Research Article

Landslide Susceptibility Zonation of the Upper Satluj River Basin, Himachal Pradesh, India, Using a Random Forest Approach

Gulshan Verma and Ram Lal
Department of Geography, Himachal Pradesh University, Shimla, Himachal Pradesh, India.
Corresponding Author: Email: vermag433@gmail.com; ramhpushimla@gmail.com

Abstract
Landslides are one of the major natural disasters in the Himalayan region that results in a great loss of infrastructure, environment, and human lives. The rugged topography, active tectonics and compound geo-environmental conditions make the Upper Satluj River Basin in the Kinnaur district susceptible to such incidences. The objective of this study is to delineate landslide susceptibility zones using Random Forest model. For this, 13 conditioning factors were taken into consideration, such as slope, aspect, relative relief, lithology, density of lineaments, distance to thrust line, rainfall, soil texture and land use/land cover, NDVI, SPI and TWI. The landslide inventory was divided into training (70%) and testing (30%) datasets for model development and validation. The Receiver Operating Characteristic curve was used to model performance and calculated a high Area Under the Curve of 0.841, which demonstrates good predictive power. The resulting landslide susceptibility zonation map categorized the basin into five classes, namely, very low, low, moderate, high, and very high. The study area has a high to very high susceptibility of about 30.5 %, which is mostly linked with steep slopes, high relative relief, weak lithology, and structurally disturbed regions. The results demonstrate the effectiveness of the RF model in handling complex, non-linear relationships in mountainous terrains. The susceptibility map offers a credible foundation to land-use planning, infrastructure development, and reduction of disaster risks.
Keywords: Landslide susceptibility; Random forest; Himalaya; Upper Satluj Basin.

Introduction

Among the most frequent and destructive natural disasters in the mountainous landscapes, landslides are especially likely to occur in highly vulnerable areas such as the Himalayas. This involves the downward movement of rock, debris, or soil under the influence of gravity. According to Varnes (1984), landslide can be defined as the movement of a mass of rock, debris or earth down a slope, whereas Cruden (1996) pointed out that these movements were governed by a combination of geological, geomorphological and hydrological factors. In fact, landslides are not caused by one factor, rather it is a complex interplay of natural causes, such as heavy rainfall, earth tremors, and slope instability, and anthropogenic factors, such as road construction, deforestation, and unplanned development. Landslides cause significant socioeconomic losses in the world. They cause an estimated US$20 billion in damages to annual economies (Klose et al., 2016) and have taken over 110,000 lives since the turn of the 20th century (Niyokwiringirwa et al., 2026). The number of non-seismically induced landslides alone has caused about 32,322 deaths between 2004 and 2010 (Petley, 2012), a rate of over 4,500 deaths per annum. The developing countries, especially the mountainous areas of Asia, the Central and South American countries play the major part in the burden, especially because of the rapid urbanization process, deforestation, and the extreme weather caused by the climate, which in turn increases vulnerability. In India, there are around 0.42 million km² of land that is vulnerable to landslides (12.6% of the total area of the country) (National Remote Sensing Centre, 2023). It has the worst impact on the Himalayan region, where the Northwest and Northeast Himalayas occupy 0.14 million and 0.18 million km², respectively (National Remote Sensing Centre, 2023). The Indian Himalayas are characterised by young weak geology, rugged topography, heavy monsoon rainfall, seismic activity, and rising anthropogenic interventions, which frequently lead to landslides that affect transportation networks, devastate settlements, and take hundreds of lives every year. Major events include the 2013 Uttarakhand disaster (more than 580 confirmed deaths and total losses of over 5,700 including missing persons; National Institute of Disaster Management, 2013) and the 2023 Himachal Pradesh monsoon-related disasters (428 deaths; Himachal Pradesh State Disaster Management Authority, 2023; Chandel et al., 2025).

The current research is based on Upper Satluj River basin in the Kinnaur district of Himachal Pradesh in the Northwest Himalayan zone. Kinnaur is extremely susceptible owing to steep slopes of Satluj River valley, poor rock formations and seismic forces. The Urni landslide is a recurring event that has posed a threat to infrastructure on multiple occasions and demonstrated the necessity of thorough evaluation (Kumar et al., 2019). Landslide susceptibility zonation (LSZ) is an important tool used to locate and map landslide prone areas. It combines various conditioning factors. Over the years, a number of methods have been devised, including qualitative/heuristic (knowledge-based) methods, which utilize expert judgment, as well as quantitative ones, including statistical models (e.g., logistic regression, frequency ratio), probabilistic methods, and sophisticated machine learning algorithms (Das et al., 2023). Machine learning methods have become popular due to their capability to describe complex non-linear relationships within heterogeneous terrains. One of them is the Random Forest (RF) algorithm introduced by Breiman (2001), which is a potent ensemble machine learning algorithm. It builds several decision trees with bootstrapped samples and combines the predictions with majority voting. RF has some strengths such as high predictive capability, insensitivity to overfitting, multicollinearity, and the ranking of variable significance, which makes it very appropriate in mapping landslide susceptibility in complex Himalayan settings. The overall aim of the research is to create landslide susceptibility zonation map as of Upper Satluj River basin in Kinnaur District using a random forest model by incorporating multi-source geospatial data obtained through remote sensing and GIS platforms. Sustainable land-use planning, infrastructure development, early warning systems, and community resilience efforts can be based on accurate susceptibility maps. Although Kinnaur district is highly vulnerable, only a limited number of studies have applied machine learning techniques for landslide susceptibility zonation in this region. Such studies are critical in the environment of accelerating climate change and rising extreme weather conditions in the Himalayas to preserve lives, ecosystems, and ensure economic well-being in susceptible districts such as Kinnaur in the long term.

