set of factors thus balances widely recognized predictors from the literature with context-specific variables that address the unique triggering conditions of the study regions.

3.2. Trained Model Referring Tokushima Prefecture Using FBCR Tool

3.2.1. Input Training Features to Train the Model

The Figure 14 illustrates the spatial distribution of cliff-type landslides (red triangles) and non-landslide locations (black dots) across Tokushima Prefecture, Japan, with the study boundary outlined in black. Landslides are predominantly concentrated in the central and western mountainous regions, particularly around Miyoshi and along the southern slopes of the northwestern highlands, often following steep terrain and ridge lines. Scattered occurrences are also visible along the eastern coastal zone, though at a lower density. The proximity of many landslides to major roads, suggests potential human-induced triggers from slope cutting and infrastructure development. In contrast, the southeastern coastal belt shows relatively few landslides, likely due to flatter topography and different geological conditions. The inclusion of both landslide and non-landslide points ensures comprehensive spatial coverage for susceptibility analysis and model validation.

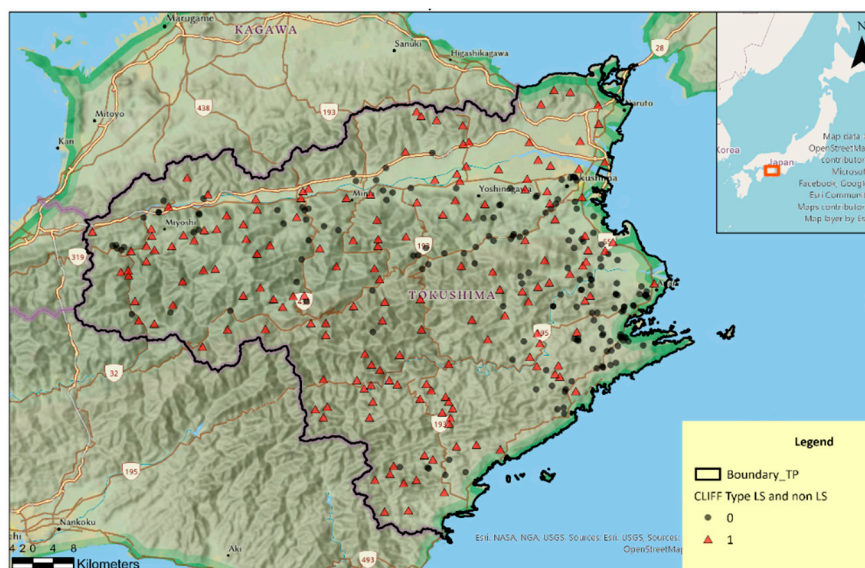


Figure 14. Cliff type LS occurred, and none occurred point layer of TP.

The Digital Elevation Model (DEM) map of Tokushima Prefecture Figure 15 below shows the variation in terrain elevation within the study boundary, outlined in black. Elevation ranges from near sea level (0 m, shown in green to blue shades) along the coastal plains to the highest peaks (up to 1,949 m, shown in red) concentrated in the central and southwestern mountainous regions. The steep, high-relief areas, predominantly in the west and central parts, transition to moderately elevated foothills before flattening out toward the eastern and northern coasts. This elevation gradient reflects the region's rugged interior and low-lying coastal zones, which play a significant role in influencing slope stability, drainage patterns, and the spatial distribution of landslides.

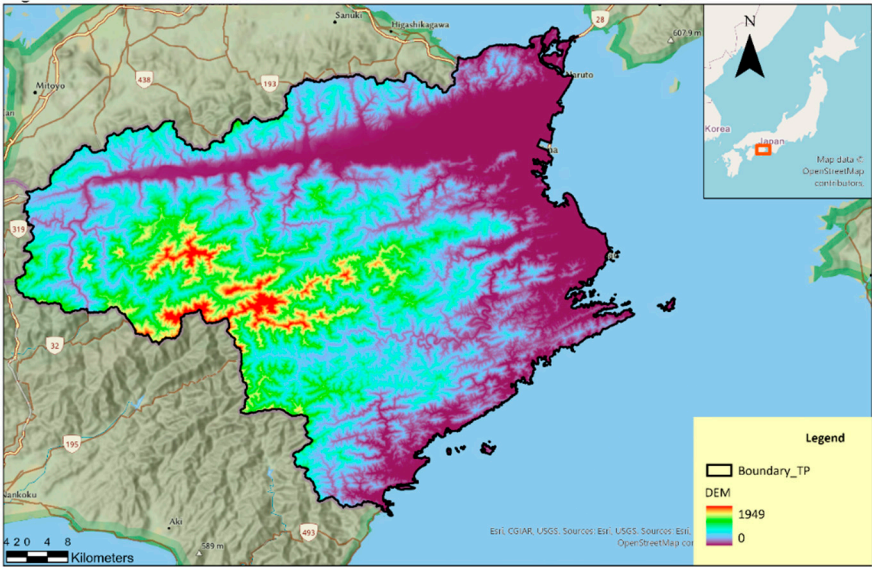


Figure 15. Digital Elevation Model of TP.

In the slope layer map (Figure 16), a graduated color scheme was employed where the color brown deepens with increasing slope, indicating variations in slope steepness.



Figure 16. Slope layer of TP.

Referring to the Profile Curvature Map in Figure 17, positive values indicate areas where surface flow is accelerating, and erosion is occurring. In contrast, negative profile curvature values suggest areas where surface flow is slowing down and deposition is taking place.

The Plan Curvature Map in Figure 18 indicates that a positive value signifies a laterally convex surface at that cell, while a negative plan curvature indicates a laterally concave surface. A value of zero denotes that the surface is linear.

The Direct Duration Radiation Map in Figure 19 represents the duration of direct solar radiation received at each location, measured in hours, indicating how long a location is exposed to direct sunlight over a specified time period.

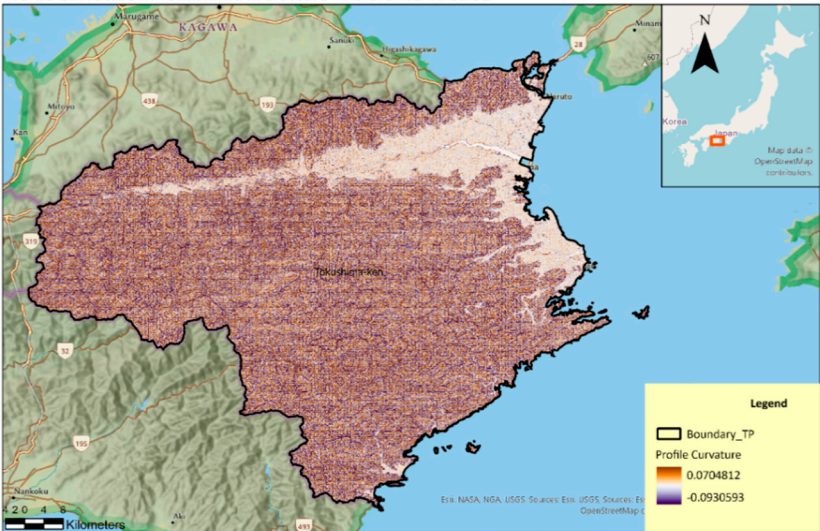


Figure 17. Profile curvature of TP.



Figure 18. Plan curvature of TP.

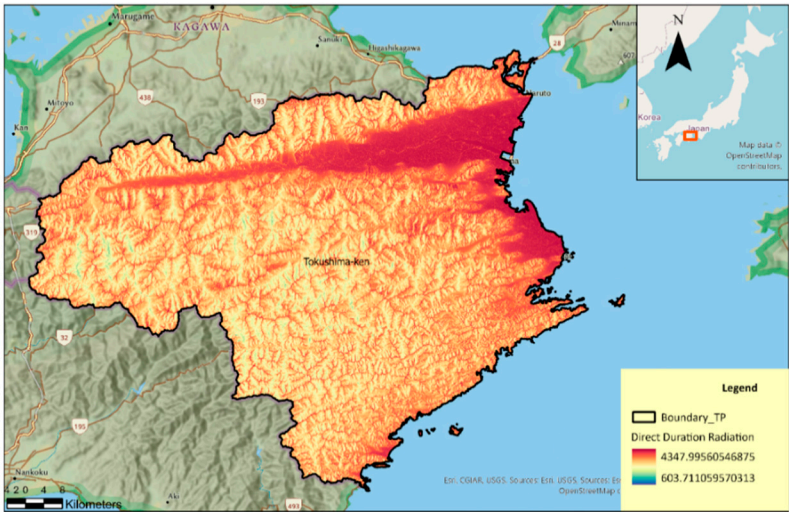


Figure 19. Direct Duration Radiation of TP.

As depicted in Figure 20, the direct radiation layer outputs the direct incoming solar radiation value for each location, measured in kilowatt-hours per square meter (kWh/m²). This data helps in understanding the amount of solar energy received directly from the sun at various surface locations.

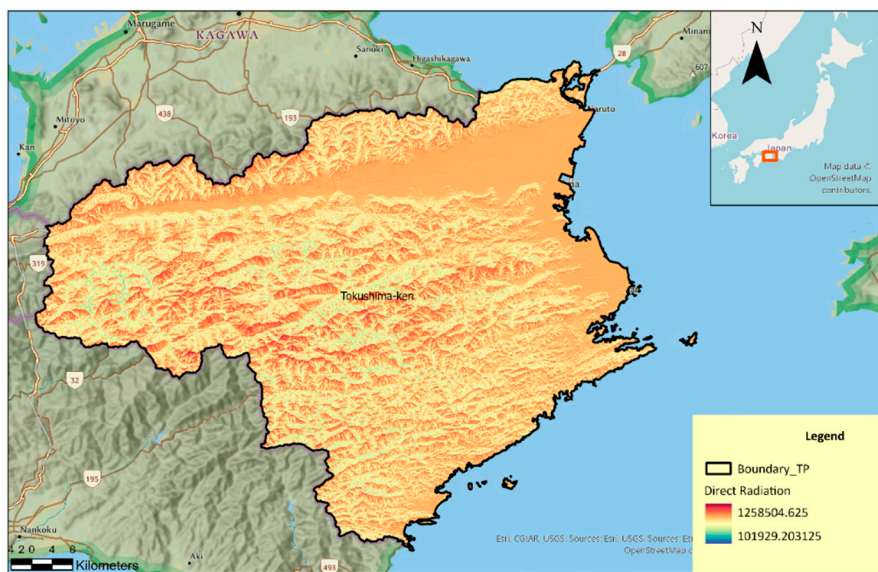


Figure 20. Direct Radiation of TP.

The Flow Direction Map in Figure 21 illustrates the direction of flow from each cell to its downslope neighbor(s), utilizing an eight-direction (D8) approach, which accounts for the eight adjacent cells into which flow could travel.

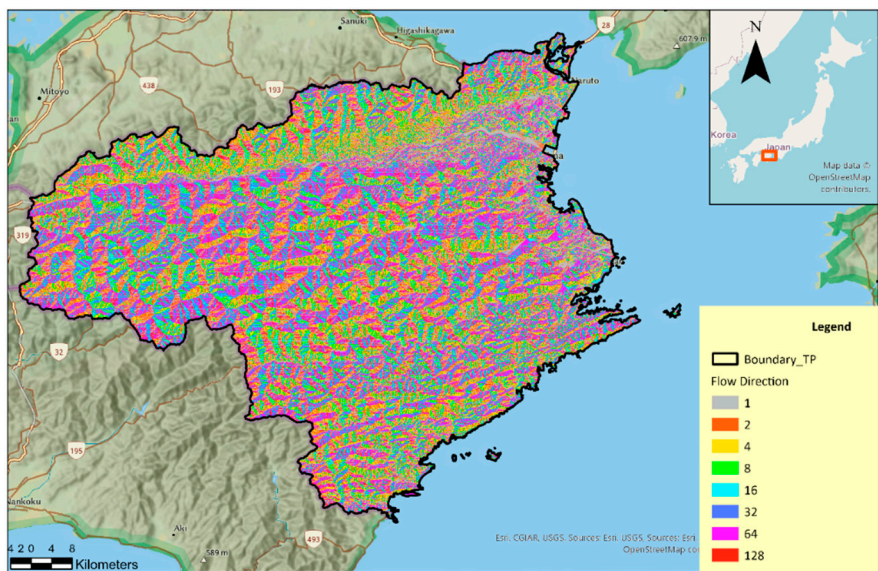


Figure 21. Flow Direction of TP.

The Flow Accumulation Map in Figure 22 represents the accumulation of flow to each cell in the output raster; higher values indicate areas with more accumulated flow, typically found in valleys or stream channels. This map features a color ramp ranging from light blue for low accumulation to dark blue for high accumulation.

