Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Machine Learning-Based Landslide Susceptibility Mapping and Environmental Controls in Hanang District, Northern Tanzania

Asnath Alberto Malekela

asnathalberto.malekela@zalf.de

Leibniz Centre for Agricultural Landscape Research (ZALF)

Research Article

Keywords: Landslides, Machine learning, Random Forest, Remote sensing, Geographic information systems (GIS)

Posted Date: July 23rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10348179/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Landslides are among the most destructive geohazards in mountainous regions, causing substantial socio-economic and environmental challenges. In Tanzania, the 2023 landslide disaster in Hanang district highlighted the need for reliable spatial tools to support landslide risk assessment and disaster preparedness. This study employed a Random Forest (RF) machine learning model, integrated with remote sensing and Geographic Information Systems (GIS), to map landslide susceptibility in the study area. A total of 859 inventory points, comprising 429 observed landslide locations and 430 non-landslide locations, were used to develop and validate the model. A comprehensive set of ten (10) environmental conditioning factors was derived from the satellite imagery and geospatial datasets, including annual rainfall, rainfall event intensity, elevation, slope, aspect, curvature, normalized difference vegetation index (NDVI), topographic wetness index (TWI), distance to rivers, and land use/land cover features. The RF model achieved an overall accuracy of 93.5% and a Kappa coefficient of 0.869, demonstrating outstanding predictive performance. Variable importance analysis identified NDVI as the dominant predictor of landslides. The susceptibility map indicated that approximately 43.92% of the district falls within the high and very high susceptibility classes, representing the area's most vulnerable to future landslides. Spatial validation further demonstrated strong model reliability, with 93.71% of the observed landslide inventory located within the high and very high susceptibility zones. These findings demonstrate that the RF model effectively predicts landslide susceptibility and provides valuable information for disaster risk reduction, land-use planning, and climate adaptation strategies.

1. Introduction

Landslides are among the most destructive natural hazards, with critical socioeconomic and environmental implications, often exacerbated by drivers of global change (Kumar, 2024; Alcántara-Ayala, 2025). Globally, rainfall-induced landslides account for a significant proportion of slope failures, particularly in mountainous and tectonically active regions where intense precipitation interacts with susceptible terrain conditions (Fidan et al., 2024; Riaz et al., 2024). The frequency and severity of landslide events have increased in many parts of the world due to the combined effects of climate variability, extreme rainfall events, land degradation, and expanding human activities in vulnerable landscapes (Amarasinghe et al., 2024; Zhao et al., 2026). This has led to a critical need for identifying areas susceptible to landslides for disaster risk reduction, land-use planning, and climate adaptation strategies. In Sub-Saharan Africa (SSA), landslides pose a growing environmental and socio-economic challenge across the continent (Mertens et al., 2016; Serwajja et al., 2024). Even though landslide hazards have historically received less attention than droughts and floods in Africa, recent catastrophic events have highlighted the growing vulnerability of mountainous regions in SSA (Nahayo et al., 2023; Kinyeru et al., 2026).

In East Africa, several countries have experienced devastating rainfall-induced landslides, resulting in fatalities, displacement, and significant economic losses (Sieber et al., 2025; Nwazelibe et al., 2026). Despite the increasing occurrence of such events, landslide susceptibility assessments remain limited in

many African countries, constraining evidence-based planning and hazard mitigation efforts. Tanzania has experienced numerous landslides, particularly in the highland and mountainous regions with steep slopes and heavy rainfall. In December 2023, extremely heavy rainfall triggered widespread landslides and mudflows, causing extensive destruction of settlements, infrastructure, agricultural land, and human lives in the Hanang district (Sieber et al., 2025; Authority, 2022). This was among the most severe rainfall-induced slope failures recorded in the country. The disaster highlighted the urgent need for spatially explicit tools to identify high-risk areas and support proactive disaster risk management.

Landslide susceptibility mapping is essential for estimating the likelihood of landslide occurrence by relating observed landslides to environmental conditioning factors. Over the years, traditional susceptibility mapping approaches have been constrained by subjective weighting and assumptions about interactions among influencing factors (Pathak et al., 2023; Kaynak, 2026). Recent advances in geospatial technologies, remote sensing, and machine learning have transformed landslide susceptibility assessment (Mohan et al., 2021; Tehrani et al., 2022). Machine learning (ML) algorithms have demonstrated superior predictive performance compared with many conventional approaches (Zhang & Wang, 2024; Nwazelibe et al., 2026). Among machine learning models, Random Forest (RF) has emerged as one of the most widely used algorithms for landslide susceptibility mapping due to its robustness, high predictive accuracy, and ability to assess variable importance (Lu et al., 2024; Kanwar et al., 2025). The algorithm constructs multiple decision trees using random subsets of observations and predictor variables, thereby improving classification performance while reducing variance. RF can effectively accommodate both continuous and categorical variables and can quantify the relative contribution of individual conditioning factors to landslide occurrence (Dai et al., 2023; Kaur et al., 2024; Nguyen & Doan, 2026).

Previous landslide susceptibility studies in Tanzania have primarily relied on conventional GIS-based multicriteria approaches to delineate hazard-prone areas (Zahor & Yamungu, 2022; Makonyo & Zahor, 2023; Muhimbula et al., 2025; Mulaya et al., 2025). Although these studies have provided valuable insights, the application of advanced machine learning algorithms for landslide susceptibility assessment remains limited. Existing studies have generally been constrained by relatively small landslide inventories, limited evaluation of environmental conditioning factors, and limited use of cloud-based geospatial computing platforms to efficiently process multi-source geospatial datasets. This study advances previous research by integrating a comprehensive landslide inventory comprising 859 locations (429 landslide and 430 non-landslide points) with ten (10) environmental conditioning factors within a Random Forest–Google Earth Engine (GEE) framework. Beyond generating a landslide susceptibility map, the study quantifies the relative importance of individual conditioning factors, thereby providing new insights into the interactions between vegetation condition, topography, hydrological characteristics, and land use in controlling landslide occurrence.

2. Methods and Materials

2.1 The study area

The study was conducted in Hanang district, located in northern Tanzania (Fig. 1). The district is characterized by highly variable terrain ranging from relatively flat lowlands to steep mountainous landscapes. Elevation varies considerably across the district, leading to pronounced differences in slope gradients, drainage patterns, and geomorphological conditions (Mulaya et al., 2025). The district experiences a semi-arid to sub-humid climate, with rainfall exhibiting strong seasonal and spatial variability. The combination of rugged topography, intense rainfall events, and active surface runoff makes the district particularly susceptible to rainfall-induced slope failures. Vegetation cover varies considerably across the district due to differences in elevation, land use, and climatic conditions (Muhimbula et al., 2025). The district was selected for this study because it represents one of Tanzania's most landslide-prone areas and experienced one of the country's most destructive rainfall-induced landslide disasters in late 2023. The extreme rainfall event triggered widespread slope failures, resulting in substantial loss of life, community displacement, infrastructure destruction, and extensive damage to agricultural land (Mulaya et al., 2025; Sieber et al., 2025). The disaster highlighted the urgent need for reliable spatial tools to identify areas vulnerable to future landslides and to support disaster preparedness, land-use planning, and climate adaptation strategies.

2.2 Landslide Inventory Development

A reliable landslide inventory is essential for model development and validation as it provides the reference information needed to distinguish landslide-affected areas from stable terrain (Cehn et al., 2025; Dai et al., 2025). In this study, the landslide inventory was developed by interpreting high-resolution satellite imagery and vegetation disturbance patterns derived from the Normalized Difference Vegetation Index (NDVI). Landslide locations were identified based on characteristic geomorphological signatures, including exposed soil, scarps, disrupted vegetation, and debris deposits, while non-landslide locations were selected from stable areas exhibiting no evidence of slope failure. A comprehensive landslide inventory comprising 859 locations (429 landslide and 430 non-landslide points) was developed and used for both model training and validation (Fig. 2).

Landslide occurrence was treated as a binary dependent variable in ML classification, indicating whether a landslide occurred. Landslide presence was assigned a value of one (1), while non-landslide locations were assigned a value of zero (0), as shown in Eq. 1:

$$Y = \begin{cases} 1, & if\ landslide\ occured\ at\ location\ i \\ 0, & if\ no\ landslide\ occured\ at\ location\ i \end{cases}$$

1

2.3 Selection of Landslide Conditioning Factors

Landslides are controlled by a complex interaction of topographic, hydrological, climatic, and environmental factors (Selamat et al., 2025). In this study, ten (10) conditioning factors were selected,

including elevation, slope, aspect, curvature, event rainfall, annual rainfall, NDVI, land use and land cover, distance to rivers, and Topographic Wetness Index (TWI).

2.3.1 Topographic wetness index (TWI)

Hydrological processes play a crucial role in landslide occurrence because soil saturation increases pore-water pressure and reduces slope stability. TWI is widely used to quantify the tendency of water accumulation across landscapes and has been identified as an important predictor of landslide occurrence in mountainous regions (Moresi et al., 2020; Leonelli et al., 2026). Areas with higher TWI values generally exhibit greater moisture retention and increased susceptibility to slope failure. TWI was used to characterize spatial patterns of moisture accumulation and potential soil saturation. The index was calculated as:

$TWI = In\left(\frac{\alpha}{tan\ \beta}\right)$ (2) Whereby, α is the upslope contributing area per unit contour length, and β represents the local slope angle (radians). Higher TWI values indicate wetter conditions and a greater potential for landslides.

2.3.2 Vegetation condition using NDVI

Vegetation cover contributes significantly to slope stabilization through root reinforcement, interception of rainfall, and regulation of soil moisture. On the contrary, degraded vegetation may increase susceptibility to erosion and slope failure (Chen et al., 2023). To quantify vegetation condition, the NDVI was calculated from Sentinel-2 imagery. NDVI is one of the most widely used indicators of vegetation health and density, and is expressed as:

$$NDVI = \frac{NIR - Red}{NIR + Red}$$

3

NIR denotes near-infrared reflectance, and Red represents red-band reflectance. NDVI values range from − 1 to + 1, with higher values indicating denser and healthier vegetation cover.