Study Area

The present study focuses on the Upper Satluj River Basin in Kinnaur district, Himachal Pradesh. This basin covers the majority of the Kinnaur district, excepting the northeast of Sangla tehsil which flows into Yamuna River system (Figure1). The total area of the study region is approximately 6,245 km². Geologically, the region is a part of the Himalayan orogenic belt and is dominated by intricate tectonic environments and weak rock structures. The primary lithological formations in the basin include the Kanwar Group (comprised of mainly quartzite and black shale), Vaikrita Group (mainly schist and gneiss), and Haimanta formations (mainly shale and slate). Together with the strong tectonic activity and the geomorphological processes, these rocks have an important impact on slope stability and landscape development in the basin. Ongoing movement of the tectonic plates along the prominent structures of the Main Central Thrust (MCT) and the South Tibetan Detachment System (STDS) has generated closely compressed and faulted bedrock, whereas a current tectonic movement has been observed to fracture slope materials (Kumar et al., 2021). This tectonic system is the source of the natural instability of the slopes within the basin.

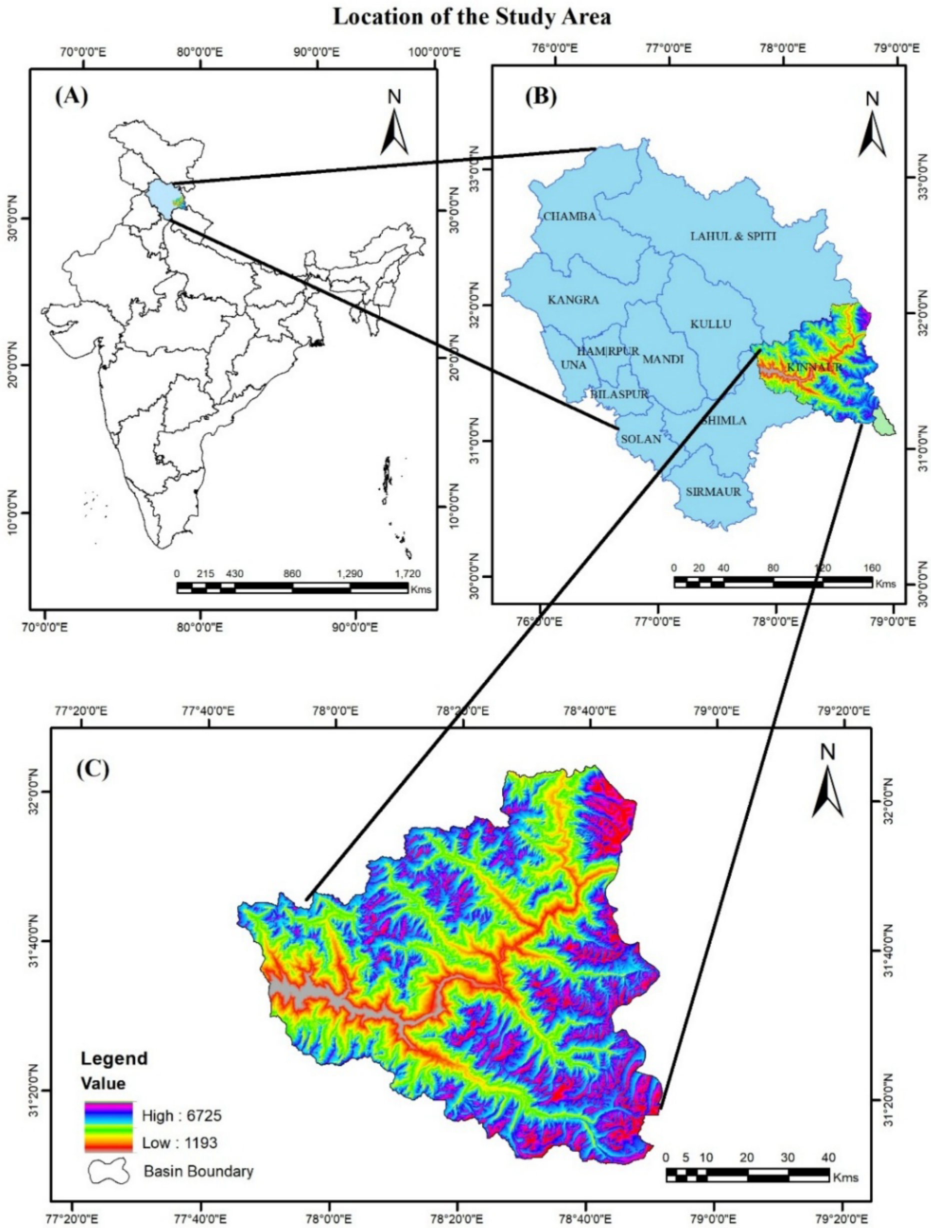

Figure 1. (A) Location of Himachal Pradesh and the Upper Satluj Basin within India; (B) Location of the Upper Satluj Basin in Himachal Pradesh; (C) Upper Satluj River Basin with elevation distribution in Kinnaur District.