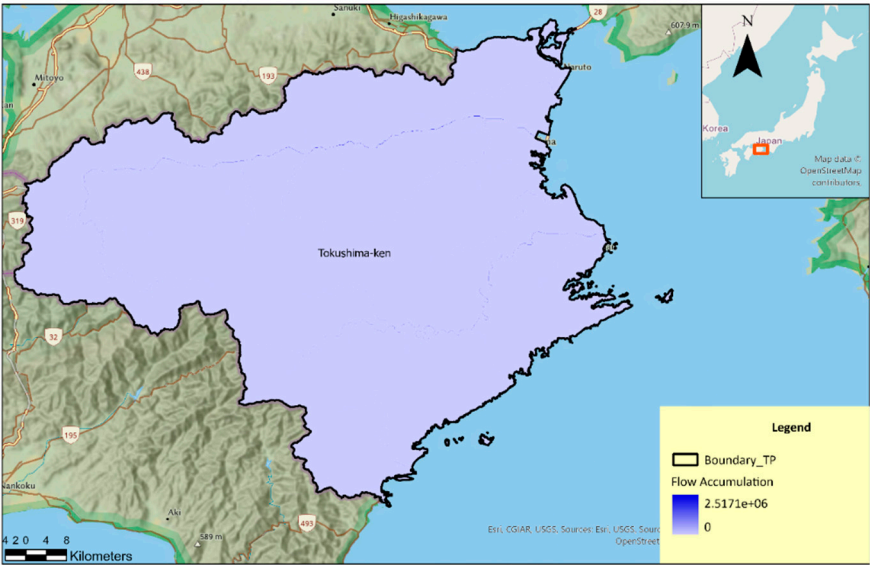


Figure 22. Flow Accumulation of TP.

Figure 23 presents the Topographic Wetness Index (TWI), which measures the spatial distribution of soil moisture. Higher TWI values suggest wetter areas, while lower values indicate drier regions. The wetness distribution is represented using a gradient color ramp, ranging from yellow for dry areas to blue for wet areas.

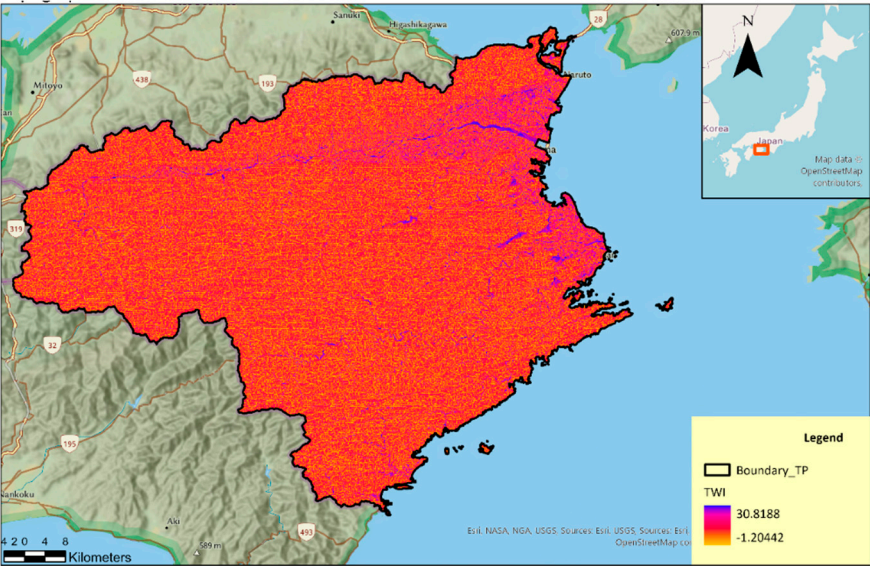


Figure 23. Topographical Wetness Index of TP.

The Topographical Position Index Map in Figure 24 shows that positive TPI values indicate a cell that is higher than its surroundings, typically representing ridges or hilltops. Negative TPI values suggest a cell that is lower than its surroundings, typically indicating valleys or depressions. When TPI values are near zero, it typically represents flat or gently sloping areas.

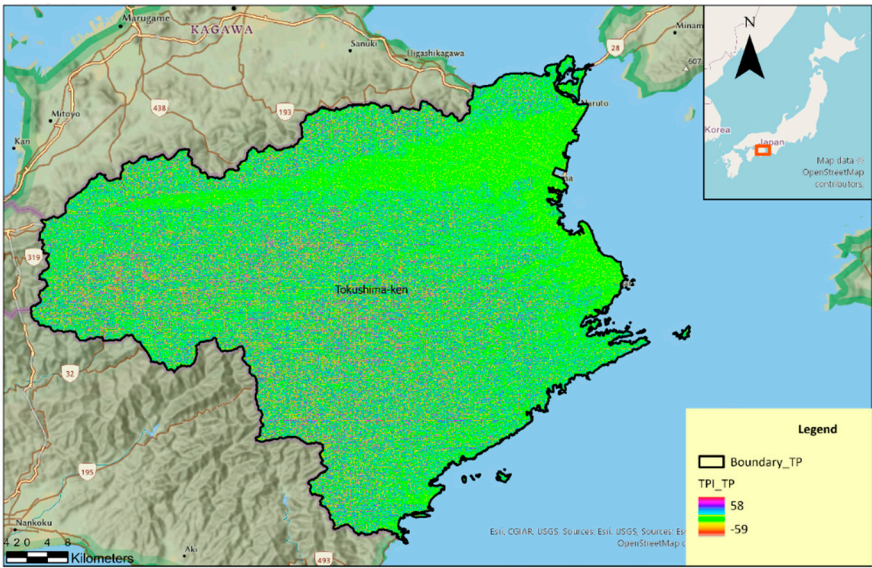


Figure 24. Topographical Position Index of TP.

In the Sediment Transportation Index Map (Figure 25), high STI values indicate areas with a high potential for sediment transport, generally corresponding to wetter conditions, while low STI values suggest areas with low sediment transport potential, typically indicating drier conditions.



Figure 25. Sediment Transportation Index of TP.

As seen in Figure 26 of the Stream Power Index (SPI) map, higher SPI values indicate areas with a greater potential for erosion, which are typically found in steep and high-flow regions. Positive values suggest an increased erosive potential, while negative values indicate areas of sediment accumulation. The distribution of stream power is depicted using a gradient color ramp, ranging from light brown for low power to dark brown for high power.

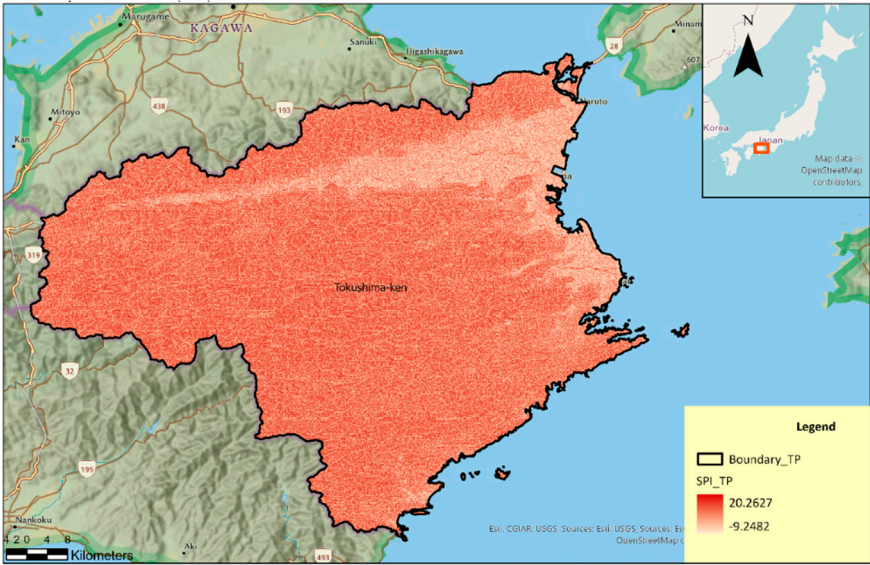


Figure 26. Stream Power Index of TP.

Figure 27 illustrates the Topographical Roughness Index (TRI), where high TRI values reflect significant elevation changes between adjacent cells, indicating rough terrain such as steep slopes, cliffs, or mountainous regions. Conversely, low TRI values represent areas with minimal elevation changes, referring to flat or gently rolling terrain.

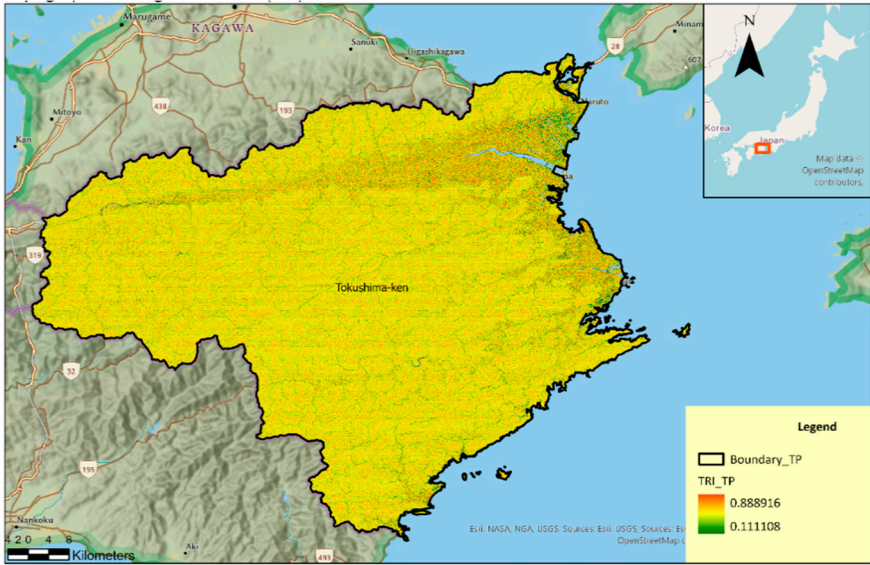


Figure 27. Topographical Roughness Index of TP.

To further characterize conditioning factors, Figures 28–31 thematic maps were prepared to represent Euclidean distances from major roads, buildings, water bodies, and active geological faults. Each map employs a graduated color scale, with areas in close proximity to the respective feature shown in red, transitioning through orange, yellow, green and finally to light pink at the greatest distances. The distance-from-roads and distance-from-buildings layers identify zones of potential human exposure and accessibility, relevant for assessing both vulnerability and post-disaster response feasibility. The distance-from-water-bodies layer captures hydrological influences, including potential erosion and flood-driven slope destabilization. The distance-from-faults layer incorporates tectonic and seismic triggers, as faults can induce ground shaking or weaken slope materials. Collectively, these datasets provide a comprehensive spatial framework for landslide

susceptibility analysis, enabling the integration of geomorphological, hydrological, seismic, and anthropogenic variables in predictive modeling.

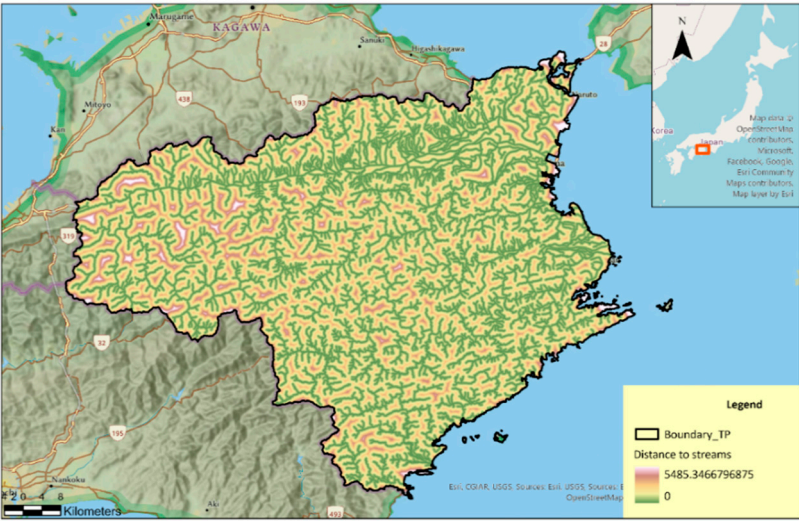


Figure 28. Distance from Streams of TP.

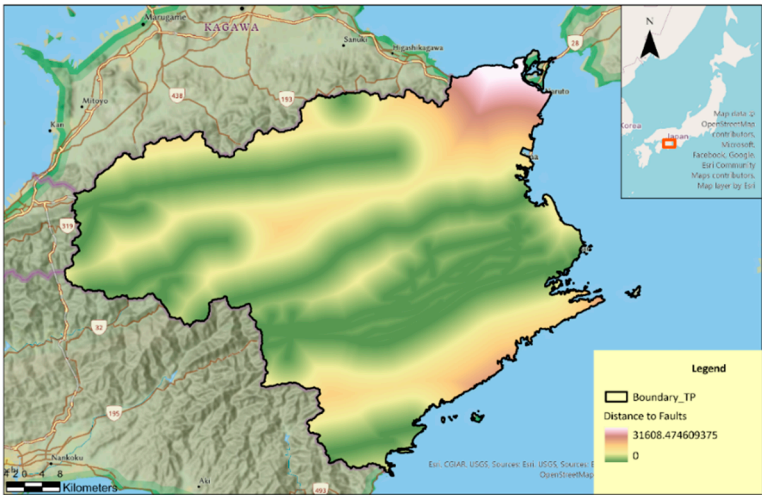


Figure 29. Distance from Faults of TP.