2.4 Random Forest Landslide Susceptibility Model

To systematically assess landslide susceptibility in Hanang district, this study adopted an RF-based modeling framework that integrates remote sensing, GIS, and machine learning techniques. As illustrated in Fig. 3, the framework combines a comprehensive landslide inventory comprising 859 validated locations (429 landslide and 430 non-landslide points) with ten environmental conditioning factors representing topographic, hydrological, climatic, vegetation, and land-use characteristics. Predictor variables were extracted and used to train an RF classifier with a 70% training and 30% testing split, using 100 decision trees to improve predictive stability. The trained model generated continuous landslide occurrence probabilities, which were subsequently classified into five susceptibility classes

ranging from very low to very high. Model performance was evaluated using overall accuracy, the Kappa coefficient, producer's accuracy, user's accuracy, and spatial validation based on the distribution of observed landslides across susceptibility classes. Also, the relative importance of each environmental conditioning factor was quantified to identify the dominant drivers of rainfall-induced landslide occurrence.

Following the workflow presented in Fig. 3, landslide susceptibility was modeled using the RF algorithm. RF is a supervised ensemble machine learning technique that constructs multiple decision trees from bootstrap samples of the training dataset, with randomly selected subsets of predictor variables at each split. In recent years, machine learning techniques have become increasingly important in landslide susceptibility assessment because they can model complex interactions among environmental variables (Zhang & Wang, 2024; Nwazelibe et al., 2026). RF has consistently demonstrated high predictive performance among the ML models, making it one of the most widely applied methods for landslide susceptibility modeling (Lu et al., 2024; Kanwar et al., 2025). In this study, the model was trained using the selected environmental conditioning factors to estimate the probability of landslide occurrence, which was subsequently classified into five susceptibility classes. The RF prediction function is expressed as:

$$\hat{Y}(x) = \text{mode}\{T_1(x), T_2(x), ..., T_B(x)\} \quad (4)$$

where $\hat{Y}(x)$ denotes the final predicted class for observation x, $T_1(x)$, $T_2(x)$, ..., $T_B(x)$ represent the predictions generated by individual decision trees, and B is the total number of trees within the RF ensemble. The function mode { } returns the class that receives the highest number of votes across all decision trees. In this study, an RF classifier comprising 100 decision trees was implemented on GEE to model landslide susceptibility.

2.5 Model Training and Validation

To evaluate the predictive performance of the RF model, the landslide inventory dataset was randomly divided into training and testing subsets. In this study, 70% of the observations were used for model training, and the remaining 30% for validation. The RF model estimates the probability of landslide occurrence at a given location based on a set of environmental conditioning factors. The conditional probability of landslide occurrence is expressed as:

$$P(Y_i = 1 \mid X_i) \quad (5)$$

where $P(Y_i = 1 \mid X_i)$ denotes the probability of landslide occurrence at location i given the environmental conditioning factors; Y_i is the binary response variable indicating landslide occurrence (1 = landslide, 0 = non-landslide); and X_i denotes the vector of predictor variables. The resulting probability surface was subsequently reclassified into five susceptibility categories: very low, low, moderate, high, and very high susceptibility.

2.6 Variable Importance Analysis

Variable importance analysis was conducted to identify the relative contribution of environmental conditioning factors to landslide occurrence. The RF model can evaluate the importance of each predictor by measuring the decrease in prediction accuracy when that predictor is randomly shuffled while all other variables remain unchanged (Taalab et al., 2018; Schlögl et al., 2025). The greater the reduction in prediction accuracy, the more important the predictor is in determining landslide susceptibility. The variable importance score was calculated as:

VI = (1/B) Σ=1 (Eperm,j − Eorig) (6)

where VI is the importance score of predictor variable j, Eperm,j is the classification error after transformation of variable j in the b tree, Eorig is the original classification error before transformation, and B is the total number of trees in the RF ensemble. Higher importance scores indicate a greater contribution of a predictor variable to landslide susceptibility.

The predictive performance of the RF model was evaluated using a confusion matrix, which compares observed landslide occurrences with model predictions. Accuracy assessment is an essential step in landslide susceptibility modeling because it quantifies the reliability of model outputs and determines the extent to which predicted susceptibility patterns correspond to actual landslide occurrences (Abdelkader & Csámer, 2025; Bhattacharya et al., 2025). Four performance metrics were employed in this study, namely overall accuracy (OA), producer's accuracy (PA), user's accuracy (UA), and the Kappa coefficient (κ).

Overall accuracy (OA) measures the proportion of observations that were correctly classified by the model and is calculated as:

OA = (TP + TN) / (TP + TN + FP + FN) (7)

where TP represents true positives, TN denotes true negatives, FP refers to false positives, and FN represents false negatives.

Producer's accuracy (PA) evaluates the ability of the model to correctly identify actual landslide locations and is expressed as:

PA = TP / (TP + FN) (8)

Higher PA values indicate lower omission errors and better landslide detection.

User's Accuracy (UA) measures the reliability of predicted landslide classes and is calculated as:

UA = TP / (TP + FP) (9)

Higher UA values indicate lower commission errors and greater confidence in the predicted landslide locations.

Although OA indicates how well the model classifies landslide and non-landslide locations, it does not account for correct classifications that may occur by chance. Therefore, the Kappa coefficient (κ) was used to provide a more reliable measure of model performance. The Kappa statistic compares the model's observed accuracy with the expected accuracy by random chance (Cohen, 1960; Osisanwo et al., 2017). Kappa coefficient is calculated as:

$$\kappa = (P_o - P_e) / (1 - P_e) \quad (10)$$

where P_o represents the observed accuracy between predicted and actual classes, and P_e denotes the expected accuracy by chance. Kappa values range from 0 to 1, with values approaching 1 indicating excellent classification performance and values near 0 indicating poor agreement beyond chance.

3. Results and Discussion

This section presents and interprets the results of the RF-based landslide susceptibility assessment for Hanang district. The findings are evaluated to determine the spatial distribution of landslide susceptibility, identify the dominant environmental factors controlling rainfall-induced landslides, and assess the model's predictive performance and spatial reliability. The results are discussed in relation to the geomorphological and environmental characteristics of the study area and compared with previous studies to highlight their scientific significance.

3.1 Spatial Distribution of Landslide Susceptibility

Landslide susceptibility mapping provides a spatial distribution of areas susceptible to landslides. The susceptibility classes in the study area revealed a considerable spatial variability in landslide risk across the district. The combined high- and very-high-susceptibility classes covered approximately 1,613.91 km², representing 43.92% of the total district area. Contrarywise, the moderate susceptibility class covered the smallest area, accounting for 11.85% of the total area (Table 1). The results suggest that a substantial portion of the study area is vulnerable to landslides, particularly in areas with steep slopes, high rainfall, sparse vegetation cover, and proximity to drainage networks (Fig. 4). The extensive coverage of highly susceptible terrain highlights the district's significant exposure to rainfall-induced slope failures and underscores the importance of spatially targeted risk-reduction interventions. These findings are consistent with previous studies demonstrating that topography, hydrological conditions, vegetation cover, and drainage characteristics are the primary controls of landslide occurrence (Greco er al., 2023; Liang, 2026; Pratama et al., 2026).

The susceptibility map reveals a distinct spatial clustering of high- and very high-susceptibility zones in the central, eastern, southeastern, and northwestern parts of the study area. These results indicate that landslide occurrence is strongly controlled by the spatial heterogeneity of environmental conditions rather than by a random distribution (Sun et al., 2023; Liu et al., 2025). The highly susceptible areas coincide with the steep mountainous terrain surrounding Mount Hanang, the escarpments of the East African rift valley, and deeply incised drainage networks, where steep slopes, high local relief, and

concentrated surface runoff collectively reduce slope stability. In contrast, the very low and low susceptibility classes are predominantly associated with valley bottoms, lowland plains, and gently undulating landscapes characterized by lower slope gradients, reduced erosional intensity, and comparatively stable geomorphological conditions. Similar spatial patterns have been reported in mountainous environments where landslide susceptibility is primarily governed by the interaction of topography, drainage characteristics, and vegetation cover (Abeysiriwardana & Gomes, 2022; Li & Duan, 2024). The concentration of highly susceptible zones along mountainous and dissected landscapes therefore confirms the dominant influence of terrain configuration on rainfall-induced slope instability.

The spatial patterns observed in this study also reflect the influence of the extreme rainfall event that affected the district in late 2023, triggering widespread landslides in the study area. Intense rainfall likely increased soil moisture, promoted infiltration, elevated pore-water pressure, and reduced soil shear strength, thereby accelerating slope failure in areas that were already geomorphologically unstable. However, the results demonstrate that extreme rainfall alone does not explain the observed landslide distribution. Instead, rainfall served as a triggering mechanism whose effects were intensified by pre-existing environmental conditions. The findings support previous studies showing that rainfall-induced landslides arise from the interaction between climatic triggers and local environmental controls rather than from rainfall intensity alone (Smith et al., 2023; Jiang et al., 2025). The identified high-susceptibility zones represent priority areas for landslide risk reduction and climate adaptation. The susceptibility map provides a scientific basis for guiding land-use zoning, infrastructure siting, watershed management, and community preparedness. As climate change is projected to increase the frequency and intensity of extreme rainfall events in many tropical mountainous regions, integrating susceptibility information into development planning and disaster management strategies is essential for reducing future landslide impacts.

Table 1
Area distribution of landslide susceptibility classes in Hanang District.