The surface is very rugged and mountainous, the heights ascending between 1,193 m in the bottom valleys, and 6,725 m in the high peaks, making the relative relief, 5,532 m, very remarkable. The Satluj River, originating from the Tibetan Plateau, forms the primary drainage system of the basin. Its larger tributaries within the basin are the Spiti, Baspa, Tidong and Ropa rivers. The hydrological networks are significant in determining the geomorphology and water resources of the area (Kuniyal et al., 2019). The Upper Satluj Basin is characterised by a significant climatic variation because of the changes in altitude. The lower elevations experience a temperate climate, while the upper reaches are characterized by cold desert conditions. (Jamwal et al., 2020). The average yearly rainfall is 500 to 1,200 mm with a huge percentage of it being snowfall at the greater heights, particularly throughout the winter season. The southwest monsoon winds become weaker as it moves westward to eastward forming clear climatic zones in the basin. The vegetation and land cover patterns are the best reflection of this climatic gradient. The lower and middle altitudes are characterized by coniferous forests with an abundance of pine, deodar and fir. Conversely, elevations that are higher have alpine grasslands, barren rocky slopes and huge glacial terrains. The valley floors, especially in and around Kalpa and Sangla, are home to apple orchards and other horticultural products which form the main agrarian economy of the region. The Upper Satluj Basin is very prone to earthquakes and frequent landslides due to the fact that it is located in the seismically active Himalayan belt. The terrain is naturally unstable due to steep slopes, poor geological formations and heavy seasonal rainfall.

Methodology

In the current research, landslide susceptibility zonation (LSZ) was created employing the machine learning algorithm of Random Forest (RF). The methodology was based on a systematic multi-step approach that involved: (i) preparation of landslide inventory, (ii) selection of geo-environmental causative variables, (iii) evaluation of multicollinearity (iv) optimization of the hyperparameters of the model and (v) validation of the final susceptibility map. The elaborate methodology which was applied in this study is as detailed in the subsequent sections.

Landslide Inventory

Landslide inventory is one of the most crucial inputs to any susceptibility mapping. It is a systematic recording of the place, nature and scale of previous landslide events in the study area thus exposing significant spatial patterns and their correlation with the underlying geo-environmental conditions. An inventory of high quality enhances both the predictability and reliability of statistical and machine learning-based models (Guzzetti et al., 2012; Fell et al., 2008). In this paper, inventory was made by integrating both the secondary data of the Bhukosh portal of Geological Survey of India and a field survey in the area. A total of 1,543 landslides were identified and mapped. The inventory was randomly divided into 70: 30 proportions to use in modelling (fig. 3). Training the Random Forest model used 70% of the landslide locations and the other 30% was reserved to validate the model. In order to maintain a balance in class representation and minimize bias, the same number of non-landslide points was randomly generated in the study area. The Receiver Operating Characteristic-Area Under the Curve (ROC-AUC) method was used to analyze the final model performance.

Causative Factors

Thirteen geo-environmental causative factors were chosen in this study and were incorporated in the Random Forest machine learning to generate the landslide susceptibility zonation map. These parameters were selected based on the literature available and are a combination of topographic, hydrological, geological, structural, and vegetation-related factors that are known to have a significant effect on slope instability in high-relief, tectonically active areas such as Kinnaur.

Slope

Slope angle is universally recognised as one of the most important factors controlling landslide occurrence because steeper gradients directly increase gravitational shear stress while simultaneously reducing the frictional resistance of the slope material (Ayalew & Yamagishi, 2005; Lee, 2005; Pradhan & Lee, 2010) (Table 1). The slope layer in the present study was derived from the ALOS PALSAR Digital Elevation Model (12.5 m resolution) (table 1) and reclassified into five classes: <15°, 15–25°, 25–35°, 35–45°, and >45° (Figure 2).

Table 1 Data sources of landslide causative factors

Parameter	Source	Scale / Resolution
Slope gradient, aspect, relative relief, TWI, SPI, drainage density	ALOS PALSAR DEM https://search.asf.alaska.edu	12.5 m
Lithology, lineament, thrust	Geological Survey of India https://bhukosh.gsi.gov.in/Bhukosh/Public	1:50,000
Soil	National Bureau of Soil Survey and Land Use Planning (NBSS&LUP)	1:50,000
Rainfall	IMD Gridded Data https://imdpune.gov.in/	0.25° x 0.25°
LULC, NDVI	Sentinel-2 https://dataspace.copernicus.eu/data-collections	10 m

Slope Aspect

Slope aspect affects local microclimate, soil moisture retention, vegetation cover, and weathering intensity, all of which indirectly influence slope stability (Wieczorek et al., 1997). North- and northeast-facing slopes in the region often retain more moisture and are therefore more prone to failure in monsoon-dominated areas. The aspect map was extracted from the same ALOS PALSAR DEM and classified into nine directional classes (fig. 4b) following the common practice (Bisht et al., 2024).

Relative Relief

Relative relief is an effective proxy of terrain ruggedness and erosion capability of mountain terrain. Greater relative relief values are more slope instability because of increased undercutting and gravitational forces (Anbalagan, 1992; Saha et al., 2002), This parameter has often been employed in susceptibility mapping. The DEM of the study area was used to derive the relative relief which was then placed into five classes using the Natural Breaks technique.

Drainage Density

Drainage density is an indicator of channel dissection and is strongly related to the surface runoff, erosion of toes, and slope saturation (Gorsevski et al., 2006; Lee and Pradhan, 2007). Highly drained areas are more prone to landslides due to increased erosion and seepage of ground water. This aspect is regularly added to the studies of the Himalayas. The ALOS PALSAR DEM was used to calculate the density of drainage using hydrological tools in the ArcGIS (fig. 4d). First, the DEM was cleaned up by filling in surface depressions (sinks) so that the flow continuity is achieved. Later, layers of flow direction and flow accumulation were created to facilitate additional hydrological analysis.

Stream Power Index (SPI)

SPI measures the erosive power of concentrated surface runoff and is especially used to detect areas in which channel incision can cause adjacent slopes to become unstable. It can be expressed mathematically as:

$$SPI = As \times \tan \beta \tag{1}$$

where β is the local slope gradient in radians and As is the catchment area (Moore et al. 1991). Having applied the above equation to the DEM the obtained map was further categorized into five classes based on the Natural Breaks (Jenks) classification method (fig. 4e).