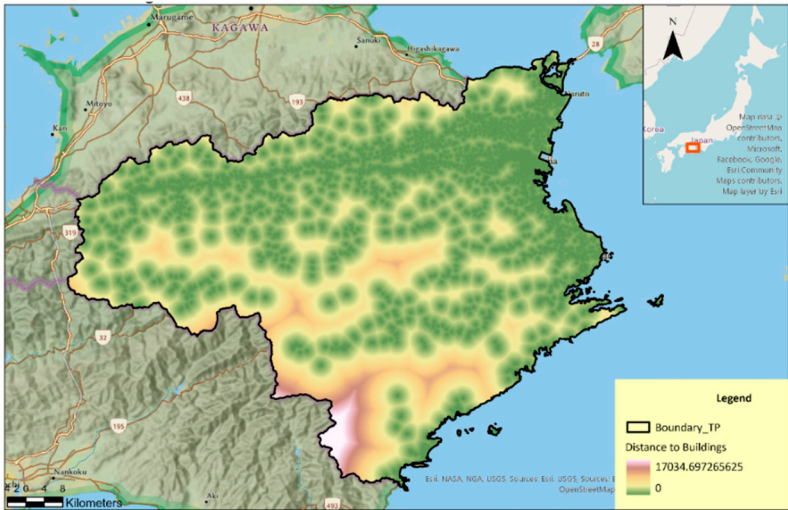


Figure 30. Distance from Buildings of TP.

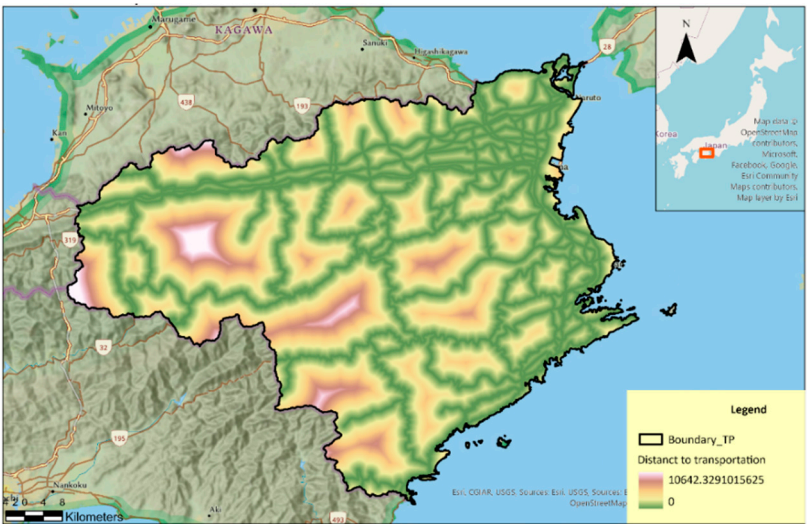


Figure 31. Distance from Roads of TP.

As shown in Figure 32, aspect also referred to as exposure, indicates the compass direction a terrain surface faces.

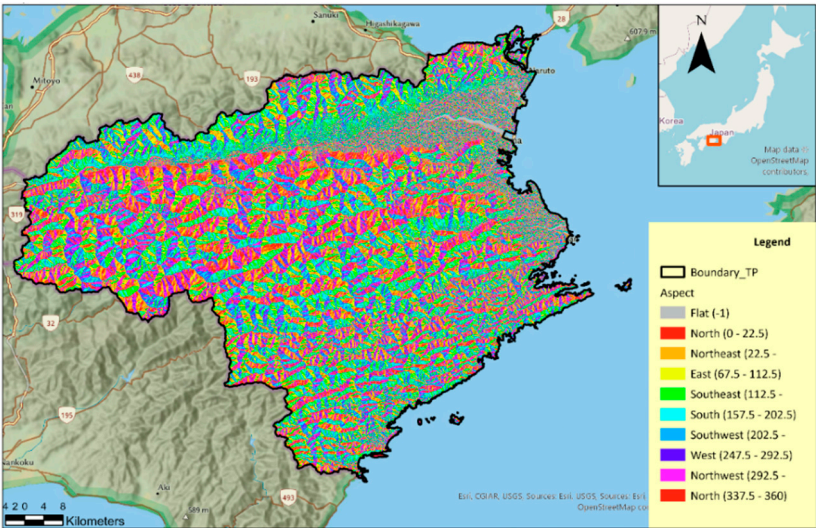


Figure 32. Aspect of TP.

Figure 33 derived from geotechnical datasets, ranges from 113 cm (light shades) to 200 cm (dark red), highlighting spatial variability in regolith depth. Thicker soils, particularly in low- to mid-slope zones, may store greater volumes of water during heavy rainfall events, increasing pore-water pressure and enhancing the likelihood of slope failure. In contrast, thin soils over bedrock in upper slopes may be more prone to shallow slides or debris flows.

Figure 34 displays the Normalized Difference Vegetation Index (NDVI). NDVI values typically range from -1 to 1. Values close to 1 indicate healthy, dense vegetation, while values close to 0 suggest barren areas, such as rock, sand, or urban environments. Negative values represent water, snow, or clouds. A color scheme is used to depict this variation, with green indicating healthy vegetation and red representing barren areas.

As in Figures 35 and 36 above, average annual rainfall data from various meteorological stations were collected, and year-by-year average annual rainfall maps were created using the IDW tool. All 20 maps were then summarized into one map using the cell statistics tool. Annual average rainfall distribution is ranging from 124.74 mm (light blue) to 286.87 mm (dark blue). Higher precipitation is concentrated in central and southwestern zones, coinciding with mountainous catchments, while

coastal lowlands receive comparatively lower rainfall. Given that intense or prolonged rainfall is a primary landslide trigger in humid mountain regions, this factor is critical for susceptibility modeling.

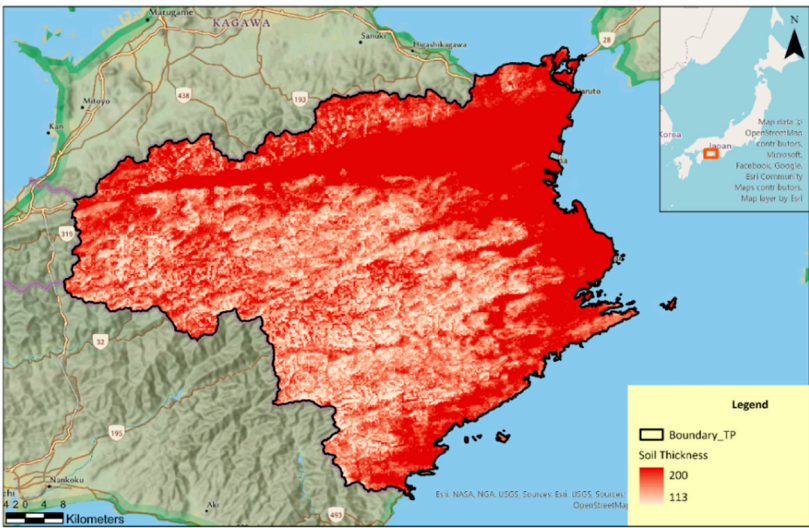


Figure 33. Soil Thickness of TP.

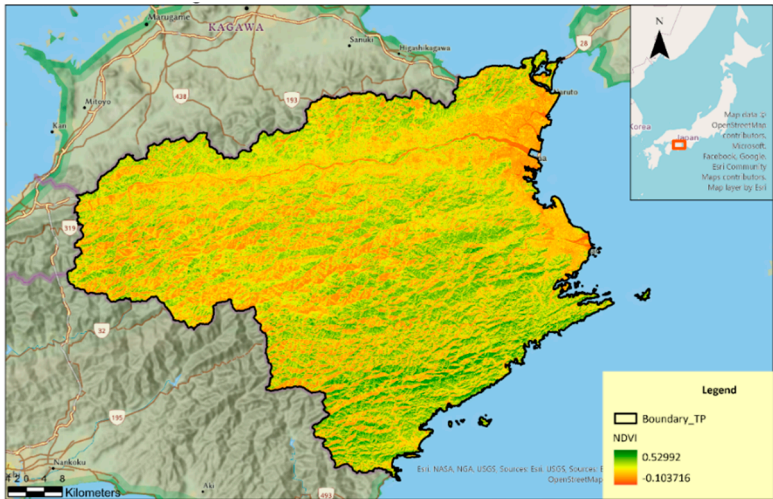


Figure 34. Normalized Difference Vegetation Index of TP.

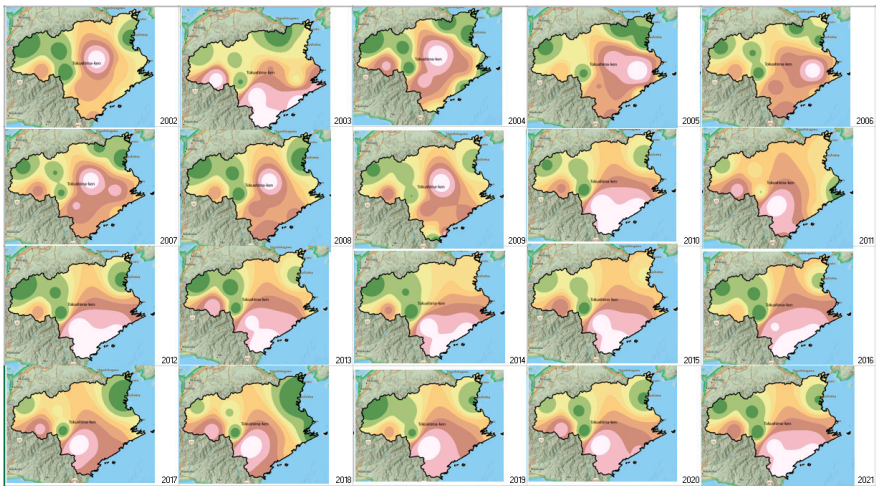


Figure 35. Year-by-year Annual Average Rainfall layers from 2002 to 2021 of Tokushima Prefecture.

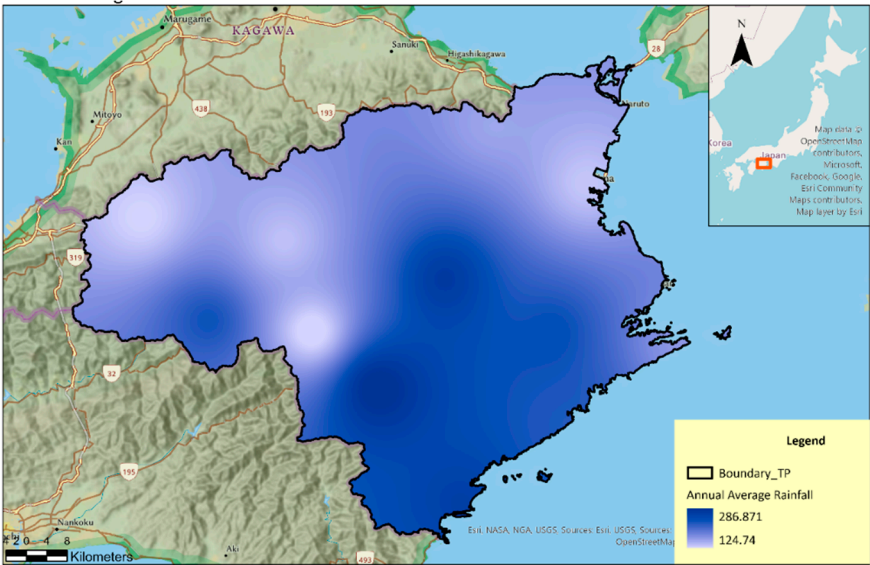


Figure 36. Sum of Annual Average Rainfall for 20 years of TP.

3.2.2. Trained Model Output

As shown in Table 6, F1-Score: Combines precision and recall into one metric, giving a balanced measure even if class distributions are uneven. Category 0: 0.82, Category 1: 0.85, Overall: 0.84 — these values are relatively high, meaning the model predicts both classes with good balance between false positives and false negatives. MCC (Matthews Correlation Coefficient) : Ranges from -1 (total disagreement) to +1 (perfect prediction), with 0 meaning no better than random guessing. MCC = 0.68 for both categories and overall — a strong positive correlation, indicating reliable classification across both classes.

Table 6. Validation Data Classification Diagnostics.

Category	F1-Score	MCC	Sensitivity	Accuracy
0	0.82	0.68	0.84	0.84
1	0.85	0.68	0.84	0.84
All	0.84	0.68	NA	0.84

Sensitivity (Recall / True Positive Rate): Measures the proportion of actual positives correctly identified. Both categories have 0.84 sensitivity, meaning the model captures 84% of true instances for each class. Accuracy : The proportion of total predictions (both positive and negative) that are correct. Accuracy = 0.84 across both categories and overall — suggesting that the model performs equally well for each class.

The histogram (Figure 37) illustrates the frequency distribution of accuracy scores obtained across multiple iterations. Accuracy values ranged from approximately 0.71 to 0.88, with the majority concentrated between 0.77 and 0.84. The model achieved a mean accuracy of 0.8012 and a median accuracy of 0.8000, indicating stable predictive performance with minimal skewness. The standard deviation of 0.03539 reflects relatively low variability across runs, suggesting that the model is robust and generalizes well to unseen data. This consistency in accuracy highlights the suitability of the selected predictor variables and model configuration for assessing cliff-type landslide susceptibility in the study area.