Susceptibility Class	Area (km²)	Percentage (%)
Very low	925.38	25.18
Low	699.81	19.05
Moderate	435.26	11.85
High	554.24	15.08
Very high	1059.67	28.84
Total	**3674.36**	**100.00**

3.2 Environmental Drivers of Landslide Susceptibility

The study used various environmental conditioning factors that influence slope instability. The contribution of environmental conditioning factors to landslide occurrence was evaluated using the RF model's variable importance function. The model was used to predict the spatial distribution of landslide susceptibility across the study area using ten (10) variables grouped into topographic, hydrological, climatic, and land-surface variables. The variable importance analysis revealed substantial differences in the contribution of environmental predictors to landslide occurrence. Among the ten conditioning factors, NDVI emerged as the most influential predictor, with an importance score of 137.39 indicating that vegetation condition plays a dominant role in controlling slope stability (Figs. 5 & 6). The strong influence of NDVI indicates that vegetation cover provides an important stabilizing effect through root reinforcement, reduction of surface runoff, and regulation of soil moisture. Areas characterized by sparse vegetation or disturbed land cover are more likely to experience increased erosion and reduced slope resistance, thereby increasing susceptibility to landslides. Similar findings have been reported in various studies where vegetation-related indicators were identified as critical determinants of slope instability (Chen et al., 2023; Zhang et al., 2023; Nguyen et al., 2025).

Slope was identified as the second most influential predictor, with an importance score of 62.22. Steeper slopes generate greater gravitational stresses and are therefore more susceptible to failure when weakened by hydrological factors such as intense rainfall. Numerous landslide susceptibility studies have similarly identified slope as one of the most important conditioning factors to landslide occurrence (Campforts et al., 2022; Qiu & Wei, 2025). Elevation ranked third with an importance score of 60.26. The influence of elevation reflects the study area's topographic complexity, particularly around Mount Hanang, where steep mountainous terrain creates favorable conditions for slope failure. Elevation may also indirectly capture variations in rainfall, drainage patterns, and geomorphological processes that influence landslide occurrence. Curvature exhibited moderate importance scores of 46.36, suggesting that terrain morphology and slope orientation influence the distribution of landslide-prone areas. Similar observations have been reported in susceptibility studies, which show that topographic and vegetation factors make a substantial contribution to landslide events (Nohan et al., 2019; Asmare, 2023; Jiang et al., 2023).

Hydrological and climatic variables also played a critical role in controlling landslide susceptibility (Figs. 5 & 7). The Topographic Wetness Index (TWI) recorded a high importance score (58.93), indicating that soil moisture accumulation is a key factor governing slope instability in the study area. High TWI values indicate areas where runoff converges and water accumulates, thereby increasing soil saturation and ultimately reducing soil shear strength. Similarly, distance to rivers achieved a high importance score (58.24), suggesting that fluvial processes, including channel incision and riverbank undercutting, substantially contribute to slope destabilization. Similar observations have been reported in previous studies demonstrating that topographic wetness and proximity to drainage strongly influence rainfall-induced landslides by regulating subsurface water movement and slope hydrology (Guo et al., 2025; Liang, 2026).

The climatic factors, specifically event rainfall, had the highest importance score (57.90), indicating that the extreme rainfall event that affected the study area in late 2023 was the principal trigger of the observed landslides. However, the comparable importance of annual rainfall (55.13) indicates that long-term rainfall regimes also contribute significantly by maintaining antecedent soil moisture conditions that increase landscape susceptibility to extreme rainfall events. The finding suggests that landslide occurrence in Hanang results from the interaction between long-term climatic conditions and short-duration extreme rainfall rather than from individual rainfall events alone. Similar findings have been reported by Fusco et al. (2022) and Perkins et al. (2025), who showed that antecedent moisture conditions largely determine the magnitude of the slope response once thresholds for intense rainfall are exceeded.

Aspect recorded a moderate importance score (45.89), indicating that slope orientation exerts an indirect influence on landslide occurrence through its effects on solar radiation, evapotranspiration, and vegetation growth, rather than serving as a primary controlling factor. Furthermore, land use and land cover exhibited the lowest importance score (27.71) among all conditioning factors. Although land cover is widely recognized as an important determinant of landslide susceptibility, its comparatively lower contribution in the present study suggests that terrain morphology, hydrological processes, vegetation condition, and rainfall dynamics exert stronger controls on rainfall-induced slope failures in Hanang district. Similar observations have been reported by other researchers who found that topographic and hydrological variables generally outperform generalized land-cover classes because they directly represent the physical mechanisms responsible for slope instability (Reichenbach et al., 2018; Arabameri et al., 2020; Hong et al., 2020; Merghadi et al., 2020). Generally, the variable importance analysis demonstrates that rainfall-induced landslides in the study area are governed by the interaction of vegetation, topographic, hydrological, and climatic factors rather than by a single environmental variable. The dominance of NDVI, together with the high importance of slope, elevation, TWI, distance to rivers, and rainfall-related variables, suggests that landslide occurrence results from coupled environmental processes. Vegetation condition influences slope protection, terrain morphology controls gravitational instability, hydrological factors regulate water accumulation, and extreme rainfall provides the triggering mechanism for slope failure. This integrated understanding extends previous landslide susceptibility studies by demonstrating how machine learning can disentangle the relative contributions of multiple interacting environmental factors, thereby improving the interpretation of rainfall-induced landslide processes.

3.3 Model Performance and Validation

The RF model demonstrated outstanding predictive performance in distinguishing landslide-prone areas from stable areas within the study district. The accuracy assessment results presented in Table 2 indicate that the model achieved an overall accuracy (OA) of 93.5% and a Kappa coefficient of 0.869. Also, producer's accuracy (PA) values of 95.0% and 91.0%, and the user's accuracy (UA) values of 93.1% and 94.0% demonstrate that the model effectively minimized omission and commission errors. The high OA and Kappa coefficient indicate that the RF model was highly effective in capturing the complex

relationships between landslide occurrences and environmental conditioning factors. According to Landis and Koch (1977), Kappa values exceeding 0.80 indicate almost perfect agreement between observed and predicted classes. Various studies on machine learning techniques for landslide susceptibility mapping have reported that the RF model outperforms many conventional statistical approaches due to its ability to model nonlinear relationships and interactions among predictor variables (Arabameri et al., 2020; Merghadi et al., 2020). The accuracy achieved in this study highlights the effectiveness of combining a comprehensive landslide inventory with cloud-based geospatial analysis.

The model performance depended on the spatial representativeness of the landslide inventory. The dataset comprised 429 landslide points and 430 non-landslide points distributed throughout the study area. From these points, 602 (70%) were used as the training sample and 257 (30%) as the test sample. Previous studies have emphasized that inventory quality is one of the most important determinants of the performance of susceptibility models (Lv et al., 2024; Chen & Fan, 2025). The results obtained in this study support this assertion and demonstrate that a well-distributed inventory can substantially enhance predictive capability.

Table 2
Accuracy assessment statistics of the Random Forest model.

Accuracy Metric	Value
Overall Accuracy	0.935
Kappa Coefficient	0.869
Producer's Accuracy	0.950, 0.910
User's Accuracy	0.931, 0.940

3.4 Spatial Validation of the Landslide Susceptibility Map

To further evaluate the spatial reliability of the RF model, the observed landslide inventory was overlaid on the predicted landslide susceptibility map, and the distribution of landslide locations across the five susceptibility classes was quantified (Fig. 8). The results demonstrate a strong spatial correspondence between the predicted susceptibility zones and the observed landslide locations. The results show that of the 429 identified landslide locations, 344 (80.19%) occurred within the very high susceptibility class, while 58 (13.52%) occurred within the high susceptibility class. Accordingly, 93.71% of all observed landslides were correctly captured within the high- and very-high-susceptibility zones, providing strong evidence that the RF model accurately identified areas most prone to slope failure. On the other side, a small percentage occurred within the moderate, low, and very low susceptibility classes, indicating that most of the observed landslides were correctly concentrated in areas predicted to be highly susceptible to landslides. The concentration of observed landslides within the high and very high susceptibility classes confirms that the model effectively captured the spatial interactions among environmental conditioning factors controlling slope instability. Areas classified as highly susceptible are characterized

by steep slopes, substantial elevation differences, reduced vegetation cover, and high rainfall intensity, all of which increase the likelihood of landslides.

The spatial validation complements the statistical evaluation of model performance. While the confusion matrix yielded an OA of 93.5% and a Kappa coefficient of 0.869, the overlay analysis demonstrates that these high accuracy metrics are also reflected spatially, with most observed landslides occurring in areas predicted to be highly susceptible. The agreement between statistical and spatial validation strengthens confidence in the RF model's predictive capability and demonstrates its suitability for regional-scale landslide susceptibility assessment. Similar findings have been reported in previous studies, which show that RF models reliably identify landslide-prone terrain by capturing complex nonlinear relationships among topographic, hydrological, climatic, and vegetation-related predictors (Hong et al., 2020; Nguyen et al., 2025; Tian et al., 2026). Studies suggest that successful susceptibility models should demonstrate strong spatial agreement between historical landslide locations and predicted high-risk areas (Reichenbach et al., 2018; Jones et al., 2021; Dai et al., 2025).

3.5 Implications for Landslide Risk Management

The findings of this study have important implications for disaster risk reduction and sustainable land management. The prominence of NDVI highlights the importance of vegetation conservation and restoration in maintaining slope stability. Similarly, the strong influence of rainfall and hydrological variables underscores the need for improved rainfall monitoring and early warning systems that can identify periods of elevated landslide hazard. The extensive area classified as highly susceptible suggests that susceptibility information should be incorporated into district-level planning processes, infrastructure development strategies, settlement expansion decisions, and watershed management programs. Generally, the study demonstrates the potential of integrating machine learning algorithms with cloud-based geospatial platforms for landslide susceptibility assessment in data-constrained environments.