Topographic Wetness Index (TWI)

TWI identifies areas prone to water accumulation and soil saturation, which significantly reduce slope shear strength. It is one of the most commonly used hydrological factors in landslide studies (Gokceoglu & Aksoy, 1996; Lee & Evangelista, 2006). The Topographic Wetness Index (TWI) was calculated according to the method suggested by Moore et al. (1991), expressed as:

$$\mathrm{TWI} = \ln\left(\frac{A_s}{\tan_\beta}\right) \tag{2}$$

where A_s represents the specific catchment area (flow accumulation), tan β denotes the slope gradient, and ln refers to the natural logarithm. The resulting TWI map was then divided into five classes using the Natural Breaks (Jenks) method of classification (fig. 4f).

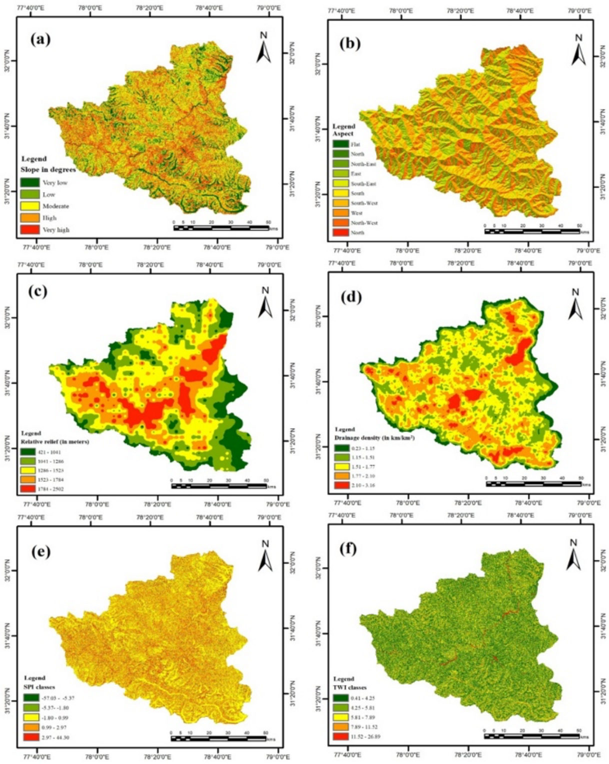

Figure 2 Causative factors considered for landslide hazard mapping: (a) Slope, (b) Aspect, (c) Relative Relief, (d) Drainage Density, (e) SPI, and (f) TWI.

Rainfall

Intensive or long-term rainfall commonly causes landslides since it causes the saturation of the soil, the pressure of pore-water, and the eventual decline in shear strength of the slope materials. The increase in pore-water pressure as rainfall increases reduces the effective stress in the slope, and thus, predisposes the slope to failure. Investigations carried out by Guzzetti et al. (2006) and Dahal and Hasegawa (2008) have shown that occurrence of landslides is closely related to areas with great rainfall (annual or seasonally). In this research, the rain distribution map of the region was drawn based on gridded information of the India Meteorological Department. The annual rainfall data during the years 2014-2024 were used and the resulting data were categorized into five groups to represent different levels of rainfall (fig. 5a).

Lithology

Lithology is the composition and the physical properties of rocks and soils in a particular locality. It is an important factor in the strength, permeability and weathering behavior of slope materials. Some rock types, especially weaker rocks like schist and shale are more susceptible to landslides since they have low cohesion and are more vulnerable to water infiltration (Gupta and Joshi, 1990). In this study, lithological data was obtained in the Geological Survey of India at 1:50,000 scale (fig. 5b). It was further divided into 12 lithological units that were analysed as a dataset (table 2).

Table 2: Different Rock Types in Various Lithological Groups.

Group Name	**Lithologic**
Kanawar	Quartzite, Black Shale with Nodules
Lilang	Dolomite, Limestone, Oolitic Limestone
Vaikrita	Schist, Gneiss, Migmatite, Quartzite, Marble
Rampur (Naraul)	White-Green Quartzite, Phyllite, Basic Flows
Jeori-Wangtu Banded Gneissic Complex	Granitods, Gneisses and Migmatites
Kulu	Streaky And Banded Gneiss.
Sanugba	Shale, Slate, Calcarenite, Limestone, Dolomite
Kuling	Calcareous Sandstone, Shale, Quartzite, Conglomerate
Haimanta	Shale, Slate, Siltstone, Quartzite
Jutogh (Undiff.)	Psammatic Gneiss, Garneti Mica Schist, Amphibolite
Rakcham Granite	Medium To Coarse Biotite Granite
Nako Granite	Granite

Source: Geological Survey of India

Soil

The type of soil directly influences cohesion, permeability and shear strength. The different soil textures react differently to saturation and thus are essential to proper modelling. The soil layer was obtained from the National Bureau of Soil Survey and Land Use Planning (NBSS&LUP) at a scale of 1:50,000 and reclassified based on soil texture (Figure 3).