The classification performance of the Random Forest model for cliff-type landslide prediction in Tokushima Prefecture is further illustrated by the predicted versus actual class distribution (Figure 38). The plot shows the proportion of correctly and incorrectly classified samples for each predicted category. Class “0” (non-landslide areas) constitutes the majority of the dataset and exhibits high classification accuracy, with a small proportion misclassified as landslides. Similarly, class “1”

(landslide-prone areas) is accurately identified in most cases, with minimal false negatives. This balanced performance across classes indicates that the model is effective at distinguishing between landslide and non-landslide areas despite the potential class imbalance, thereby reinforcing its reliability for susceptibility mapping.

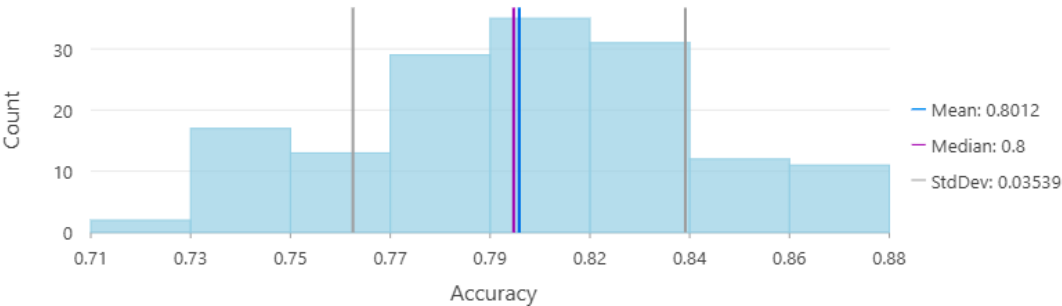


Figure 37. Validation Accuracy Graph of the trained Model sequentially.

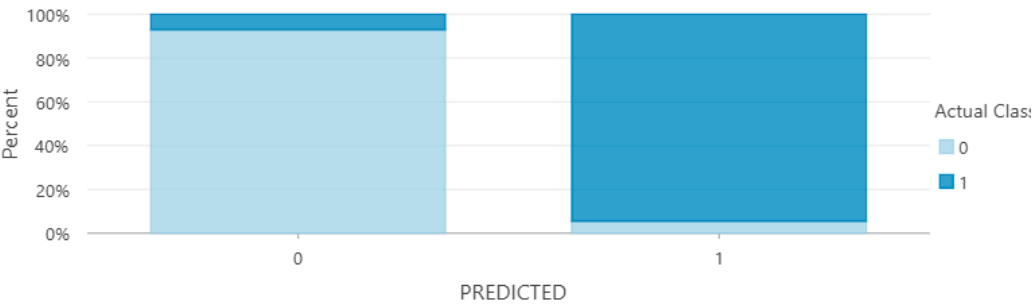


Figure 38. Prediction Performance Graph:.

Figure 39 illustrates the classification results of the Random Forest model for cliff-type landslide occurrence in Tokushima Prefecture, represented through a categorical confusion matrix. The vertical axis corresponds to the actual occurrence of landslides (1 = occurrence, 0 = non-occurrence), while the horizontal axis represents the model’s predicted labels. The color intensity denotes the count of samples within each category, with darker shades indicating higher frequencies. The model successfully predicts a large proportion of non-landslide areas (upper-left cell) and landslide-prone areas (lower-right cell), with relatively fewer misclassifications in the off-diagonal cells. These results indicate a strong agreement between predicted and actual classes, further supporting the model’s robustness and reliability in landslide susceptibility assessment.

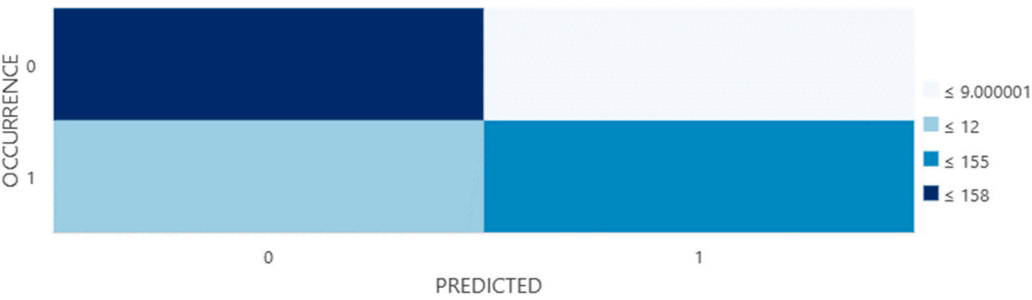


Figure 39. Confusion Matrix.

Figure 40 presents the distribution of predicted versus actual classes for the Random Forest model applied to cliff-type landslide occurrence mapping in Tokushima Prefecture. The horizontal

axis represents the actual class labels (CAT_0 = non-occurrence, CAT_1 = occurrence), while the vertical axis indicates the percentage of samples. The color segments within each bar denote the proportion of predicted classes (CAT_0 in light blue and CAT_1 in dark blue). The results demonstrate that a high proportion of non-occurrence areas (CAT_0) were correctly classified, with only a small fraction being misclassified as landslide occurrences. Similarly, the majority of landslide-prone locations (CAT_1) were accurately identified, with limited misclassification into the non-occurrence category. This distribution highlights the model’s effectiveness in distinguishing between stable and landslide-susceptible zones.

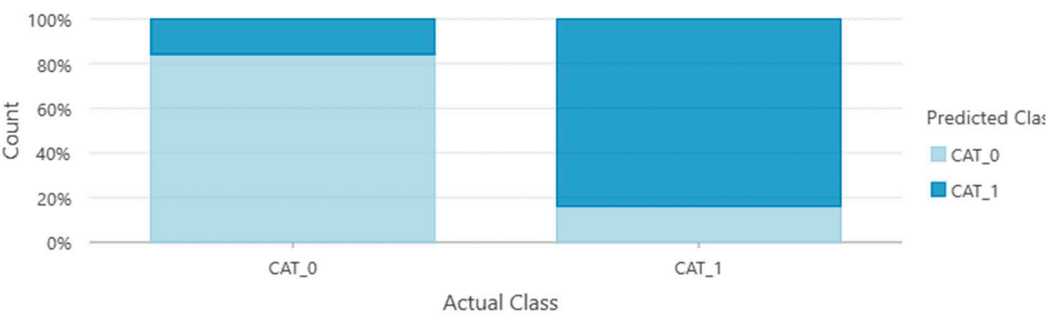


Figure 40. Validation Performance Graph of the trained Model.

Figure 41 illustrates the relative importance of landslide conditioning factors as determined by the Random Forest model, measured using the importance gain metric. The Digital Elevation Model (DEM) emerged as the most influential predictor, followed by distance from transportation networks, soil thickness, slope, and distance from buildings. Profile curvature and TRI measures also exhibited notable contributions. Climatic and hydrological variables, such as direct duration radiation, annual average rainfall, and the topographic wetness index, ranked moderately in importance. In contrast, factors such as NDVI, distance from faults, and aspect contributed relatively less to the model. The results indicate that topographic and proximity-related parameters play a dominant role in predicting landslide susceptibility in the study area, whereas vegetation and fault-related factors exert comparatively lower influence.

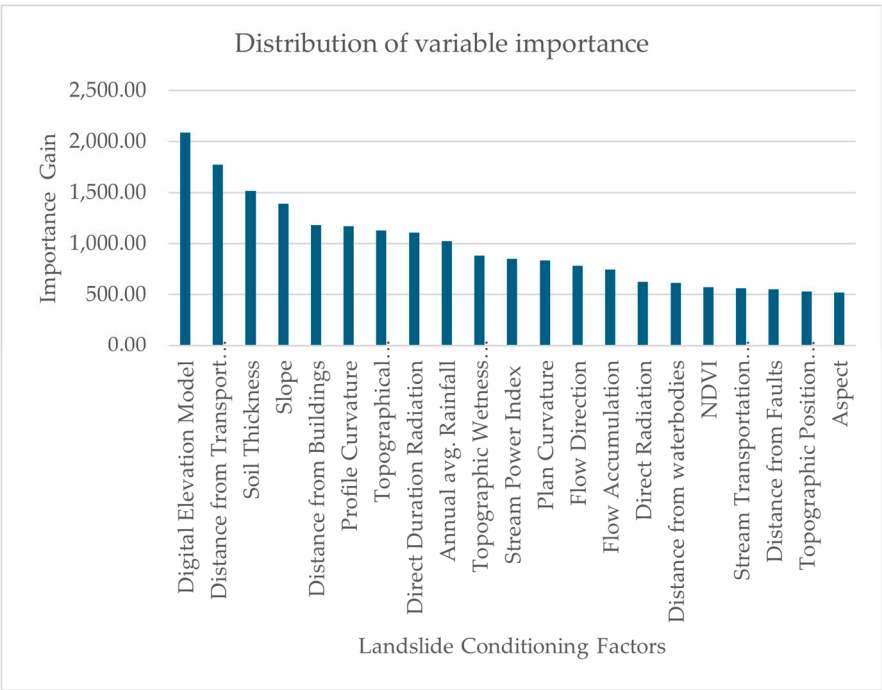


Figure 41. Distribution of Variable Importance of the trained Model.

3.2. Predict to Tokushima Prefecture Subarea

3.2.1. Predicted Raster Layer

Using the Forest-based and Boosted Classification and Regression tool in GIS Pro, the prediction type was changed from "train only" to "predict to raster." The raster output was saved in a geodatabase. The subarea of Tokushima Prefecture (TP)—where significant cliff landslides (LS) occurred in the past—was selected for the prediction, as illustrated in Figure 42. All 21 layers were generated using the clip raster tool in GIS. Twenty-one explanatory variable layers (Figure 43) were created for the TP subarea and were matched with the explanatory raster in the trained model using the FRBC tool.

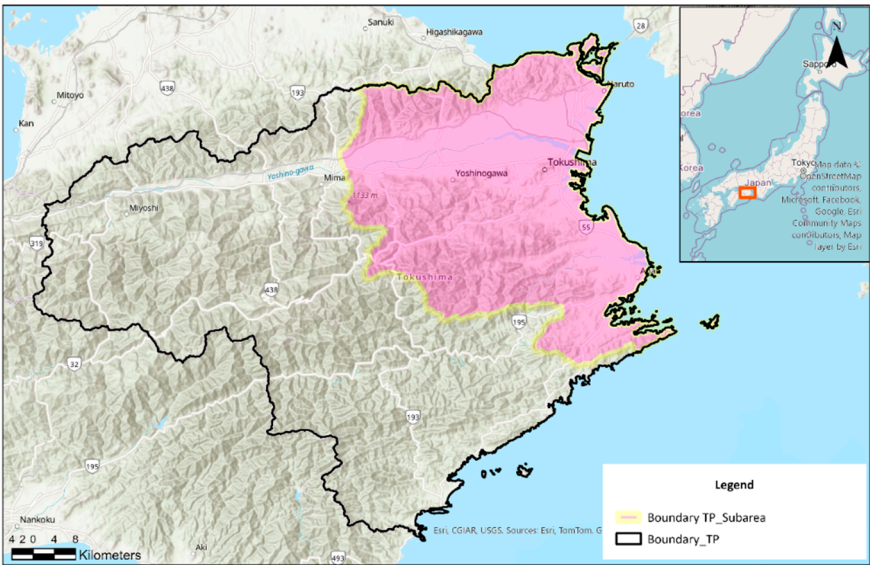


Figure 42. Selected subarea of TP for the prediction.

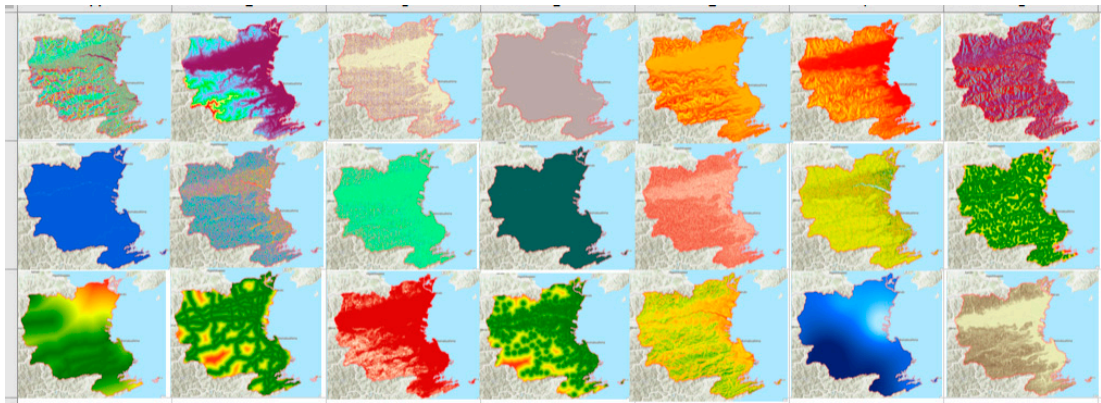


Figure 43. Aspect, DEM, Profile Curvature, Plan Curvature, Direct Duration Radiation, Direct Radiation, Flow Direction, Flow Accumulation, TWI, TPI, STI, SPI, TRI, Distance from Streams, Distance from Faults, Distance from Buildings, Distance from Roads, Soil thickness, NDVI, Annual Average and Slope of TP subarea sequentially.