4. Conclusion

The study employed an RF machine learning model integrated with remote sensing and GIS to assess landslide susceptibility in Hanang district, northern Tanzania. The results demonstrate that the RF algorithm effectively captured the complex interactions among topographic, hydrological, vegetation, and climatic variables governing landslide occurrence. The generated susceptibility map revealed distinct spatial heterogeneity in landslide hazard, with the high and very high susceptibility classes occupying a large share of the district and accounting for most historical landslide occurrences. Spatial validation further confirmed the model's reliability by demonstrating strong agreement between the predicted susceptibility classes and the observed landslide distribution. Together with the model's high predictive performance, the findings confirm the reliability of the RF approach for landslide susceptibility assessment. The analysis established that vegetation condition (NDVI) was the most influential conditioning factor controlling landslide susceptibility in the study area, followed by slope, elevation,

topographic wetness index, distance to rivers, and rainfall-related variables. The relatively lower contribution of land use/land cover suggests that topographic and climatic controls currently exert a stronger influence on slope instability than generalized land-cover classes, although local land management practices remain important contributors to landslide initiation. Generally, this study demonstrates that machine learning, when combined with remote sensing and geospatial analysis, provides a robust and scientifically reliable approach for landslide susceptibility mapping. The resulting susceptibility map constitutes an important evidence base for enhancing disaster preparedness, strengthening climate resilience, and supporting risk-informed development planning in landslide-prone regions. Future research should focus on integrating multi-temporal remote sensing and near-real-time rainfall data to develop dynamic landslide susceptibility and early-warning systems. Also, applying explainable artificial intelligence techniques could further support more effective disaster risk reduction and climate adaptation strategies.

Declarations

Acknowledgements

I extend my gratitude to the Alexander von Humboldt Foundation for granting me the Georg Forster Research Fellowship for Postdocs in Germany, which made it possible for me to work on this paper. I also thank the Mwalimu Nyerere Memorial Academy for granting me permission to take the Postdoc fellowship in Germany.

Author contributions

AAM was involved in the study's conception and design, methodology, data analysis, and the drafting of this paper. The author read and approved the final manuscript.

Funding details.

This work was supported by the Alexander von Humboldt Foundation under a Georg Forster Research Fellowship for Postdocs.

Data availability

No datasets were generated or analysed during the current study.

Code availability

Not applicable.

AI disclosure

This manuscript benefited from the use of ChatGPT, developed by OpenAI, which was employed to assist with grammar correction, language refinement, and clarity improvements. All conceptual work and data

analysis, interpretation, and final content decisions were made by the author.

Ethics approval

Not applicable.

Consent to participate

Not applicable.

Consent to publish

Not applicable.

Competing interests

The author declares no competing interests.

References

1. Abdelkader, M. M., & Csámer, Á. (2025). Comparative assessment of machine learning models for landslide susceptibility mapping: a focus on validation and accuracy. Natural Hazards, 121(9), 10299–10321.
2. Abeysiriwardana, H. D., & Gomes, P. I. (2022). Integrating vegetation indices and geo-environmental factors in GIS-based landslide-susceptibility mapping: using logistic regression. Journal of Mountain Science, 19(2), 477–492.
3. Alcántara-Ayala, I. (2025). Landslides in a changing world. Landslides, 22(9), 2851–2865.
4. Amarasinghe, M. P., Kulathilaka, S. A. S., Robert, D. J., Zhou, A., & Jayathissa, H. A. G. (2024). Risk assessment and management of rainfall-induced landslides in tropical regions: A review. Natural Hazards, 120(3), 2179–2231.
5. Arabameri, A., Saha, S., Roy, J., Chen, W., Blaschke, T., & Tien Bui, D. (2020). Landslide susceptibility evaluation and management using different machine learning methods in the Gallicash River Watershed, Iran. Remote Sensing, 12(3), 475.
6. Asmare, D. (2023). Application and validation of AHP and FR methods for landslide susceptibility mapping around choke mountain, northwestern ethiopia. Scientific African, 19, e01470.
7. Authority, T. M. (2022). Statement on the status of Tanzania climate in 2022. https://www.meteo.go.tz/uploads/publications/sw1628770614-TMA% 20BOOK, 202020, 20-2021.
8. Bhattacharya, S., Ali, T., Chakravortti, S., Pal, T., Majee, B. K., Mondal, A., ... & Chakrabortty, R. (2025). Application of machine learning and deep learning algorithms for landslide susceptibility assessment in landslide prone Himalayan Region. Earth Systems and Environment, 9(2), 1427–1445.

9. Campforts, B., Shobe, C. M., Overeem, I., & Tucker, G. E. (2022). The art of landslides: How stochastic mass wasting shapes topography and influences landscape dynamics. Journal of Geophysical Research: Earth Surface, 127(8), e2022JF006745.
10. Chen, C., & Fan, L. (2025). Selection of contributing factors for predicting landslide susceptibility using machine learning and deep learning models. Stochastic environmental research and risk assessment, 39(10), 4169–4194.
11. Chen, B., Shui, W., Liu, Y., & Deng, R. (2023). Analysis of slope stability with different vegetation types under the influence of rainfall. Forests, 14(9), 1865.
12. Chen, B., Maurer, J., & Gong, W. (2025). Applications of UAV in landslide research: a review: B. Chen et al. Landslides, 22(9), 3029–3048.
13. Cohen, J. (1960). A coefficient of agreement for nominal scales. Educational and psychological measurement, 20(1), 37–46.
14. Dai, X., Chen, J., Zhang, T., & Xue, C. (2025). Integrated landslide risk assessment via a landslide susceptibility model based on intelligent optimization algorithms. Remote Sensing, 17(3), 545.
15. Dai, X., Zhu, Y., Sun, K., Zou, Q., Zhao, S., Li, W., ... & Wang, S. (2023). Examining the spatially varying relationships between landslide susceptibility and conditioning factors using a geographical random forest approach: A case study in Liangshan, China. Remote Sensing, 15(6), 1513.
16. Fidan, S., Tanyaş, H., Akbaş, A., Lombardo, L., Petley, D. N., & Görüm, T. (2024). Understanding fatal landslides at global scales: a summary of topographic, climatic, and anthropogenic perspectives. Natural Hazards, 120(7), 6437–6455.
17. Fusco, F., Bordoni, M., Tufano, R., Vivaldi, V., Meisina, C., Valentino, R., ... & De Vita, P. (2022). Hydrological regimes in different slope environments and implications on rainfall thresholds triggering shallow landslides. Natural Hazards, 114(1), 907–939.
18. Greco, R., Marino, P., & Bogaard, T. A. (2023). Recent advancements of landslide hydrology. Wiley Interdisciplinary Reviews: Water, 10(6), e1675.
19. Guo, J., Wang, Y., & Li, Y. (2025). Topographic controls on the initiation and transport of landslide-triggered debris flows. Geomorphology, 486, 109901.
20. Hong, H., Liu, J., & Zhu, A. X. (2020). Modeling landslide susceptibility using LogitBoost alternating decision trees and forest by penalizing attributes with the bagging ensemble. Science of the total environment, 718, 137231.
21. Jiang, H., Zou, Q., Zhou, B., Jiang, Y., Cui, J., Yao, H., & Zhou, W. (2023). Estimation of shallow landslide susceptibility incorporating the impacts of vegetation on slope stability. International Journal of Disaster Risk Science, 14(4), 618–635.
22. Jiang, J., Hu, Y., Zheng, D., & Lyu, L. (2025). Disaster risk assessment of collapses and landslides in a hilly coastal city: the role of rainfall triggers and the disaster-inducing environment. Natural Hazards, 121(18), 21683–21704.
23. Jones, J. N., Boulton, S. J., Bennett, G. L., Stokes, M., & Whitworth, M. R. (2021). Temporal variations in landslide distributions following extreme events: Implications for landslide susceptibility

modeling. Journal of Geophysical Research: Earth Surface, 126(7), e2021JF006067.

24. Kanwar, M., Pokharel, B., & Lim, S. (2025). A new random forest method for landslide susceptibility mapping using hyperparameter optimization and grid search techniques. International Journal of Environmental Science and Technology, 22(11), 10635–10650.
25. Kaur, R., Gupta, V., & Chaudhary, B. S. (2024). Landslide susceptibility mapping and sensitivity analysis using various machine learning models: a case study of Beas valley, Indian Himalaya. Bulletin of Engineering Geology and the Environment, 83(6), 228.
26. Kaynak, T. (2026). A systematic framework for the integration of feature selection and artificial intelligence in landslide susceptibility assessment. Natural Hazards, 122(2), 91.
27. Kinyeru, E. T. N., Harris, M. N., & Jansz, J. (2026). Why do landslides occur in Kenya, and what can be done to mitigate their occurrences?. Natural Hazards, 122(4), 155.
28. Kumar, P. (2024). Social and economic impact in the landslide prone zones and related policies. In Landslides in the Himalayan region: risk assessment and mitigation strategy for sustainable management (pp. 499–529). Singapore: Springer Nature Singapore.
29. Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. biometrics, 159–174.
30. Leonelli, G., Francese, R. G., & Chelli, A. (2026). Active landslide portions contribute to surface water concentration: Insights from GIS analysis and field data in the Northern Apennines. Earth Systems and Environment, 10(1), 453–472.
31. Li, Y., & Duan, W. (2024). Decoding vegetation's role in landslide susceptibility mapping: An integrated review of techniques and future directions. Biogeotechnics, 2(1), 100056.
32. Liang, W. L. (2026). Topographic and hydrological controls on a shallow landslide: Prefailure evidence from a forested headwater catchment. Earth Surface Processes and Landforms, 51(7), e70344.
33. Liu, Y., Han, X., Cui, X., Li, X., & Jin, H. (2025). Spatial heterogeneity of landslide distribution and its drivers in the Yangtze River Basin: a remote sensing and GIS-based multi-factor analysis. Frontiers in Environmental Science, 13, 1576700.
34. Lu, Z., Liu, G., Song, Z., Sun, K., Li, M., Chen, Y., ... & Zhang, W. (2024). Advancements in technologies and methodologies of machine learning in landslide susceptibility research: current trends and future directions. Applied Sciences, 14(21), 9639.
35. Lv, Z., Wang, S., Yan, S., Han, J., & Zhang, G. (2024). Landslide susceptibility assessment based on multisource remote sensing considering inventory quality and modeling. Sustainability, 16(19), 8466.
36. Makonyo, M., & Zahor, Z. (2023). GIS-based analysis of landslides susceptibility mapping: a case study of Lushoto district, north-eastern Tanzania. Natural Hazards, 118(2), 1085–1115.
37. Merghadi, A., Yunus, A. P., Dou, J., Whiteley, J., ThaiPham, B., Bui, D. T., ... & Abderrahmane, B. (2020). Machine learning methods for landslide susceptibility studies: A comparative overview of algorithm performance. Earth-Science Reviews, 207, 103225.