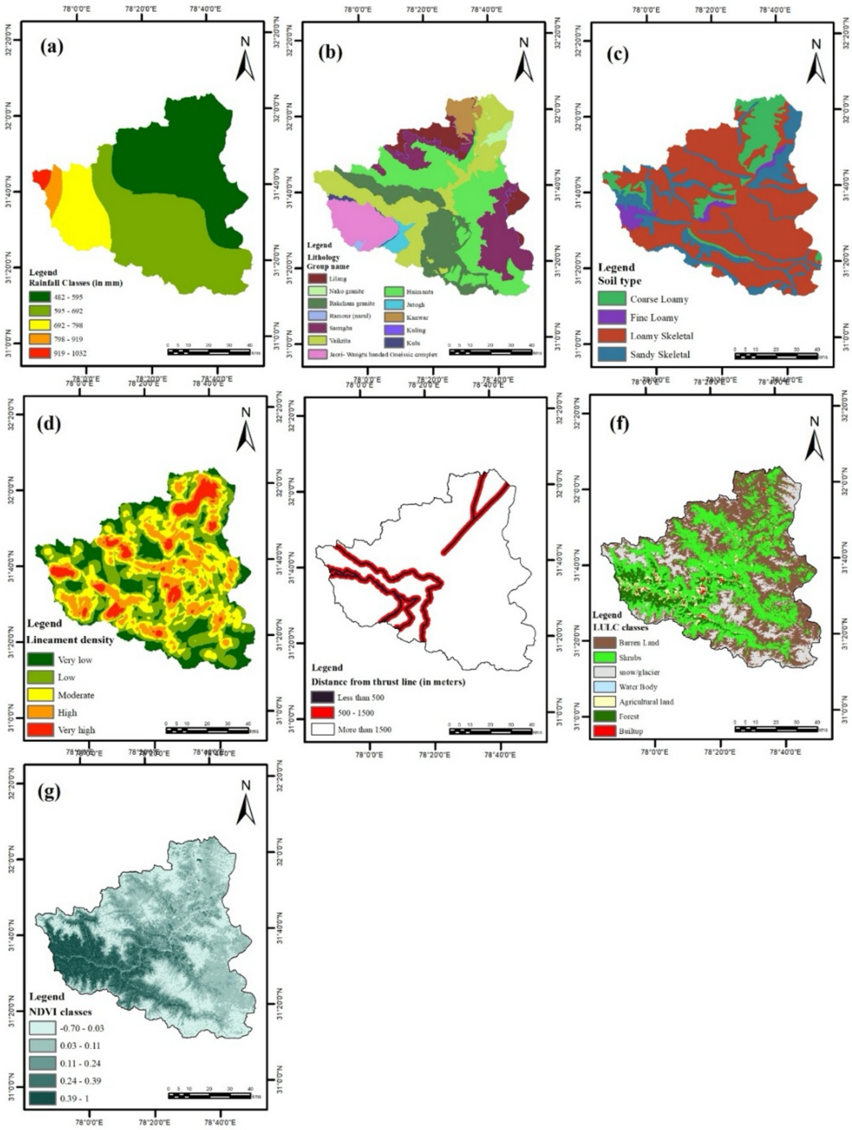

Figure 3 Causative factors considered for landslide hazard mapping: (a) Rainfall, (b) Lithology, (c) Soil, (d) Lineament Density, (e) Distance from thrust line, (f) LULC, and (g) NDVI.

Distance to Thrust Line

Regions that are found to be in close proximity to the large thrust faults are regions of high deformation, that is, the rock masses are highly sheared, fractured and weakened. These conditions cause slopes to be much more unstable and hence these areas are more prone to landslides (Anbalagan, 1992). The effects of the thrust proximity were included in the current study by adding the factor of distance to thrust, which was divided into three buffer areas: 0–500 m, 500–1500 m, and more than 1500 m.

Land Use/Land Cover

It is important to note that LULC is influential in determining the stability of the slope and the process of surface erosion, root reinforcement, and the degree of human-induced disturbances. Barren soils and urbanized areas tend to be more susceptible to landslides than a forest area, mostly because roots are bound with less strength and because there is more surface runoff (Saha et al. 2005; Pradhan and Lee, 2010). In this work, supervised classification of Sentinel-2 images (October 2023) was used to prepare the LULC map.

Normalized Difference Vegetation Index

NDVI measures the density and health of vegetation, which offers mechanical root strength and decreases surface erosion. Lower NDVI values typically correspond to higher susceptibility (Gupta & Joshi, 1990; Lee & Pradhan, 2007). The NDVI band was calculated with Sentinel-2 images (fig 5f) based on the following formula:

$$NDVI = (NIR - Red) / (NIR + Red) \qquad (3)$$

Multicollinearity Analysis

A multicollinearity analysis has been conducted prior to including the thirteen causative factors into the Random Forest model based on Variance Inflation Factor (VIF) and Tolerance statistics. It is a necessary step in landslide susceptibility study, especially when a combination of geo-environmental variables in complicated topographies such as the Himalayas is considered (Chauhan et al., 2025; Joshi & Bhandary, 2025). High multicollinearity occurs when factors are strongly inter-correlated, which can distort the model's ability to accurately assess the individual contribution of each variable, inflate variance in predictions, and reduce the overall reliability and interpretability of the results (Dormann et al., 2013). Variance Inflation Factor (VIF) measures how much the variance of a regression coefficient is inflated due to multicollinearity, while Tolerance is its reciprocal (1/VIF) and indicates the proportion of variance in a factor not explained by the other variables (O'Brien, 2007).

Landslide Susceptibility Zonation Model

LSZ has traditionally been carried out using a variety of modelling approaches that can be broadly grouped into three main categories: qualitative, semi quantitative and quantitative (Das et al., 2023). Qualitative and semi quantitative rely on expert judgment and multi-criteria decision-making techniques such as the Analytic Hierarchy Process (AHP) to assign weights to causative factors; they are simple and useful when data are limited but remain subjective (Merghadi et al., 2020; Shano et al., 2020). Statistical models, including Frequency Ratio (FR) and Logistic Regression (LR), quantify the spatial relationship between past landslides and conditioning factors through bivariate or multivariate analysis and have been widely applied in the Himalaya because of their transparency and ease of interpretation (Wang et al., 2020). Deterministic models use physically based equations (e.g., the infinite slope stability model) to calculate factor of safety; they are highly accurate at local scales but require detailed geotechnical data and are rarely feasible for large basins (Merghadi et al., 2020).