The map in Figure 44 illustrates the spatial distribution of predicted cliff-type landslide (LS) susceptibility within the TP subarea of Tokushima Prefecture, as derived from the applied predictive model. The study area boundaries are delineated in grey. The model outputs are categorized into two classes: Cliff LS Area (pink) and Non-Cliff LS Area (green).

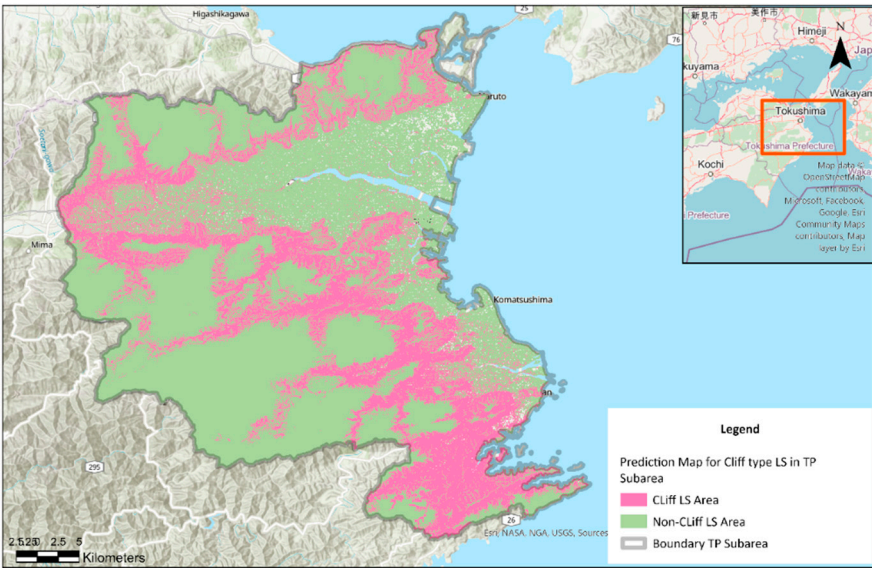


Figure 44. Predicted Map for Cliff type LS in subarea 01 of TP by the Model.

3.2.2. Validate the Predicted Layer Referring TP Inventory; Only 70% Points Used to Train the Model

The map in Figure 45 presents the spatial distribution of predicted cliff-type landslide (LS) susceptibility within the TP subarea, alongside the validation of these predictions using an independent landslide inventory. The susceptibility model classifies the landscape into Cliff LS Area (pink) and Non-Cliff LS Area (green) based on conditioning factors such as slope gradient, soil thickness, distance to roads and buildings, profile curvature, direct duration radiation, and rainfall. The overlay of known cliff-type landslide locations from the inventory (yellow dots) allows visual and quantitative assessment of model accuracy. Correctly predicted cliff-type landslide locations are indicated by blue dots, representing 112 out of 118 inventory points (94.9% agreement).

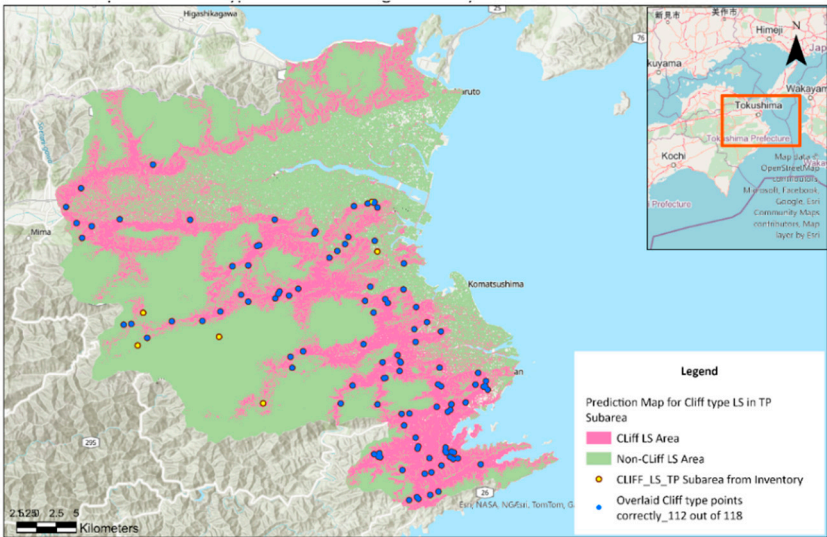


Figure 45. Validation of predicted Cliff type LS area referring TP inventory in subarea.

3.3. Predict to Kegalle District

3.3.1. Predicted Raster Layer

The Forest-based and Boosted Classification and Regression (FRBC) tool in GIS Pro was used to change the prediction type from "train only" to "predict to raster." The raster output was saved in the

geodatabase. A total of 21 layers were created using GIS tools, following the same procedures used to create the Training Prediction (TP) layers described earlier. These 21 explanatory variable layers (see Figure 46) were then loaded into the trained model in the FRBC tool under the "match explanatory raster" option.

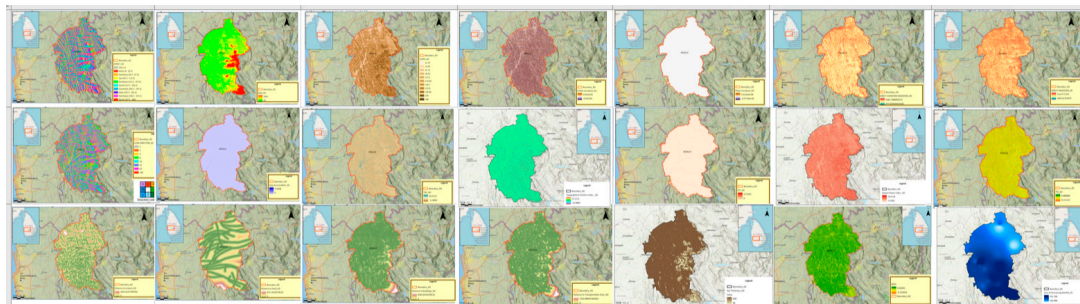


Figure 46. Aspect, DEM, Slope, Profile Curvature, Plan Curvature, Direct Duration Radiation, Direct Radiation, Flow Direction, Flow Accumulation, TWI, TPI, STI, SPI, TRI, Distance from Streams, Distance from Faults, Distance from Buildings, Distance from Roads, Soil thickness, NDVI and Annual Average of KD sequentially.

The map (Figure 47) depicts the spatial distribution of predicted cliff-type landslide (LS) susceptibility across Kegalle District, derived from the applied predictive model using relevant geomorphological, geological, and environmental conditioning factors. The district boundary is outlined in black, while the model outputs classify the terrain into Cliff LS Area (pink) and Non-Cliff LS Area (green).

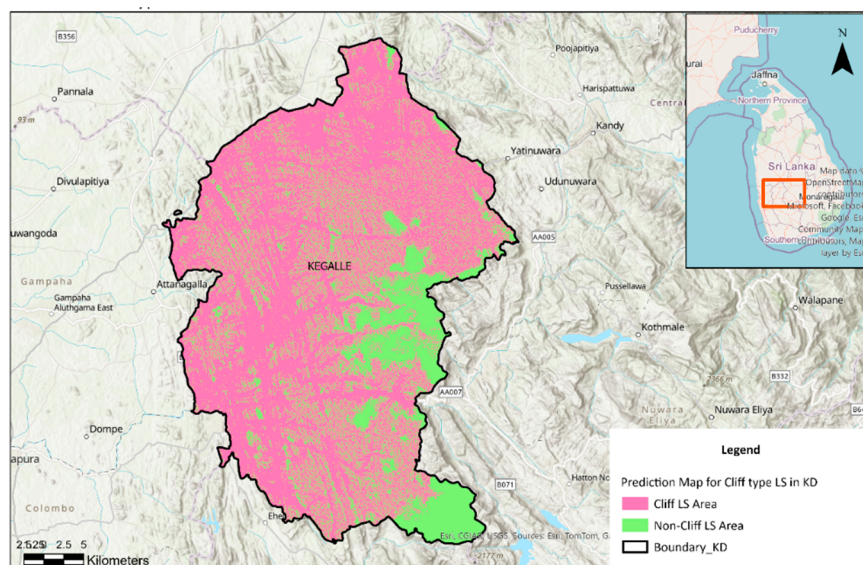


Figure 47. Predicted Map for Cliff type LS in KD by the Model.

3.3.2. Validate the Predicted Layer Referring KD, Recently Modified Inventory Which Is Before the Addition of LS Classification to the Inventory

The Figure 48 presents the predicted cliff-type landslide (LS) susceptibility for KD overlaid with a recently modified landslide inventory that includes multiple landslide types: cutting failures, ground settlement, landslides, rock falls, and slope failures/cliffs. The model output classifies the district into Cliff LS Area (pink) and Non-Cliff LS Area (green), with the boundary of KD shown in black. The distribution of inventory points demonstrates that a large proportion of recorded slope failure/cliff events coincide spatially with areas predicted as high susceptibility. Non-cliff landslide

types are scattered throughout both susceptibility zones, illustrating that while the model is specialized for cliff-type failures, other failure types may also occur in similar geomorphic settings.

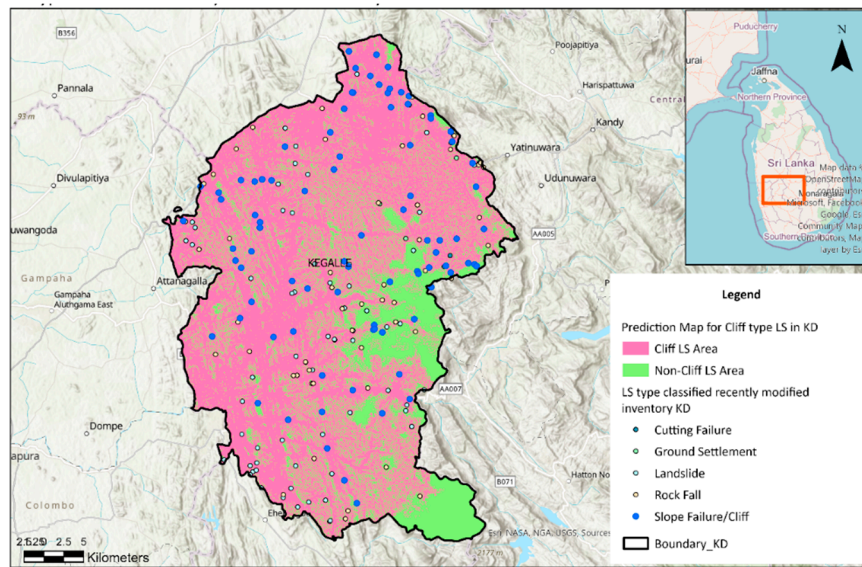


Figure 48. LS type classified recently modified inventory KD map.

The Figure 49 map focuses specifically on the validation of the cliff-type prediction by overlaying only cliff-type LS points from the inventory (brown dots). Correctly predicted points, indicated in blue, show that 72 out of 89 observed cliff-type landslides (80.9%) fall within areas classified by the model as high susceptibility. This high degree of correspondence confirms the model's strong predictive performance for this landslide type in KD.

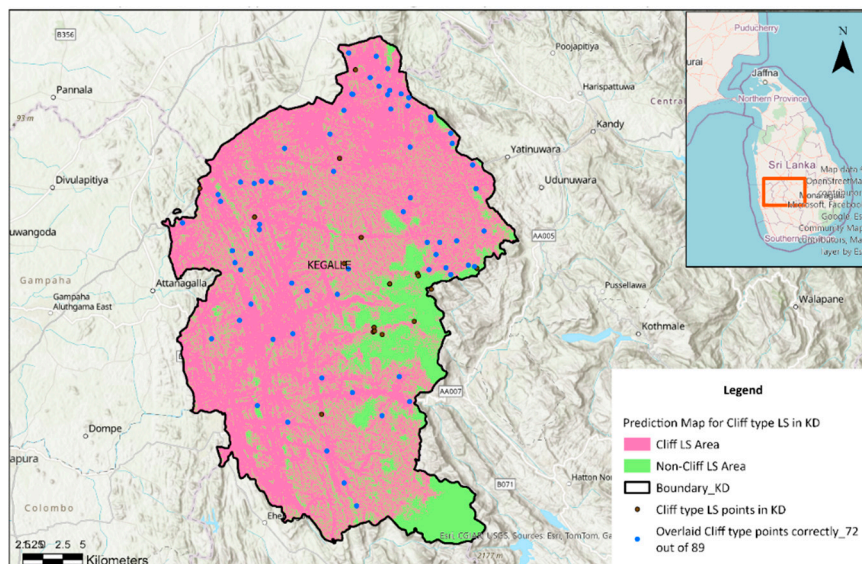


Figure 49. Validation of predicted Cliff type LS area referring recently modified KD Inventory map.

3.3.3. Predict Cliff Type LS Before Modifying the Inventory Incidents in Kegalle District

Figure 50 compares the model-predicted cliff-type landslide susceptibility in KD with the earlier landslide inventory, which contains only spatial point locations without classification of landslide type. The predicted susceptibility is categorized into Cliff LS Area (pink) and Non-Cliff LS Area (green), with the district boundary outlined in black. Inventory points are shown in black, while those

falling within the predicted Cliff LS Area are highlighted in blue. The analysis shows that 115 out of 294 recorded landslide locations from the earlier inventory (39.1%) fall within areas predicted as high susceptibility for cliff-type failures. This percentage matches with the slope failure occurrence percentage referring to the Figure 1.