38. Mertens, K., Jacobs, L., Maes, J., Kabaseke, C., Maertens, M., Poesen, J., ... & Vranken, L. (2016). The direct impact of landslides on household income in tropical regions: A case study from the Rwenzori Mountains in Uganda. Science of the Total Environment, 550, 1032–1043.
39. Mohan, A., Singh, A. K., Kumar, B., & Dwivedi, R. (2021). Review on remote sensing methods for landslide detection using machine and deep learning. Transactions on Emerging Telecommunications Technologies, 32(7), e3998.
40. Moresi, F. V., Maesano, M., Collalti, A., Sidle, R. C., Matteucci, G., & Scarascia Mugnozza, G. (2020). Mapping landslide prediction through a GIS-based model: A case study in a catchment in southern Italy. Geosciences, 10(8), 309.
41. Muhimbula, J., Sumari, N. S., & Balz, T. (2025). Landslide Susceptibility Assessment Using AHP, Frequency Ratio, and LSI Models: Understanding Topographical Controls in Hanang District, Tanzania. GeoHazards, 6(3), 58.
42. Mulaya, E., Gama, R., Kimani, C., Joseph, P., Kazimoto, E., & Mshiu, E. (2025). The triggering factors on the 3rd December 2023 catastrophic landslide in the Hanang area, northern Tanzania: The triggering factors of catastrophic landslide: Mulaya et al. Landslides, 22(3), 727.
43. Nahayo, L., Peng, C., Lei, Y., & Tan, R. (2023). Spatial understanding of historical and future landslide variation in Africa. Natural Hazards, 119(1), 613–641.
44. Nguyen, K. A., Jiang, Y. J., & Chen, W. (2025). Identifying Slope Hazard Zones in Central Taiwan Using Emerging Hot Spot Analysis and NDVI. Sustainability, 17(16), 7428.
45. Nguyen, B. Q. V., & Doan, V. L. (2026). The role of conditioning factors in machine learning-based landslide spatial probability. Environmental Geotechnics, 1–20.
46. Nohani, E., Moharrami, M., Sharafı, S., Khosravi, K., Pradhan, B., Pham, B. T., ... & M. Melesse, A. (2019). Landslide susceptibility mapping using different GIS-based bivariate models. Water, 11(7), 1402.
47. Nwazelibe, V. E., Egbueri, J. C., Unigwe, C. O., Agbasi, J. C., Ayejoto, D. A., & Abba, S. I. (2023). GIS-based landslide susceptibility mapping of Western Rwanda: an integrated artificial neural network, frequency ratio, and Shannon entropy approach. Environmental Earth Sciences, 82(19), 439.
48. Nwazelibe, V. E., Abdulkarim, M., Agbasi, J. C., Okpala, E. C., Pande, C. B., Abba, S. I., & Egbueri, J. C. (2026). Deep learning models for comprehensive landslide susceptibility study considering climate change and sustainable development implications in Western Province of Rwanda. Natural Hazards, 122(10), 432.
49. Osisanwo, F. Y., Akinsola, J. E., Awodele, O., Hinmikaiye, J. O., Olakanmi, O., & Akinjobi, J. (2017). Supervised machine learning algorithms: classification and comparison. International Journal of Computer Trends and Technology (IJCTT), 48(3), 128–138.
50. Pathak, Y. P., Prakash, I., & Dholakia, M. B. (2023). Conventional and modern approaches in landslide susceptibility mapping: A methodological review. Int J Appl Eng Technol, 5, 2633–4828.
51. Perkins, J., Oakley, N. S., Collins, B. D., Corbett, S. C., & Burgess, W. P. (2025). Characterizing the scale of regional landslide triggering from storm hydrometeorology. Natural Hazards and Earth

System Sciences, 25(3), 1037–1056.

52. Pratama, G. M., Gomi, T., Noviandi, R., Ritonga, R. P., Fathani, T. F., & Wilopo, W. (2026). Land Cover and Land Use Controls on Landslide Morphometry and Occurrence in a Heterogeneous Mountain Watershed. GeoHazards, 7(1), 31.
53. Qiu, H., & Wei, Y. (2025). Landslide geomorphology: pattern, process and stability. Journal of Earth Science, 36(1), 327–332.
54. Reichenbach, P., Rossi, M., Malamud, B. D., Mihir, M., & Guzzetti, F. (2018). A review of statistically-based landslide susceptibility models. Earth-science reviews, 180, 60–91.
55. Riaz, M. T., Basharat, M., Ahmed, K. S., Sirfraz, Y., Shahzad, A., & Shah, N. A. (2024). Failure mechanism of a massive fault–controlled rainfall–triggered landslide in northern Pakistan. Landslides, 21(11), 2741–2767.
56. Schlögl, M., Spiekermann, R., & Steger, S. (2025). Towards a holistic assessment of landslide susceptibility models: insights from the Central Eastern Alps. Environmental Earth Sciences, 84(4), 113.
57. Selamat, S. N., Majid, N. A., & Taha, M. R. (2025). Multicollinearity and spatial correlation analysis of landslide conditioning factors in Langat River Basin, Selangor. Natural Hazards, 121(3), 2665–2684.
58. Serwajja, E., Kisira, Y., & Bamutaze, Y. (2024). 'Better to die of landslides than hunger': Socio-economic and cultural intricacies of resettlement due to climate-induced hazards in Uganda. International Journal of Disaster Risk Reduction, 101, 104242.
59. Sieber, S., Malekela, A. A., & Chobo, J. S. (2025). The Influence of Major Climatic Elements on the Devastated Geo-Hydrological Disasters: A Case of Hanang Disaster of 3rd December 2023 in Tanzania. Atmos. Clim. Sci, 15, 106–125.
60. Smith, H. G., Neverman, A. J., Betts, H., & Spiekermann, R. (2023). The influence of spatial patterns in rainfall on shallow landslides. Geomorphology, 437, 108795.
61. Sun, H., Li, W., Scaioni, M., Fu, J., Guo, X., & Gao, J. (2023). Influence of spatial heterogeneity on landslide susceptibility in the transboundary area of the Himalayas. Geomorphology, 433, 108723.
62. Taalab, K., Cheng, T., & Zhang, Y. (2018). Mapping landslide susceptibility and types using Random Forest. Big Earth Data, 2(2), 159–178.
63. Tehrani, F. S., Calvello, M., Liu, Z., Zhang, L., & Lacasse, S. (2022). Machine learning and landslide studies: recent advances and applications. Natural Hazards, 114(2), 1197–1245.
64. Tian, C., Liu, S., Lan, H., & Li, L. (2026). Interpretable Multi-Temporal Landslide Susceptibility Assessment Using Random Forest and Tree-SHAP in the Eastern Himalayan Syntaxis. Remote Sensing, 18(11), 1842.
65. Zahor, Z., & Yamungu, N. E. (2022). Geographical Information Systems (GIS) and Remote Sensing (RS) Analysis for Landslides Susceptibility Mapping. University of Dar es Salaam Library Journal, 17(2), 72–93.

66. Zhang, J., Ma, X., Zhang, J., Sun, D., Zhou, X., Mi, C., & Wen, H. (2023). Insights into geospatial heterogeneity of landslide susceptibility based on the SHAP-XGBoost model. Journal of environmental management, 332, 117357.
67. Zhang, Q., & Wang, T. (2024). Deep learning for exploring landslides with remote sensing and geo-environmental data: Frameworks, progress, challenges, and opportunities. Remote Sensing, 16(8), 1344.
68. Zhao, Kailiang, Haijun Qiu, Ya Liu, Guoqing Yang, Zijing Liu, Wanqi Luo, Yaru Zhu, Guoyu Wang, and Mohib Ullah. "Projected response of global rainfall-induced landslides to future climate change." Journal of Hydrology (2026): 135698.