Machine-learning models have been especially popular in recent years due to their ability to reproduce complex non-linear relationships without making strong assumptions about the distribution of data; one example of such an ensemble method is the Random Forest (RF) technique, which has repeatedly demonstrated better performance in the Himalayan terrain (Merghadi et al., 2020; Taalab et al., 2018). The Random Forest (RF) algorithm was used to create the LSZ map of the Upper Satluj River Basin in the current study due to its strength, high predictive accuracy, and the ability to deal with complicated and non-linear associations between various conditioning factors (Akinci et al., 2020; Park and Kim, 2019). RF algorithm is an ensemble learning algorithm, originally proposed by Breiman (2001), which builds large decision trees based on bootstrapped samples of the input data and randomly chosen subsets of variables on each node split. The data on landslide inventory was taken as the dependent variable in this study whereas various geo-environmental and anthropogenic conditioning factors were taken as the independent variables. The data was randomly separated into training (70%), and validation (30%) to guarantee the assessment of the model is objective. Bootstrap sampling (with replacement) was used to train each decision tree of the RF model to increase the variety of the model and decrease the variance. The implementation of the model was carried out in R environment. To achieve a better model optimization and validation, the caret package was utilized, and it simplified the systematic cross-validation and hyperparameter tuning. The most important hyperparameters that were looked at in the creation of the models were the number of trees (ntree) and the random selection of number of variables at each split (mtry). The ntree parameter was varied between 100 and 500 to identify the optimal number of trees required for stable model performance. It was noted that the out-of-bag (OOB) error flattened off around 500 trees; hence, the final model was built with 500 trees to achieve a balance between computational efficiency and predictive accuracy. The mtry hyperparameter was optimally determined by a systematic grid search between 2 and 15. The best mtry value was set to 3 since it gave the highest Area Under the Curve (AUC) in the cross-validation indicating better performance of the model. The default method of sampling was a bootstrap sampling with replacement, which created different training subsets per tree and thus increased the robustness of models. The out-of-bag (OOB) error estimate and independent validation data were used to assess model performance. The resulting susceptibility map was better interpreted and represented spatially by categorizing it into five groups: very low, low, moderate, high, and very high using the Natural Breaks (Jenks) classification system.

Model Performance

Validation of the model was performed to assess predictive performance and reliability of the model by using the Receiver Operating Characteristic (ROC) curve and the Area Under the Curve (AUC) measure. ROC curve shows the correlation between the true positive rate (sensitivity) and false positive rate (1-specificity) at various threshold values and gives a complete measure of the classification performance (Chung and Fabbri, 2003). The AUC of 0.5 to 1.0 shows that the model is able to differentiate correctly between landslide and non-landslide sites with the values nearer to 1 being more predictive. To validate the model, 30% of the landslide inventory was used for testing.

Results and Discussion

In order to determine whether multicollinearity is present in the 13 chosen landslide conditioning factors, the Variance Inflation Factor (VIF) and Tolerance values were calculated. A value that is below 0.1 is usually deemed as a high level of redundancy, and a VIF above 10 is an indication of severe problems of multicollinearity (Yi et al.,2020). The findings showed no significant multicollinearity between the variables and all VIF values were below 2 and Tolerance values were always above 0.8. Aspect had the lowest value of collinearity (VIF = 1.03), and Stream Power Index (SPI) and Topographic Wetness Index (TWI) had slightly higher but nevertheless acceptable values (VIF = 1.21 each), which is a medium value of correlation (Table 3). No variable was dropped since none of the VIF values was greater than the generally accepted level of 5.

Table 3 Multicollinearity Analysis of Causative Factors

Collinearity Statistics		
Variables	VIF	Tolerance
Aspect	1.03	0.972
Drainage Density	1.13	0.884
Lineament Density	1.08	0.926
Lithology	1.16	0.863
NDVI	1.22	0.819
Rainfall	1.27	0.785
Relative Relief	1.1	0.911
SPI	1.21	0.83
Distance from Thrust	1.17	0.853
TWI	1.21	0.828
LULC	1.12	0.897
Slope	1.13	0.887
Soil	1.19	0.842

The landslide hazard spatial distribution reveals that the very low (23.8) and low (22.3) hazard areas are confined mainly to the southern area of the study area whereas the high (20.3) and very high (10.2) hazard areas are found in the northern, northeastern and central regions (Table 4; fig. 6). This trend is indicative of active Himalayan geomorphology. The controlling factors are slope with almost 60% of very steep slopes (>45°) located in high hazard regions and more than 95 percent gentle slopes (<15°) being stable. This trend has been observed in other basins nearby, though the correlation is greater in the Upper Satluj Basin because there is greater local relief (Singh et al., 2024). The landslide susceptibility is also highly dependent on lithology. Mostly the Kulu Group is linked with high hazard areas (approximately 65%) because of the fractured and weathered metamorphic rocks (Gupta and Joshi, 1990). Likewise, Jutogh Group and Jeori-Wangtu Banded Gneissic Complex are highly susceptible in nature because of the low metasedimentary and gneissic formations (Saha et al., 2002; Pradhan and Lee, 2010). Conversely, more stable units like Lilang Group, Nako Granite, Kuling formation are majorly restricted to low hazard areas because they are stronger and less permeable. These trends are also supported by relief and soil features. Most high and very high hazard areas are highly related to high relative relief areas and over 94 percent of low-relief areas belong to very low and low hazard areas, in line with Das et al. (2023). Fine loamy soils are more susceptible (approximately 61% in the high hazard areas), but sandy skeletal and loamy skeletal soils are comparatively more resilient.