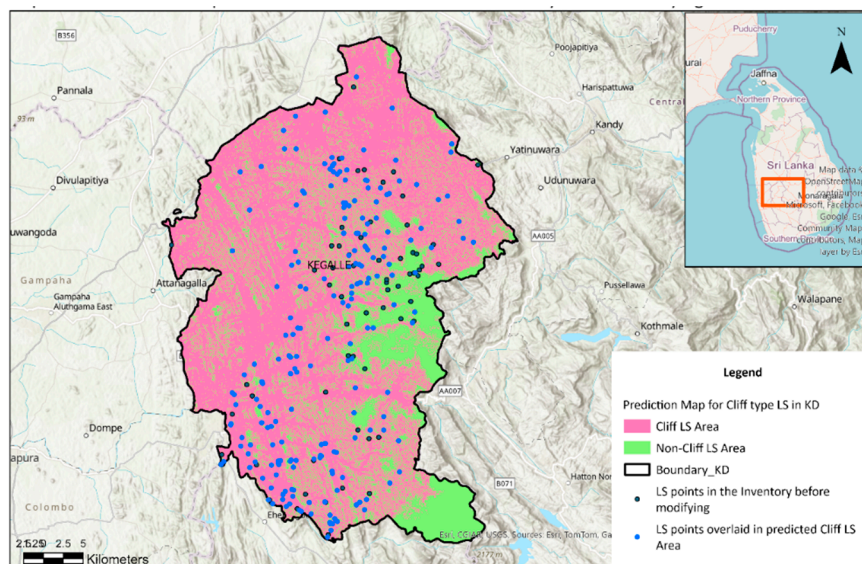


Figure 50. LS points overlaid on the predicted cliff type LS area referring KD inventory before modifying recently.

4. Discussion

Landslides are a pervasive geohazard worldwide, causing extensive human, infrastructural, and environmental losses, with their occurrence governed by complex interactions among climatic, geological, geomorphological, and anthropogenic factors. The prevalence of slope failures, debris flows, rockfalls, and other landslide types is influenced by local environmental conditions and classification practices; however, a comparison between Kegalle District, Sri Lanka, and Tokushima Prefecture, Japan, reveals broadly similar classification patterns despite regional differences. The adoption of type-specific inventories, such as the recently modified NBRO database in Sri Lanka, enhances hazard modelling by enabling more precise distinction between failure mechanisms and supporting the design of targeted countermeasures. Literature comparisons indicate that commonly prioritized landslide conditioning factors (LCFs)—including land use, rainfall, slope, elevation, and NDVI indices—are widely acknowledged as influential, yet the integration of less frequently applied but contextually significant variables, such as plane curvature, soil thickness, and distance from earthquake-induced epicenters, can strengthen predictive accuracy when adapted to local settings.

To determine the most suitable analytical approach for exploring the relationship between LCFs and specific landslide types, this study employed the Forest-Based and Boosted Classification and Regression (FBCR) tool in ArcGIS Pro. RF, widely used in susceptibility mapping, consistently demonstrates high predictive accuracy, with approximately one-third of recent studies reporting superior results compared to other machine learning methods. While RF offers robust ensemble-based predictions, the gradient boosting method further enhances model performance by sequentially correcting misclassifications, incorporating regularisation, and reducing overfitting.

The choice of study and reference areas—Kegalle District (KD) in Sri Lanka and Wakayama (WP) and Tokushima (TP) prefectures in Japan—was informed by the availability of type-specific inventories and broadly comparable elevation and rainfall ranges. Earthquake-induced landslides assessed separately given the differing seismic hazard contexts of Japan and Sri Lanka. Inventory analysis revealed no earthquake-triggered landslides in TP but 12 cliff-type events in WP between

2002 and 2022. These findings, supported by literature-based magnitude thresholds for landslide triggering, underscore the importance of region-specific factor selection and methodological adaptation in enhancing the relevance and reliability of landslide-type prediction models.

The relative importance of each LCF was first evaluated against trends reported in previous studies (Figure 10). Consistent with earlier work, DEM, slope, and land use emerged as highly influential, while factors such as soil thickness and building proximity—less frequently highlighted in the literature—showed strong importance in this study, reflecting local anthropogenic influences. Moderate-importance variables, including direct duration radiation, terrain ruggedness index (TRI), and profile curvature, contributed meaningfully to the model’s discriminatory power, highlighting the multifaceted influence of both geomorphic form and solar energy input on slope stability.

In modelling the relationship between LCFs and cliff-type landslide occurrence, 24 explanatory variables derived from satellite imagery, digital elevation models, thematic maps, and official geospatial datasets were processed to a consistent spatial resolution and projection. Many of these terrain-derived indices—such as slope, curvature metrics, TWI, SPI, TRI, and TPI—are well established in the literature for their geomorphic relevance. The gradient boosting algorithm was selected for model training due to its iterative bias-reduction process, with parameter tuning (e.g., number of trees, lambda, gamma) applied to optimize predictive performance while avoiding overfitting. Geology, soil type, and land use were excluded from the final model owing to their low importance scores, reflecting the benefits of data-driven feature selection in enhancing model generalizability.

The model results confirm that cliff-type landslides can be effectively predicted using terrain-derived explanatory variables in combination with advanced ensemble learning algorithms. The performance metrics—accuracy (0.84), sensitivity (0.84), F1 score (0.84), and Matthews correlation coefficient (MCC) (0.68)—indicate a high level of predictive reliability, consistent with or exceeding the performance of similar studies employing Random Forest and gradient boosting techniques for landslide susceptibility mapping. The relatively low standard deviation (0.03) across multiple runs further suggests model stability and generalizability. The prediction performance of 95% for both landslide and non-landslide classes in TP highlights balanced classification ability, reducing the risk of bias toward either class.

The variable importance analysis identified DEM, distance from roads, soil thickness, slope, and proximity to buildings as the most influential factors. While slope and elevation consistently appear among top predictors in global literature as shown in below Figure 51, the prominence of soil thickness and building proximity in this study reflects local geomorphic and anthropogenic influences—particularly in densely settled terrain where slope cutting and surface loading can exacerbate instability. Hydrological indices such as Direct Duration Radiation (DDR), Terrain Ruggedness Index (TRI), and Profile Curvature also emerged as moderately influential, each contributing ~7% to the model’s gain (Table 7), reinforcing the role of topographic form and energy in slope failure processes.

Table 7. Variable Importance of moderate LCF.

Variable	Importance (Gain) Percentage %	Importance (Weight) Percentage %
DDR	7	7
TRI	7	8
Profile Curvature	7	6
Annual average rainfall	6	5
Flow Accumulation	4	5
DR	4	3
TWI	4	3
Flow Direction	3	3
Distance to Streams	3	3
SPI	3	3

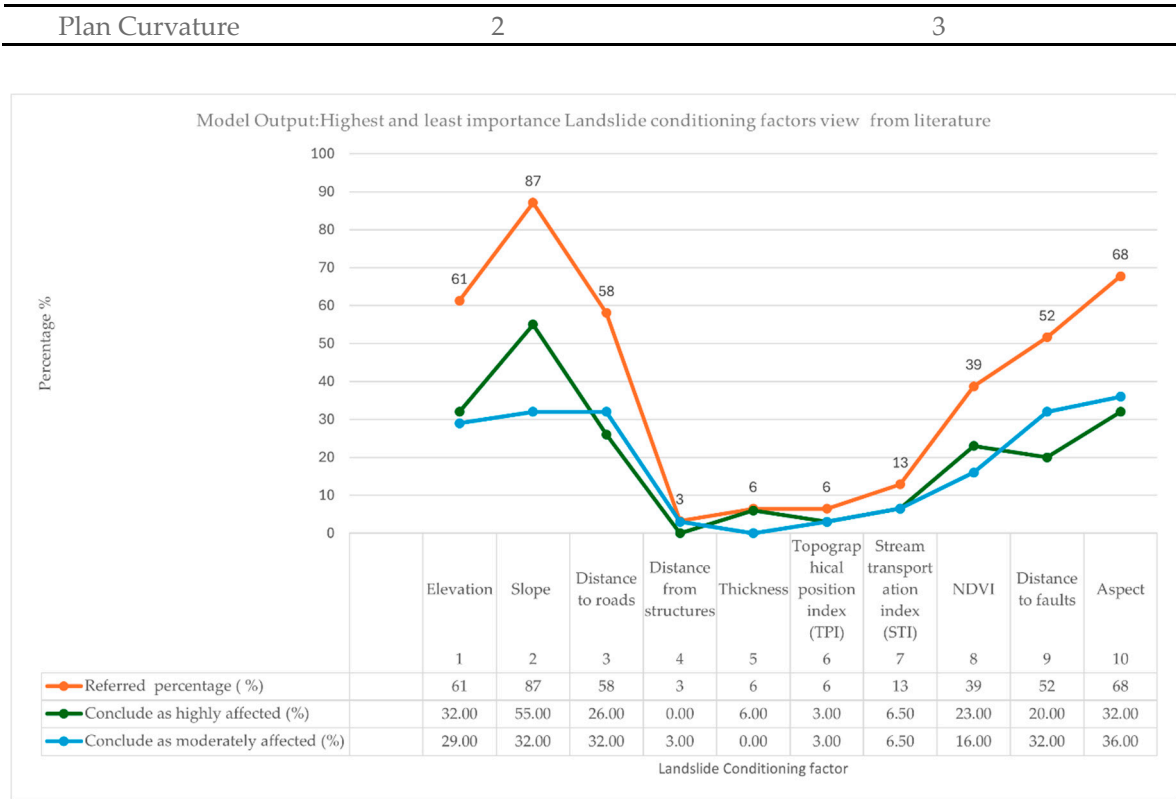


Figure 51. Model Output: Highest and least importance Landslide conditioning factors view from literature [7–16,24–35].

Validation outcomes demonstrate the added value of using type-specific inventories. In TP, 112 of 118 cliff-type landslide points (95%) overlapped with predicted high-susceptibility zones, while in KD the recently modified inventory yielded a match rate of 80.1% (72 out of 89 points). In contrast, validation against KD’s earlier non-classified inventory resulted in only 39.1% (115 out of 294 points) alignment, underscoring the critical importance of detailed landslide type information for accurate hazard modelling. This finding echoes observations that type-specific susceptibility mapping significantly improves the spatial precision of hazard delineation and the relevance of mitigation strategies.

Spatial overlay analysis between predicted cliff-type landslide zones and individual triggering factors further reinforced the model interpretation. The thematic maps (Figures 52–56 below) illustrate how high-susceptibility areas spatially correspond with Elevation, Slopes, Soil thickness zones, proximity to roads and buildings, rainfall and areas of high direct duration radiation.

The DEM comparison shows in Figure 52 that the majority of predicted cliff-type landslide zones coincide with areas in the mid-elevation range (15.138–250.114 m), highlighting the role of intermediate relief zones in fostering cliff instability. These elevations may represent transitional zones between valley floors and ridge tops, where slope gradients, drainage convergence, and weathering profiles combine to produce unstable conditions.

Soil thickness analysis in Figure 53 shows that many predicted landslide areas overlap with zones of maximum soil depth (182.048–200 cm). Thicker soils may promote instability by increasing gravitational load and water retention, thus raising pore-water pressures during heavy rainfall events.

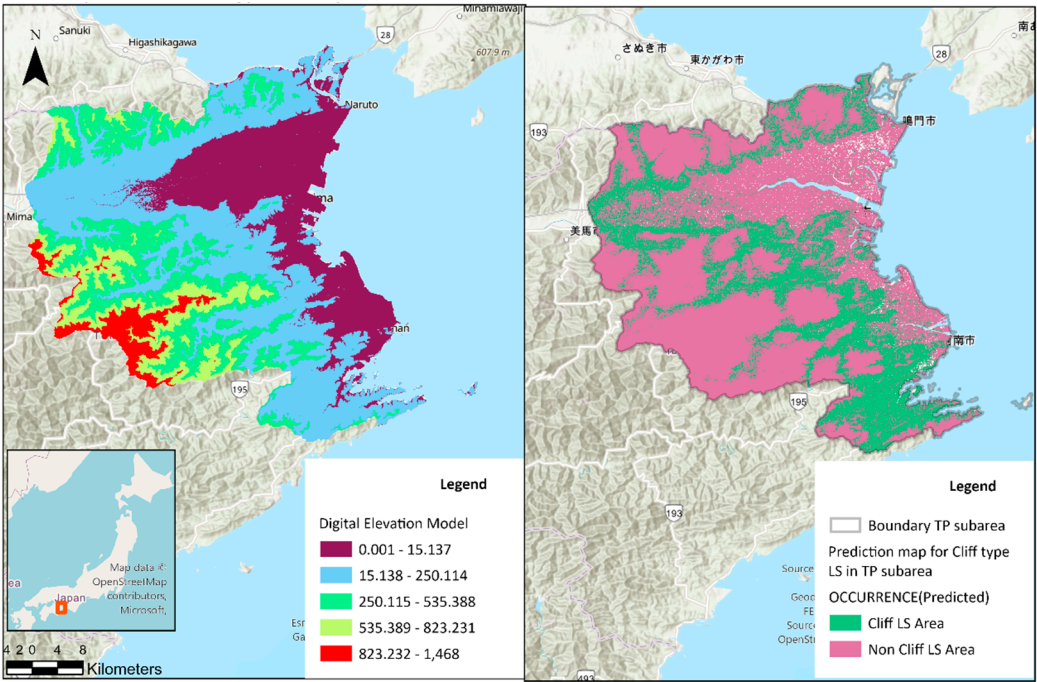


Figure 52. Comparison of DEM and Predicted Cliff LS area.