Figures

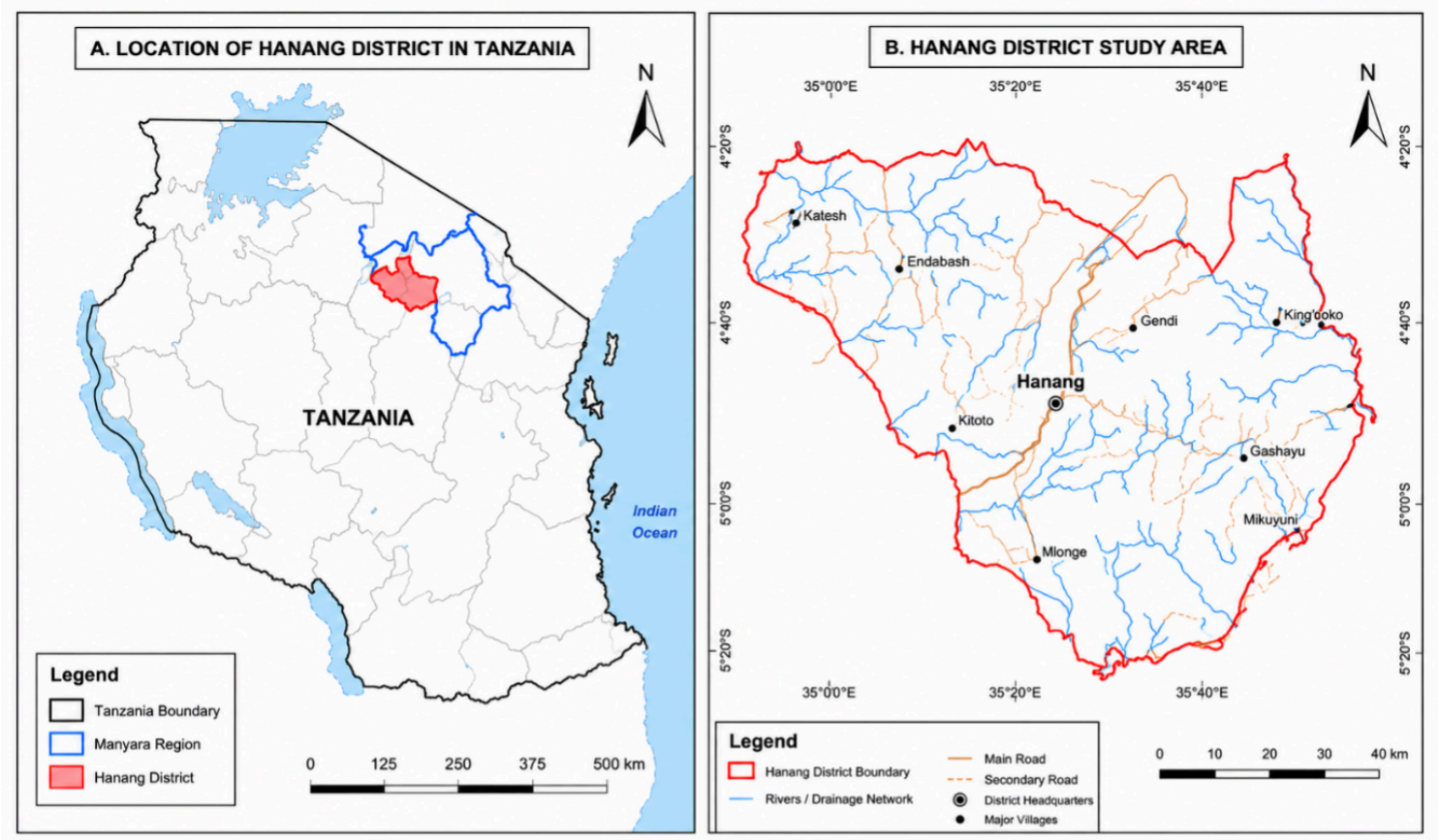


Figure 1

Location of Hanang district in northern Tanzania, showing the national location map and study area boundary with drainage network

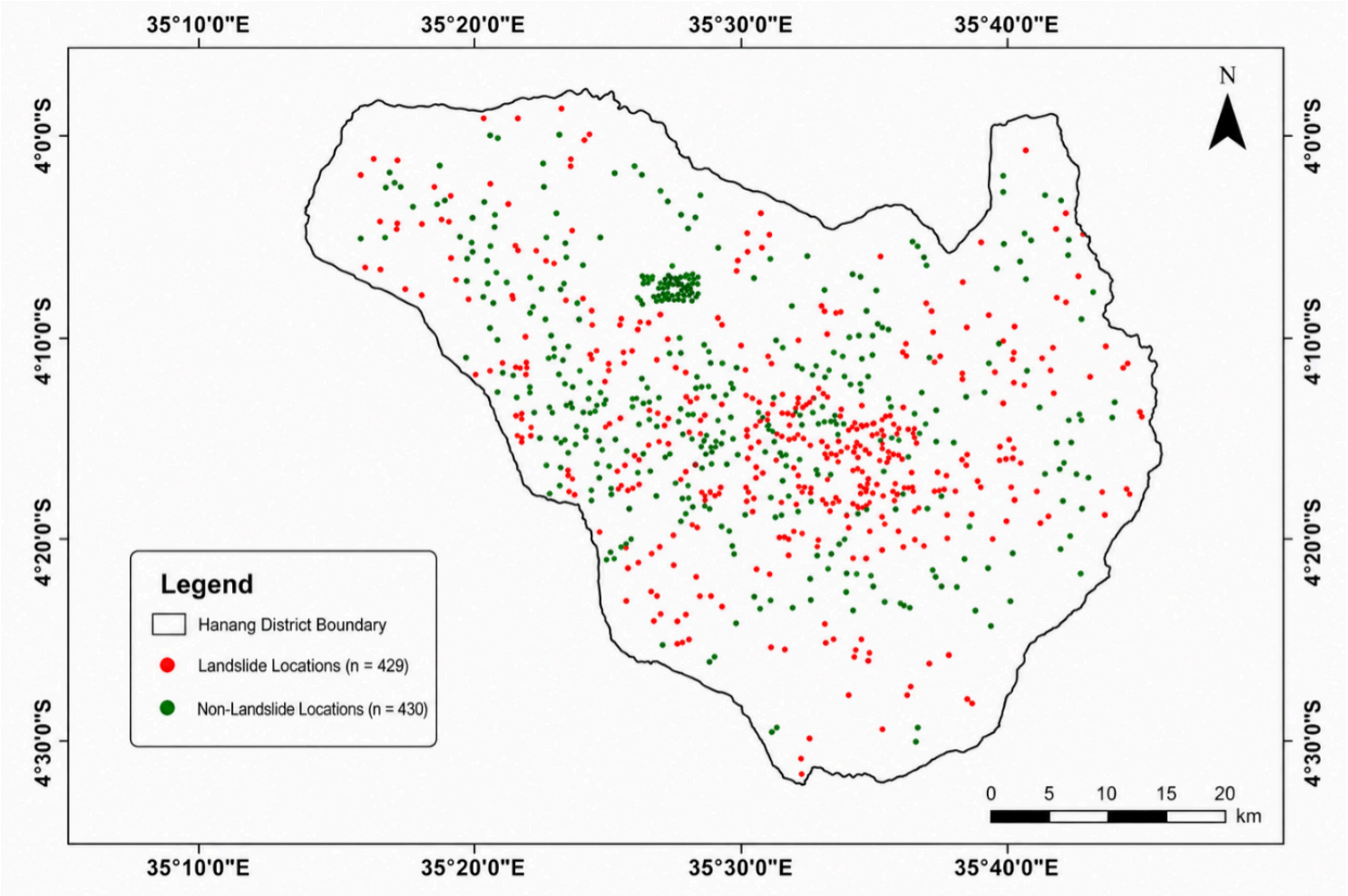


Figure 2

Inventory of landslide and non-landslide locations used for RF landslide susceptibility modeling in Hanang district. The inventory comprised 429 landslide locations and 430 non-landslide locations across the study area.

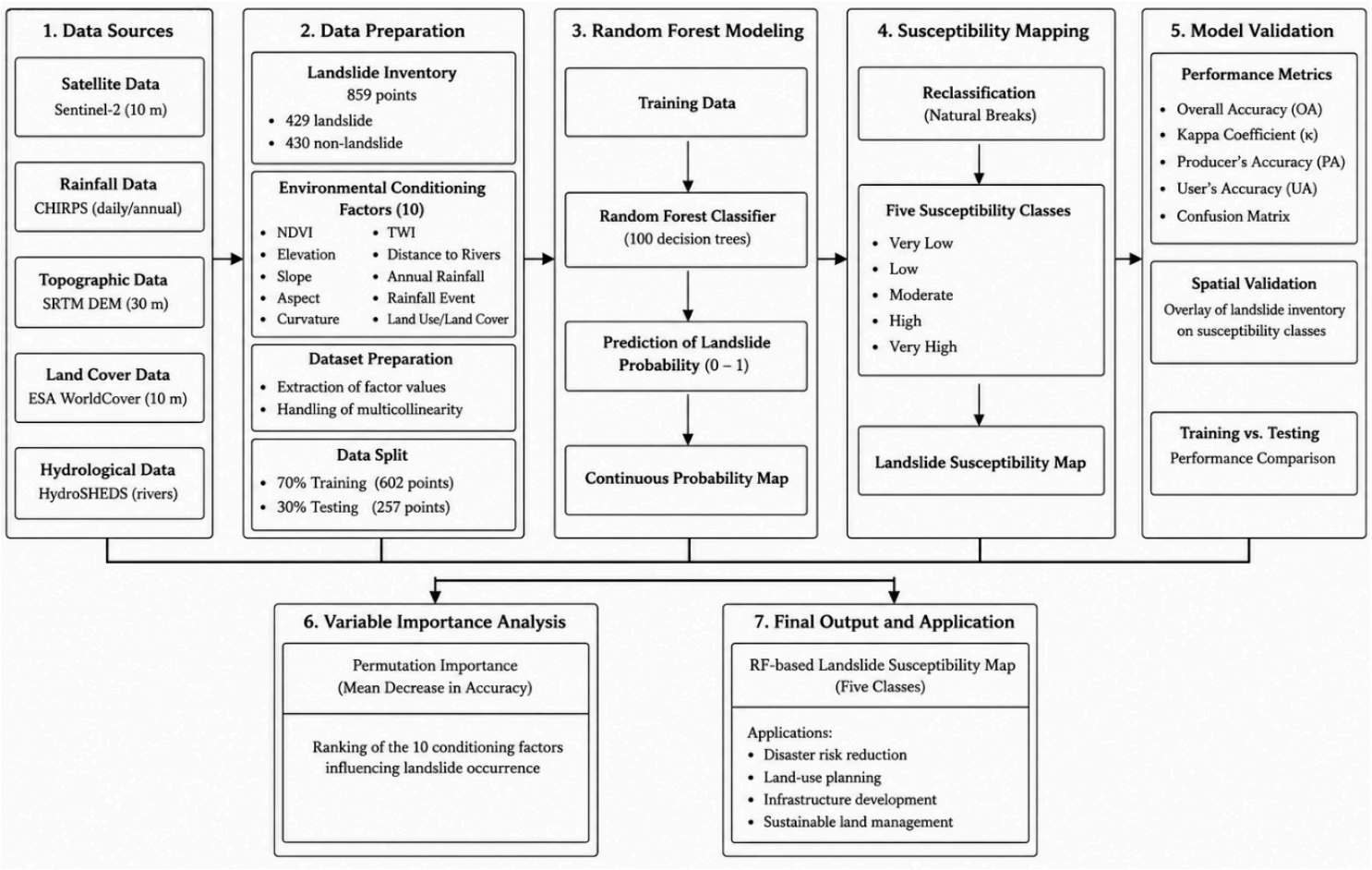


Figure 3

Methodological framework for RF-based landslide susceptibility mapping used in this study.