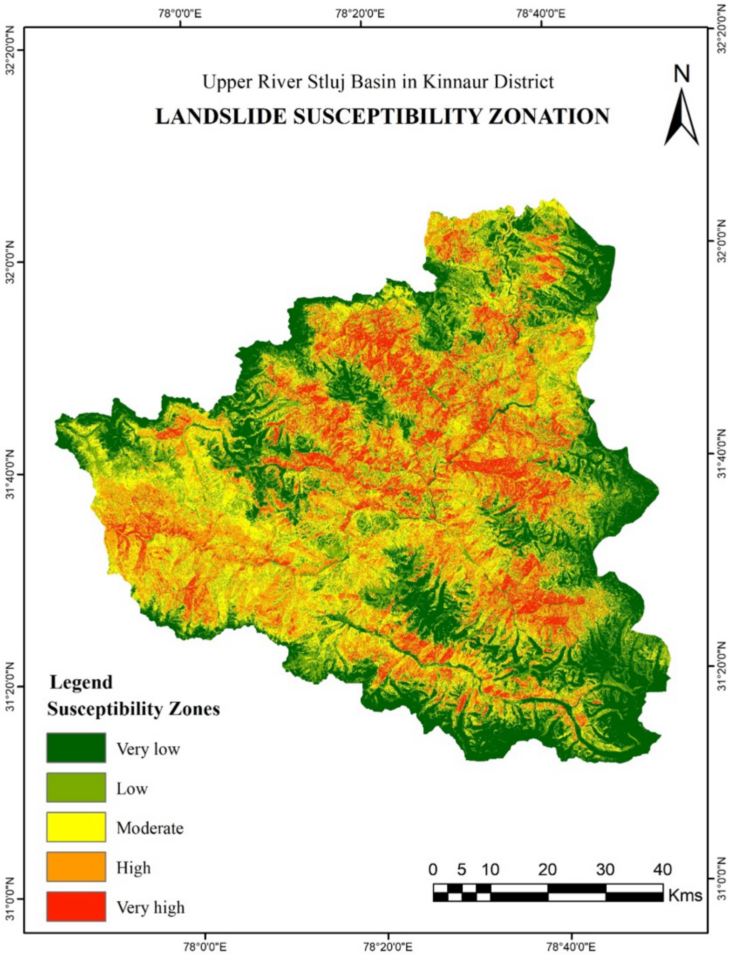

Figure 4. Landslide Susceptibility Zonation map

Table 4 Landslide Susceptibility Zones

Susceptibility Zone	Description	Area (in km)	Area (%)
I	Very Low Susceptibility	1488	23.82
II	Low Susceptibility	1390	22.26
III	Moderate Susceptibility	1460	23.38
IV	High Susceptibilit	1269	20.32
V	Very High Susceptibility	639	10.23

Random Forest model showed a high predictive power with ROC–AUC of 0.841 (Figure 5), which means that this model has a high accuracy in the process of identifying areas at risk of landslides. The resulting value of the AUC is comparable to the outcomes of other studies of the same nature, such as Liu et al. (2023) (AUC ≈ 0.87), Badapalli et al. (2025) (AUC ≈ 0.89) and Abu El-Magd et al. (2021) (AUC ≈ 0.86), demonstrating.