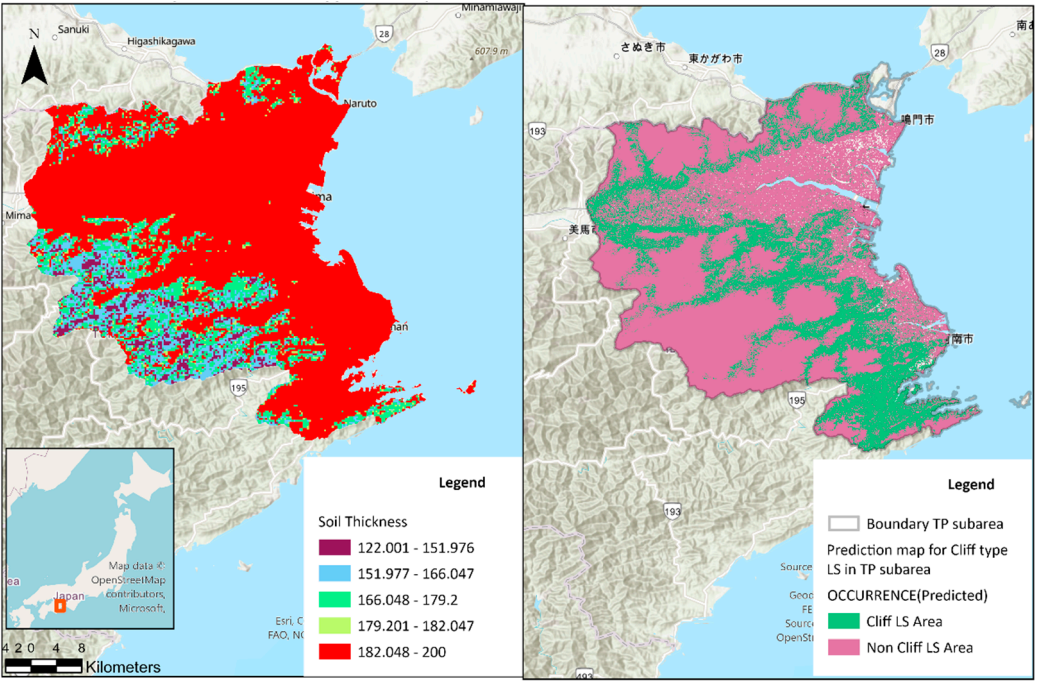


Figure 53. Comparison of Soil thickness and Predicted Cliff LS area.

The comparison between the slope classification (Figure 54) and the predicted cliff-type landslide distribution (right panel) reveals a clear spatial correspondence between the predicted cliff-type landslide zones and areas with moderate slope gradients (3.5–28.497°), represented in yellow. This suggests that cliff-type landslides in the Tokushima subarea are more frequently associated with moderately steep slopes rather than extremely steep terrain (>28.498°).

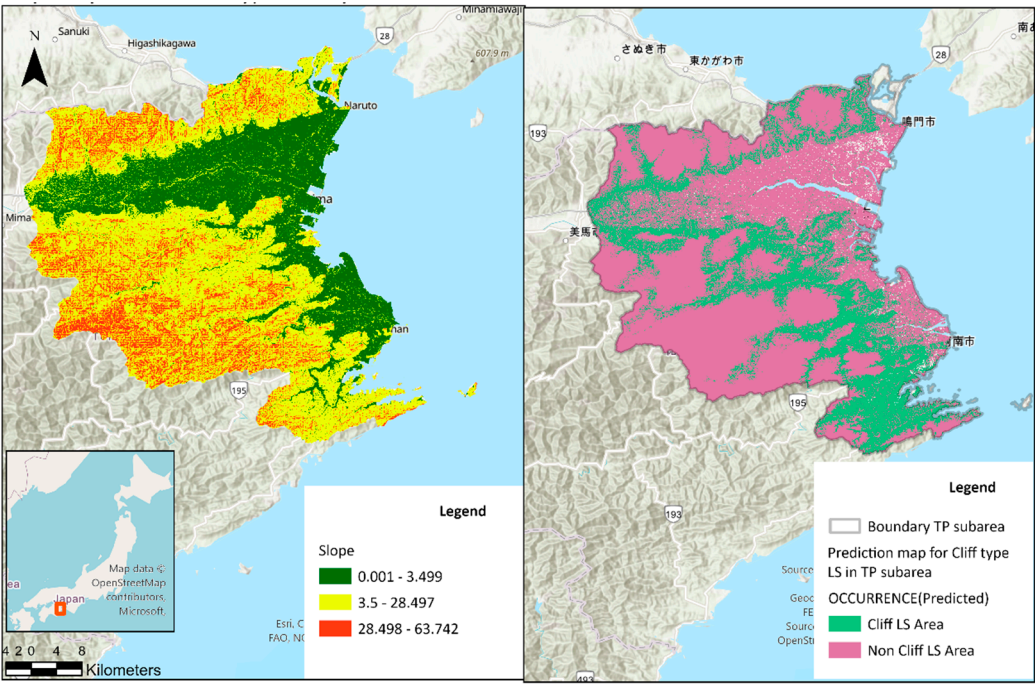


Figure 54. Comparison of slope and Predicted Cliff LS area.

Overlaying transportation networks (Figure 55) revealed that many predicted cliff-type landslide areas intersect or lie in close proximity to road corridors. This relationship is consistent with findings from previous studies that road construction, cut slopes, and associated drainage modifications can locally reduce slope stability and accelerate failure processes. Similarly, the distribution of buildings shows (Figure 56) a notable concentration within or near predicted landslide-prone zones, suggesting heightened exposure of infrastructure to slope hazards. This proximity underscores the potential socio-economic implications of such failures and the necessity for integrated land-use and hazard management planning.

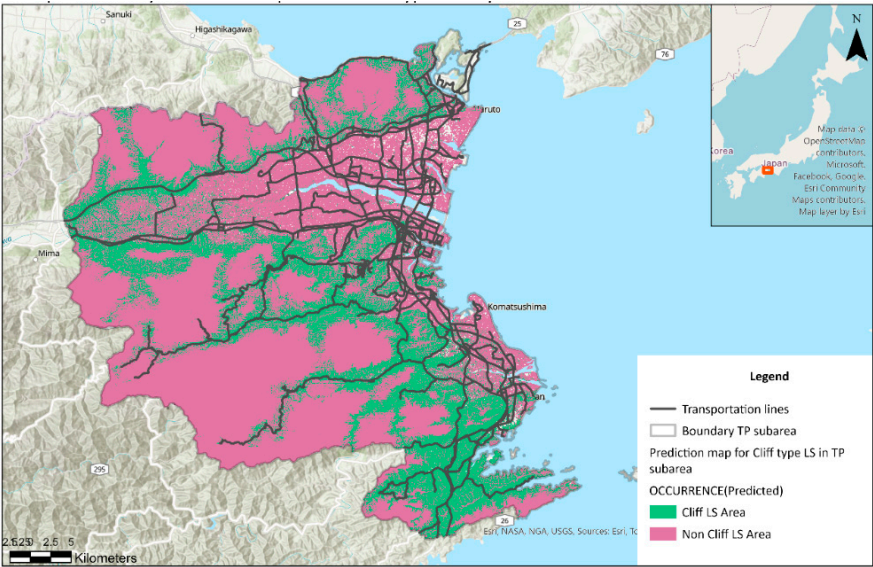


Figure 55. Transportation line layer overlaid with Predicted Cliff LS area.

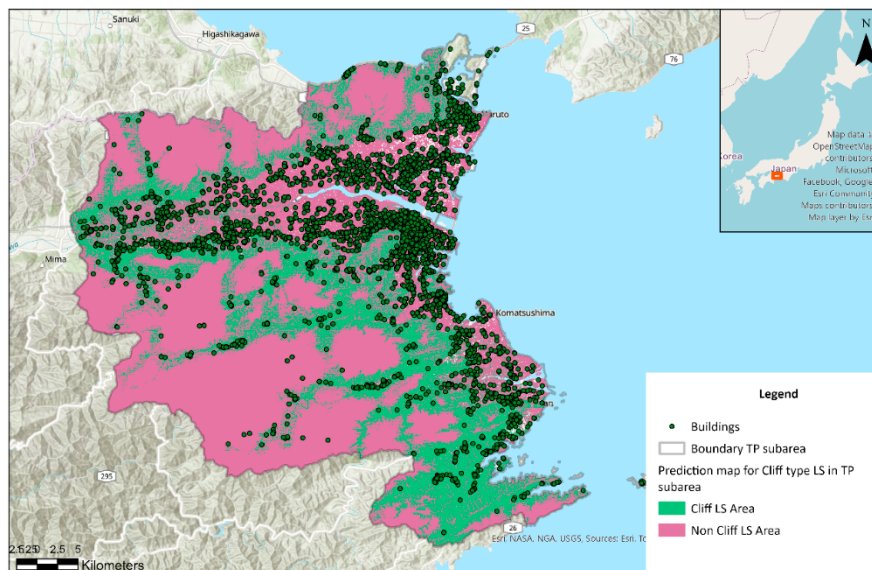


Figure 56. Building layer overlaid with Predicted Cliff LS area.

Collectively, these results demonstrate that a carefully selected set of LCFs, combined with advanced machine learning algorithms, can produce high-accuracy, transferable models for cliff-type landslide susceptibility. The approach enables more targeted hazard management, particularly when applied to inventories with detailed landslide type classifications, thereby bridging the gap between susceptibility modelling and engineering-scale countermeasure design.

5. Conclusion

This study demonstrates the effectiveness of advanced machine learning methods—specifically Random Forest and Gradient Boosting—in modelling cliff-type landslide susceptibility using a diverse set of landslide conditioning factors (LCFs). By integrating terrain-derived indices, hydrological parameters, proximity measures, and selected anthropogenic variables, the developed models achieved high predictive accuracy, stability, and balanced classification performance. The results confirm that inventories containing detailed landslide type classifications greatly improve model precision, with validation outcomes showing substantially higher agreement rates compared to non-classified datasets.

The variable importance analysis revealed that both widely recognised factors (e.g., slope, elevation) and context-specific variables (e.g., soil thickness, proximity to buildings) significantly contribute to cliff-type landslide prediction. Spatial overlay analysis highlighted the strong alignment between high-susceptibility zones and critical triggering conditions such as steep slopes, reduced soil cover, and areas influenced by human infrastructure. These findings underscore the importance of combining universally relevant predictors with locally significant factors to achieve optimal model performance.

The methodological approach developed in this research can be adapted to other landslide types and geographic contexts, provided that suitable inventory data and environmental covariates are available. Beyond academic contributions, the study offers practical value for hazard management, land-use planning, and the design of targeted countermeasures. Future work should focus on integrating temporal triggers such as rainfall intensity-duration thresholds and seismic parameters, as well as testing model transferability across regions with similar geomorphological characteristics.

Limitations:

While the model demonstrates strong performance metrics, several limitations should be acknowledged. First, the quality and completeness of the landslide and earthquake inventories used

for training may contain biases, and some events that could influence the model’s learning process might have gone undetected or misclassified. Additionally, the environmental variables utilized were restricted to the available spatial layers, potentially overlooking relevant but unmeasured factors, such as the Soil Water Index and groundwater levels.

Although the model is based on numerical and range-based factors, which reduces sensitivity to regional differences, there remains a significant risk that predictions may become unreliable when approaching the model’s extrapolation limits. In this study, the model was trained using data from Japan and applied to Sri Lanka; while the numerical nature of the conditioning factors enables interpolation and limited extrapolation, the possibility of reduced accuracy under extreme or unrepresented value ranges cannot be ruled out.

Lastly, although the analysis incorporated the Cliff landslide (LS) point layer and 21 landslide conditioning factors (LCFs), only the earthquake point layer was considered beyond the Tokushima Prefecture boundary, with a buffer applied outside the study area. This represents a limitation of the study, as the other 20 LCFs were not extended beyond Tokushima Prefecture. Expanding the spatial coverage of all conditioning factors, rather than restricting them to the prefectural boundary, could provide a more comprehensive assessment and improve the robustness of the model results.

Recommendations:

To enhance the comprehensiveness and practical utility of landslide susceptibility modeling, it is recommended to integrate temporal variables, such as rainfall intensity, earthquake occurrences, and historical landslide timing, into future models. Including these time-dependent factors would allow for predictions not only of spatial risk but also of the likely timing of landslide events. Additionally, future studies should focus on developing models specifically designed for areas affected by earthquakes, as seismic activity significantly contributes to triggering landslides, especially in steep and unstable terrain. Furthermore, it is advisable to expand the modeling approach to include various types of landslides, such as debris flows, rockfall.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Acknowledgments: The authors would like to thank all the people and organizations who contributed to make this research successfully, especially the Professors at Saitama University, Mr.K.Handa from JICA, Professor Maros Finka from Slovakia, Dr. Asiri and Mr. Sivanantharajah from Sri Lanka, Dr. Tina and Dr. Ziad from UK and Mr. Jack Horton from USA who gave valuable insight.