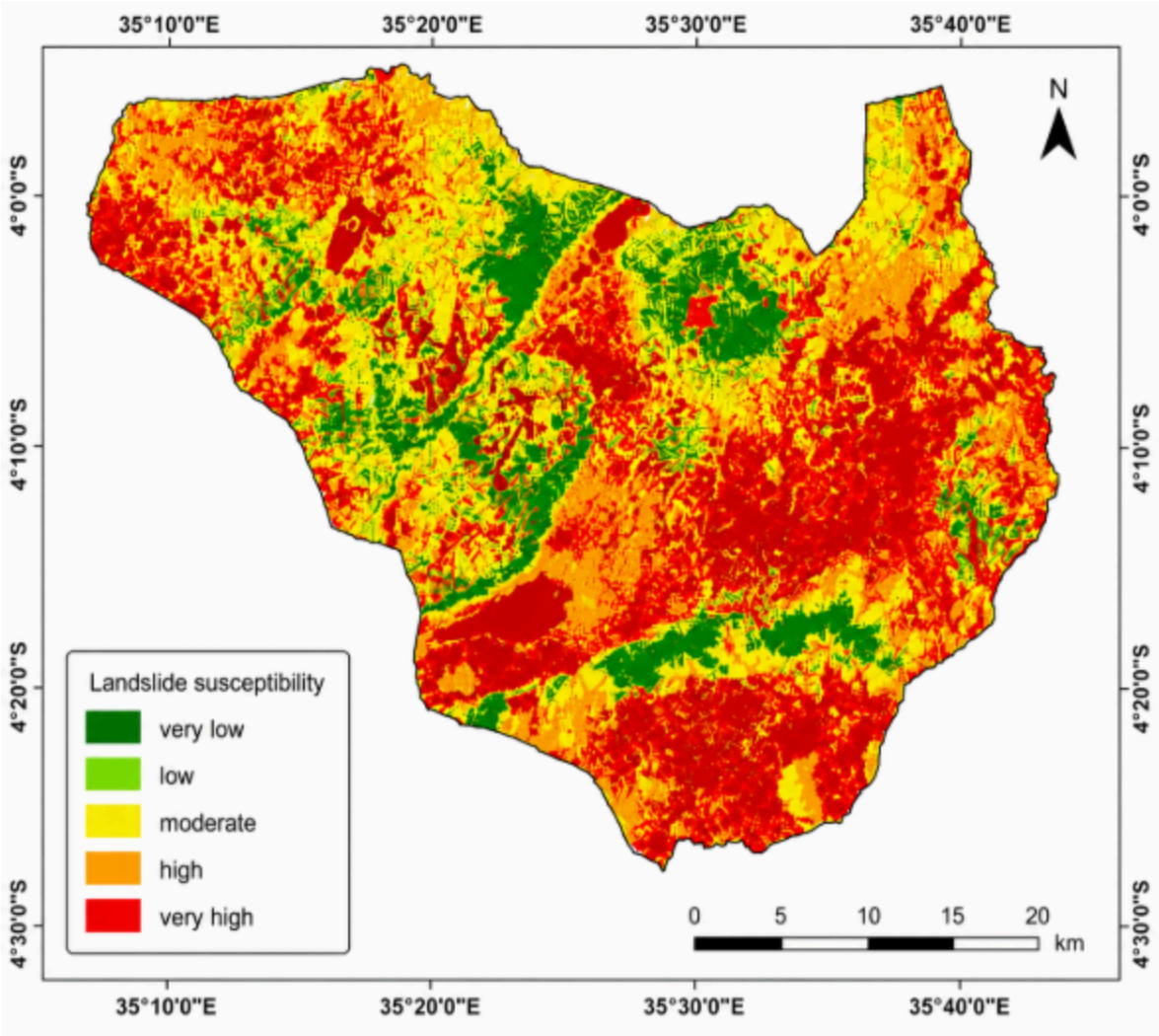


Figure 4

Landslide Susceptibility Map of Hanang District

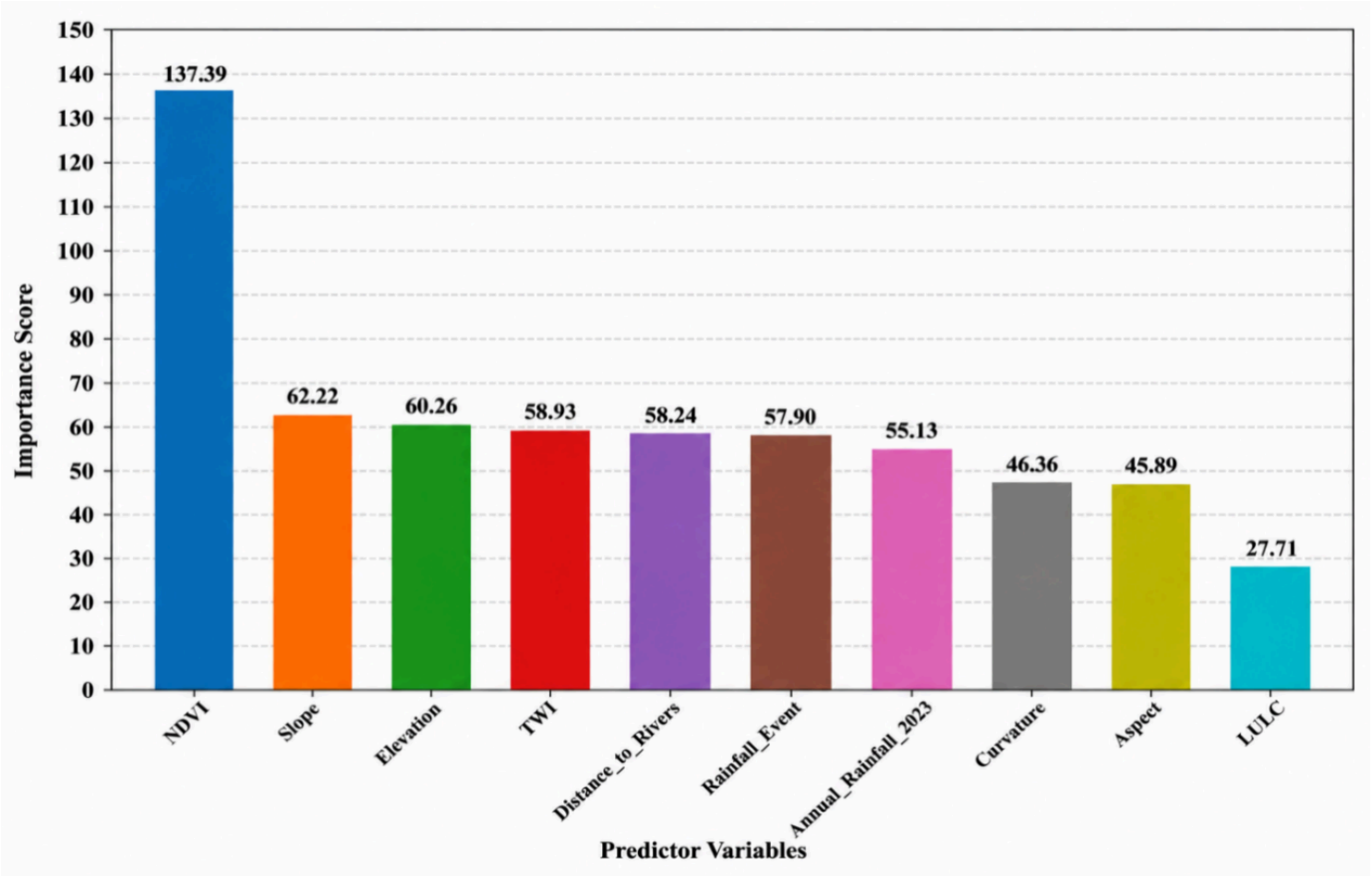


Figure 5

Variable importance from RF Model

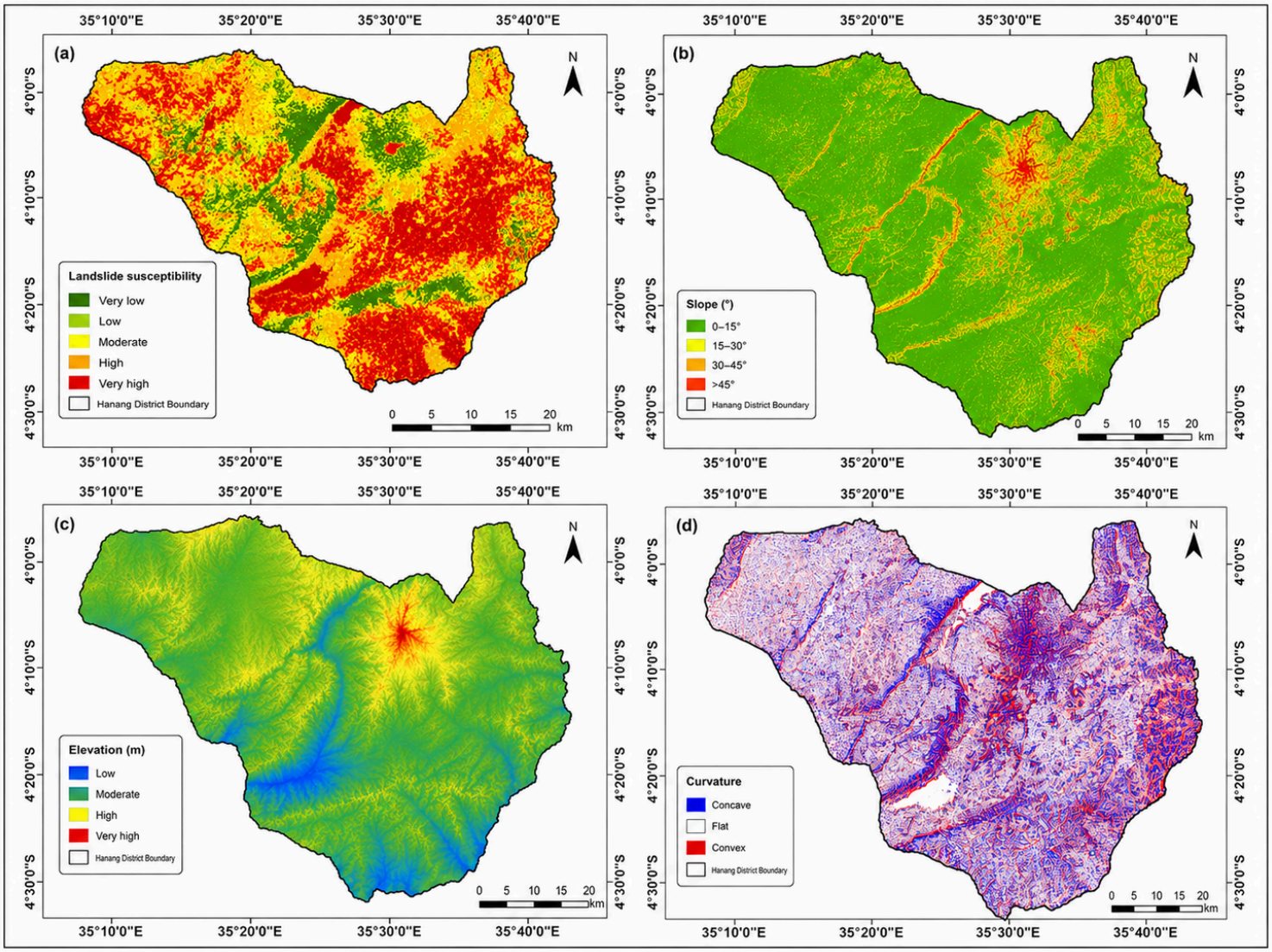


Figure 6

Spatial distribution of environmental conditioning factors: NDVI (a), Slope (b), Elevation (c), and Curvature (d)

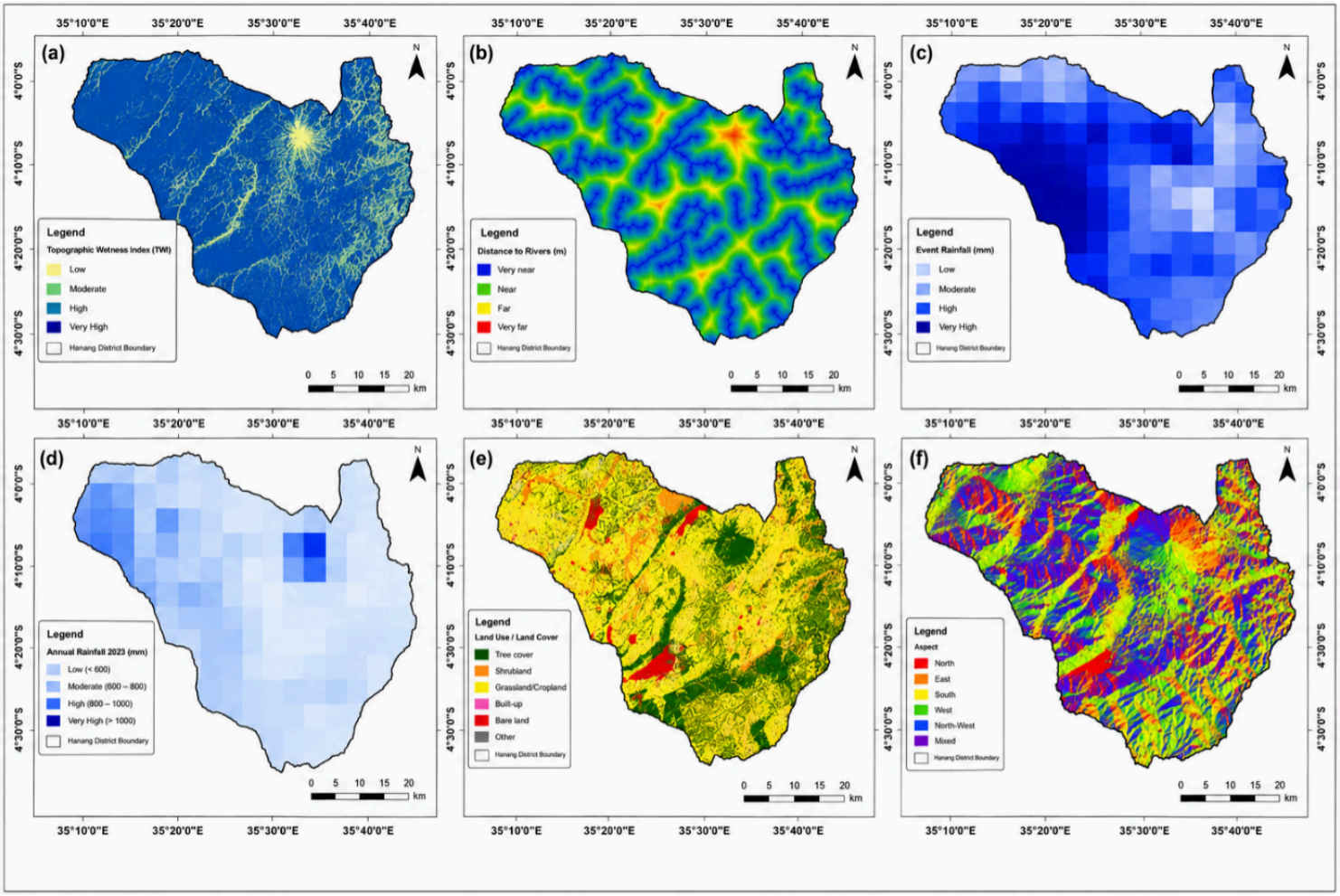
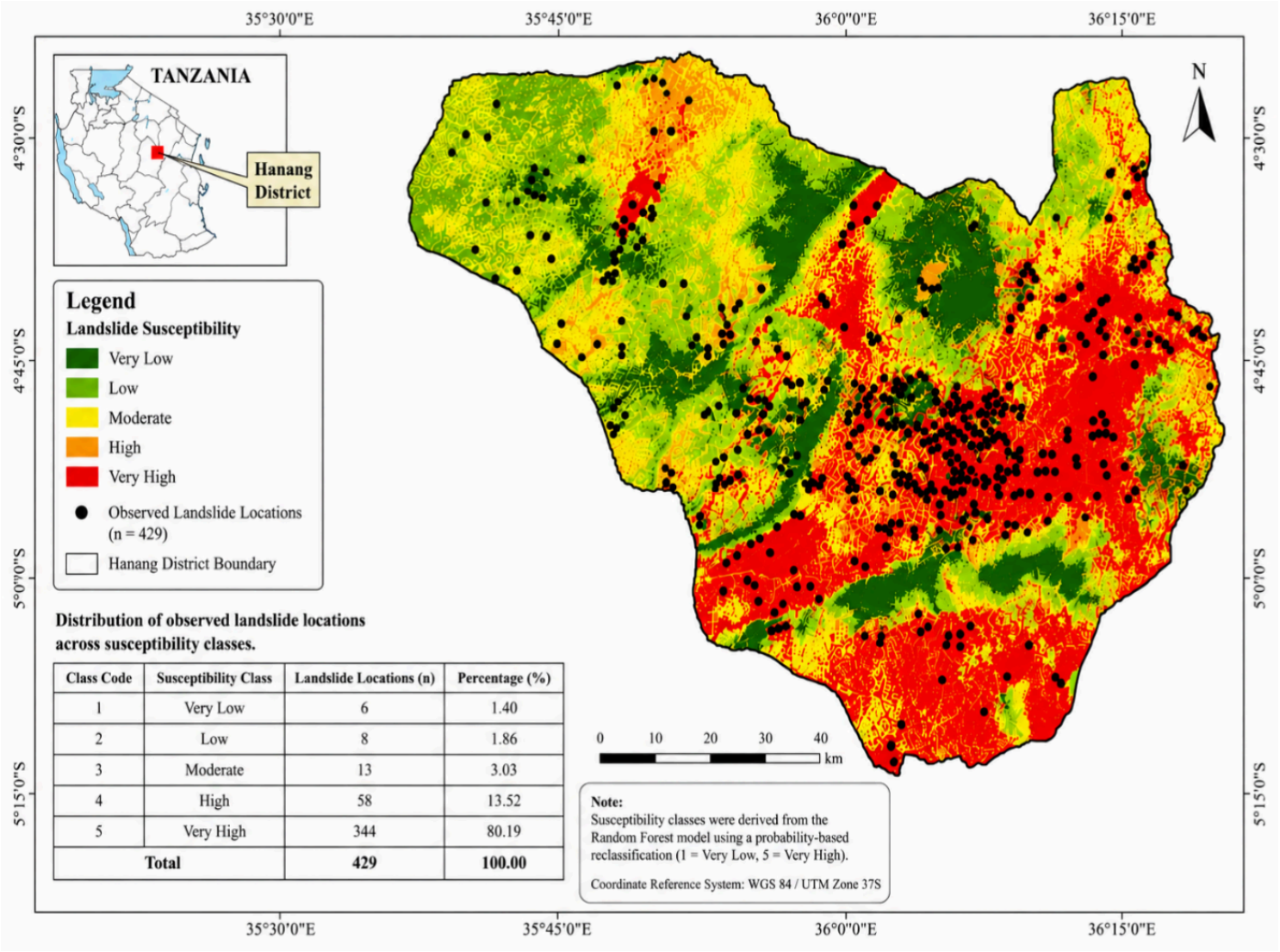


Figure 7

Environmental conditioning factors: TWI (a), distance to rivers (b), event rainfall (c), annual rainfall (d), LULC (e), and Aspect (f).



Distribution of observed landslide locations across susceptibility classes.

Class Code	Susceptibility Class	Landslide Locations (n)	Percentage (%)
1	Very Low	6	1.40
2	Low	8	1.86
3	Moderate	13	3.03
4	High	58	13.52
5	Very High	344	80.19
Total		**429**	**100.00**

Figure 8

Spatial Overlay of Observed Landslide Inventory on the Predicted Landslide Susceptibility Map