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

Abbreviations

The following abbreviations are used in this manuscript:

- DDR Direct Duration radiation
- DR Direct radiation
- FRCB Forest-based and Boosted Classification and Regression
- KD Kegalle District
- LS Landslide
- LCF Landslide Conditioning Factors
- MCC Matthew’s correlation coefficient
- NDVI Normalized difference vegetation index
- NBRO National Building Research Organization
- RA Reference Area
- STI Sediment transportation index
- SPI Stream power index
- SA Study Area
- TP Tokushima Prefecture
- TPI Topographical position index

TRI Topographical roughness index
 TWI Topographical wetness index
 WP Wakayama Prefecture

References

1. 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*, 80309. <https://doi.org/10.3390/geosciences10080309>.
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. *Geoscience Frontiers* **2021**, *12*, 857–876. <https://doi.org/10.1016/j.gsf.2020.09.004>.
3. Pennington, C.; Dijkstra, T.; Lark, M.; Dashwood, C.; Harrison, A.; Freeborough, K. Antecedent precipitation as a potential proxy for landslide incidence in South West UK. In Proceedings of World Landslide Forum, Vol. 3, 2014.
4. Haque, U.; Blum, P.; da Silva, P.F.; Andersen, P.; Pilz, J.; Chalov, S.R.; Malet, J.P.; Auflič, M.J.; Andres, N.; Poyiadji, E.; et al. Fatal landslides in Europe. *Landslides* **2016**, *13*, 1545–1554. <https://doi.org/10.1007/s10346-016-0689-3>.
5. Project News Letter Project for Capacity Strengthening on Development of Non-Structural Measures for Landslide Risk Reduction Workshop on Landslide Inventory Sheet (WG1). (2020). <https://www.facebook.com/Project.SABO/>
6. Junichi, K., & Naoki, I. (2020). Outline of measures for sediment disaster by the Sabo department of MLIT, Japan. In Landslides (Vol. 17, Issue 11, pp. 2503–2513). Springer Science and Business Media Deutschland GmbH. <https://doi.org/10.1007/s10346-020-01554-9>
7. 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. <https://doi.org/10.3390/rs15051376>.
8. Miao, F.; Zhao, F.; Wu, Y.; Li, L.; Török, Á. Landslide susceptibility mapping in Three Gorges Reservoir area based on GIS and boosting decision tree model. *Stoch. Environ. Res. Risk Assess.* **2023**. <https://doi.org/10.1007/s00477-023-02394-4>.
9. 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**. <https://doi.org/10.1002/gj.4683>.
10. Kohno, M.; Higuchi, Y. Landslide Susceptibility Assessment in the Japanese Archipelago Based on a Landslide Distribution Map. *ISPRS Int. J. Geo-Inf.* **2023**, *12*, 2. <https://doi.org/10.3390/ijgi12020037>.
11. Dhungana, G.; Ghimire, R.; Poudel, R.; Kumal, S. Landslide susceptibility and risk analysis in Benighat Rural Municipality, Dhading, Nepal. *Nat. Hazards Res.* **2023**, *3*, 170–185. <https://doi.org/10.1016/j.nhres.2023.03.006>.
12. Shah, N.A.; Shafique, M.; Ishfaq, M.; Faisal, K.; van der Meijde, M. Integrated approach for landslide risk assessment using geoinformation tools and field data in Hindukush Mountain Ranges, Northern Pakistan. *Sustainability* **2023**, *15*, 3102. <https://doi.org/10.3390/su15043102>.
13. Achour, Y.; Saidani, Z.; Touati, R.; Pham, Q.B.; Pal, S.C.; Mustafa, F.; Balik Sanli, F. Assessing landslide susceptibility using a machine learning-based approach to achieving land degradation neutrality. *Environmental Earth Sciences* **2021**, *80*, 17. <https://doi.org/10.1007/s12665-021-09889-9>.
14. Qin, Y.; Yang, G.; Lu, K.; Sun, Q.; Xie, J.; Wu, Y. Performance evaluation of five GIS-based models for landslide susceptibility prediction and mapping: A case study of Kaiyang County, China. *Sustainability* **2021**, *13*, 11441. <https://doi.org/10.3390/su13116441>.
15. Saha, S.; Roy, J.; Hembram, T.K.; Pradhan, B.; Dikshit, A.; Maulud, K.N.A.; Alamri, A.M. Comparison between deep learning and tree-based machine learning approaches for landslide susceptibility mapping. *Water* **2021**, *13*, 2664. <https://doi.org/10.3390/w13192664>.

16. Pham, Q.B.; Achour, Y.; Ali, S.A.; Parvin, F.; Vojtek, M.; Vojteková, J.; Al-Ansari, N.; Achu, A.L.; Costache, R.; Khedher, K.M.; Anh, D.T. A comparison among fuzzy multi-criteria decision making, bivariate, multivariate and machine learning models in landslide susceptibility mapping. *Geomatics, Nat. Hazards Risk* **2021**, *12*, 1741–1777. <https://doi.org/10.1080/19475705.2021.1944330>.
17. Bandara, R.M.S.; Jayasingha, P. Landslide disaster risk reduction strategies and present achievements in Sri Lanka. *Geosciences Research* **2018**, *3*, 3. <https://doi.org/10.22606/gr.2018.33001>.
18. Hemasinghe, H.; Rangali, R.S.S.; Deshapriya, N.L.; Samarakoon, L. Landslide susceptibility mapping using logistic regression model (a case study in Badulla District, Sri Lanka). *Procedia Eng.* **2018**, *212*, 1046–1053. <https://doi.org/10.1016/j.proeng.2018.01.135>.
19. Vakhshoori, V.; Pourghasemi, H.R.; Zare, M.; Blaschke, T. Landslide susceptibility mapping using GIS-based data mining algorithms. *Water* **2021**, *11*, 2292. <https://doi.org/10.3390/w11112292>.
20. Project for Capacity Strengthening on Development of Non-Structural Measures for Landslide Risk Reduction in Sri Lanka: National Building Research Organization. Final Report. 2022.
21. 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. <https://doi.org/10.1007/s11069-021-05186-7>.
22. Shinohara, Y.; Watanabe, Y. Differences in factors determining landslide hazards among three types of landslides in Japan. *Nat. Hazards* **2023**, *118*, 1689–1705. <https://doi.org/10.1007/s11069-023-06075-x>.
23. Kumarihamy, R.M.K.; Nianthi, K.W.G.R.; Shaw, R. Land Cover Changes and Landslide Risk in Sri Lanka 2022, pp. 413–433. https://doi.org/10.1007/978-981-16-7314-6_18.
24. Goetz, J.N.; Brenning, A.; Petschko, H.; Leopold, P. Evaluating machine learning and statistical prediction techniques for landslide susceptibility modeling. *Comput. Geosci.* **2015**, *81*, 1–11. <https://doi.org/10.1016/j.cageo.2015.04.007>.
25. Guo, Z.; Ferrer, J.V.; Hürlimann, M.; Medina, V.; Puig-Polo, C.; Yin, K.; Huang, D. Shallow landslide susceptibility assessment under future climate and land cover changes: A case study from southwest China. *Geoscience Frontiers* **2023**, *14*, 4. <https://doi.org/10.1016/j.gsf.2023.101542>.
26. He, Q.; Shahabi, H.; Shirzadi, A.; Li, S.; Chen, W.; Wang, N.; Chai, H.; Bian, H.; Ma, J.; Chen, Y.; 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. <https://doi.org/10.1016/j.scitotenv.2019.01.329>.
27. He, Q.; Xu, Z.; Li, S.; Li, R.; Zhang, S.; Wang, N.; Pham, B.T.; Chen, W. Novel entropy and rotation forest-based credal decision tree classifier for landslide susceptibility modeling. *Entropy* **2019**, *21*, 2. <https://doi.org/10.3390/e21020106>.
28. Hu, H.; Wang, C.; Liang, Z.; Gao, R.; Li, B. Exploring complementary models consisting of machine learning algorithms for landslide susceptibility mapping. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 639. <https://doi.org/10.3390/ijgi10100639>.
29. Kaihara, S.; Tadakuma, N.; Saito, H.; Nakaya, H. Influence of below-threshold rainfall on landslide occurrence based on Japanese cases. *Natural Hazards* **2023**, *115*, 2307–2332. <https://doi.org/10.1007/s11069-022-05639-7>.
30. Kalantar, B.; Ueda, N.; Saeidi, V.; Ahmadi, K.; Halin, A.A.; Shabani, F. Landslide susceptibility mapping: Machine and ensemble learning based on remote sensing big data. *Remote Sens.* **2020**, *12*, 11. <https://doi.org/10.3390/rs12111737>.
31. Liang, Z.; Wang, C.; Khan, K.U.J. Application and comparison of different ensemble learning machines combining with a novel sampling strategy for shallow landslide susceptibility mapping. *Stoch. Environ. Res. Risk Assess.* **2021**, *35*, 1243–1256. <https://doi.org/10.1007/s00477-020-01893-y>.
32. Mosaffaie, J.; Salehpour Jam, A.; Sarfaraz, F. Landslide risk assessment based on susceptibility and vulnerability. *Environ. Dev. Sustain.* **2023**. <https://doi.org/10.1007/s10668-023-03093-4>.
33. Nachappa, T.G.; Ghorbanzadeh, O.; Gholamnia, K.; Blaschke, T. Multi-hazard exposure mapping using machine learning for the state of Salzburg, Austria. *Remote Sens.* **2020**, *12*, 17. <https://doi.org/10.3390/RS12172757>.

34. Palliyaguru, S.T.; Liyanage, L.C.; Weerakoon, O.S.; Wimalaratne, G.D.S.P. Random forest as a novel machine learning approach to predict landslide susceptibility in Kalutara District, Sri Lanka. In Proceedings of the 20th Int. Conf. on Advances in ICT for Emerging Regions (ICTer), 2020, pp. 262–267. <https://doi.org/10.1109/ICTer51097.2020.9325460>.
35. Stanley, T.A.; Kirschbaum, D.B.; Benz, G.; Emberson, R.A.; Amatya, P.M.; Medwedeff, W.; Clark, M.K. Data-driven landslide nowcasting at the global scale. *Front. Earth Sci.* **2021**, *9*, 640043. <https://doi.org/10.3389/feart.2021.640043>.
36. 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. <https://doi.org/10.1016/j.catena.2016.11.032>.
37. Dang, V.H.; Hoang, N.D.; Nguyen, L.M.D.; Bui, D.T.; Samui, P. A novel GIS-based random forest machine algorithm for the spatial prediction of shallow landslide susceptibility. *Forests* **2020**, *11*, 1. <https://doi.org/10.3390/f11010118>.
38. Esri Academy. Available online: <https://www.esri.com/training/> (accessed on 10 July 2025).
39. Forest-based and Boosted Classification and Regression (Spatial statistics). Available online: <https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-statistics/forestbasedclassificationregression.htm> (accessed on 10 July 2025).
40. Sri Lankans shouldn't repeat the same mistake - JICA. Available online: <https://www.dailymirror.lk/opinion/sri-lankans-shouldn-t-repeat-the-same-mistake-jica/172-92179> (accessed on 10 July 2025).
41. USGS Earthquake Hazards Program. Available online: <https://earthquake.usgs.gov/earthquakes/search/> (accessed on 10 July 2025).
42. Keefer, D.K. Landslides Caused by Earthquakes. *Geol. Soc. Am. Bull.* **1984**, *95*, 406–421. [https://doi.org/10.1130/0016-7606\(1984\)95<406:LCBE>2.0.CO;2](https://doi.org/10.1130/0016-7606(1984)95<406:LCBE>2.0.CO;2)
43. Papadopoulos, G. A., & Plessa, A. Magnitude-distance relations for earthquake-induced landslides in Greece. In *Engineering Geology (Vol. 58)* **2000**. www.elsevier.nl/locate/enggeo
44. USGS EarthExplorer. Available online: <https://earthexplorer.usgs.gov/> (accessed on 10 July 2025).
45. ISRIC World Soil Information. Available online: <https://data.isric.org/geonetwork/srv/eng/catalog.search#/metadata/f36117ea-9be5-4afd-bb7d-7a3e77bf392a> (accessed on 10 July 2025).
46. Download site for digital land information. Available online: <https://nlftp.mlit.go.jp/ksj/index.html> (accessed on 10 July 2025).
47. AMeDAS. Available online: <https://www.jma.go.jp/jma/en/Activities/amedas/amedas.html> (accessed on 10 July 2025).
48. ISRIC World Soil Information. Available online: https://www.isric.org/explore/soilgrids/faq-soilgrids#What_happened_to_the_maps_of_soil_types (accessed on 10 July 2025).
49. 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. <https://doi.org/10.1130/G37530.1>.
50. Democratic Socialist Republic of Sri Lanka. Project for Capacity Strengthening on Development of Non-Structural Measures for Landslide Risk Reduction in Sri Lanka. 2022.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.