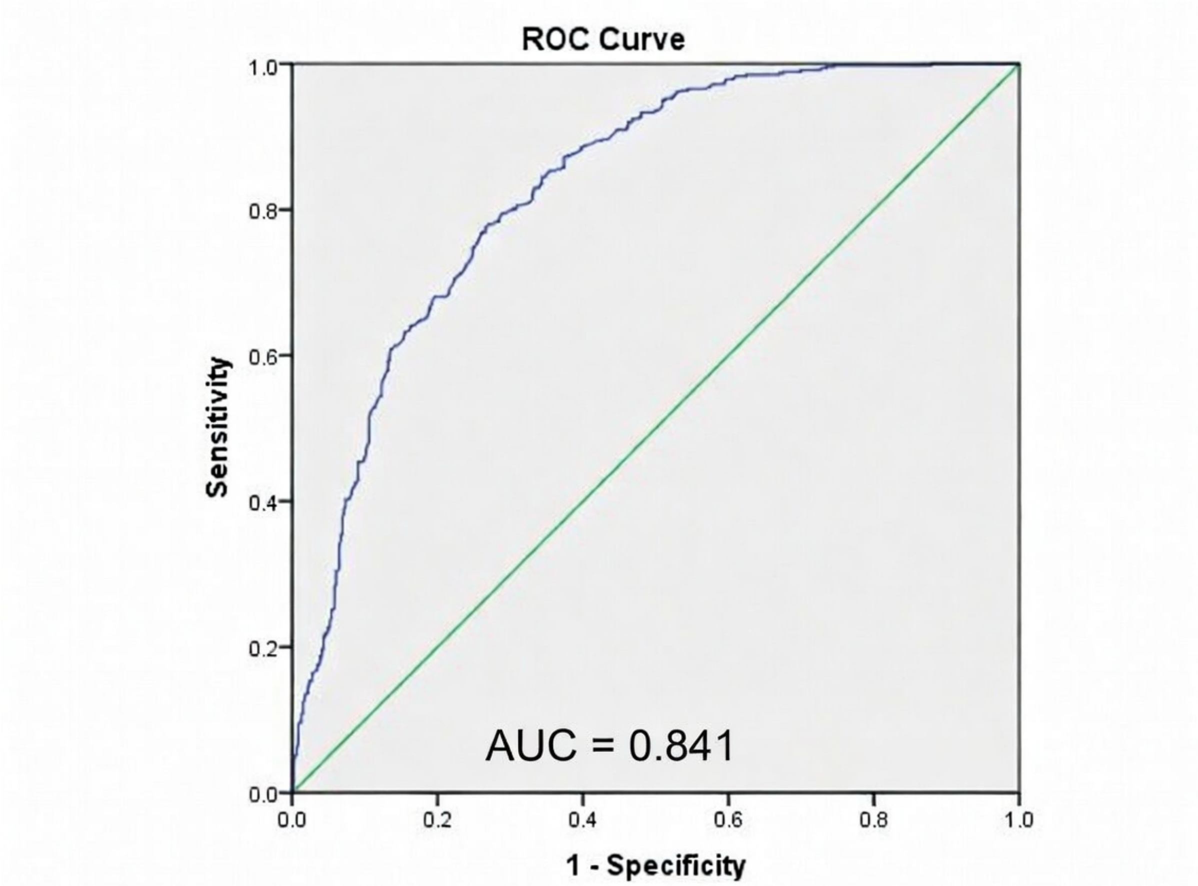

Figure 5 Validation through ROC – AUC

Conclusion

The current research enhances the knowledge of the landslide susceptibility in the Upper Satluj River Basin using a GIS-based Random Forest (RF) model. The model describes the intricate relationships between geo-environmental and human-made factors governing slope instability in this delicate Himalayan landscape. The resultant landslide susceptibility zonation (LSZ) map shows that a significant part of the basin has moderate to very high hazard zones and almost one-third of the basin has been classified as highly vulnerable. These areas are mainly related to steep slopes, high relative relief, structurally weak lithological units, and tectonically active and drainage concentrated areas. The implications of the results on practical implications on risk mitigation and planning are significant. The LSZ map will be a valuable decision-support information in terms of land-use planning as it allows to identify high-risk zones and influence the development of infrastructure, such as roads and hydropower projects, in the areas that are not threatened. It also underpins specific mitigation measures like slope stabilization practices and introduction of localized early warning mechanisms, thus making communities resilient. The RF model showed good predictive capabilities (AUC = 0.841) which proved that the model is appropriate when assessing landslide susceptibility in complex mountainous areas. The fact that it can deal with non-linear relationships and heterogeneous data allows it to be especially useful in Himalayan settings. But the model relies on the quality of input data and presupposes a steady conditioning factor over time. All in all, the paper provides a credible geospatial model of landslide danger evaluation in the Upper Satluj Basin and leads to the advancement of scientific knowledge and effective mitigation of risks. This study can be expanded by future studies through hybrid models that combine machine learning with physically based models, and high-resolution, multi-temporal data to more effectively represent dynamic triggering processes like rainfall and seismic activity. This would enhance the accuracy and usefulness of landslide susceptibility models in mountainous areas that are complicated.

To sum up, the research offers a scalable method of determining landslide-prone regions and aids in sound decision-making towards sustainable development and reduction of disaster risks in the Himalayan region.

References

Abu El-Magd, S. A., Ali, S. A., Pham, Q. B., & et al. (2021). Spatial modeling and susceptibility zonation of landslides using random forest, naïve bayes and K-nearest neighbor in a complicated terrain. Earth Science Informatics, 14(4), 1227–1243.

Akinci, H., Kilicoglu, C., & Dogan, S. (2020). Random forest-based landslide susceptibility mapping in coastal regions of Artvin, Turkey. *ISPRS International Journal of Geo-Information*, *9*(9), 553.

Anbalagan, R. (1992). Landslide hazard evaluation and zonation mapping in mountainous terrain. *Engineering Geology, 32*(4), 269-277.

Ayalew, L., & Yamagishi, H. (2005). The application of GIS-based logistic regression for landslide susceptibility mapping in the Kakuda-Yahiko Mountains, Central Japan. *Geomorphology, 65*(1-2), 15-31.

Badapalli, P. K., Nakkala, A. B., Kottala, R. B., Gugulothu, S., Hasher, F. F. B., Mishra, V. N., & Zhran, M. (2025). Landslide susceptibility level mapping in Kozhikode, Kerala, using machine learning-based random forest, remote sensing, and GIS techniques. *Land*, *14*(7), 1453.

Bisht, S., Rawat, K. S., & Singh, S. K. (2024). Earth observation data and GIS based landslide susceptibility analysis through frequency ratio model in lesser Himalayan region, India. *Quaternary Science Advances*, *13*, 100141.

Breiman, L. (2001). Random forests. *Machine learning*, *45*(1), 5-32.

Chandel, A. S. (2025). Geospatial and statistical assessment of monsoon-induced disasters in Himachal Pradesh: insights from the 2023 floods and landslides. *All Earth*, *37*(1), 1-48.

Chauhan, V., Gupta, L., & Dixit, J. (2025). Landslide susceptibility assessment for Uttarakhand, a Himalayan state of India, using multi-criteria decision making, bivariate, and machine learning models. *Geoenvironmental Disasters*, *12*(1), 2.

Chung, C. J. F., & Fabbri, A. G. (2003). Validation of spatial prediction models for landslide hazard mapping. *Natural hazards*, *30*(3), 451-472.

Cruden DM, Varnes, DJ (1996) Landslide types and processes Landslides, investigation and mitigation: special report. 247:36-75.

Dahal, R. K., & Hasegawa, S. (2008). Representative rainfall thresholds for landslides in the Nepal Himalaya. *Geomorphology, 100*(3-4), 429-443.

Das, J., Saha, P., Mitra, R., Alam, A., & Kamruzzaman, M. (2023). GIS-based data-driven bivariate statistical models for landslide susceptibility prediction in Upper Tista Basin, India. *Heliyon*, *9*(5).

Das, S., Sarkar, S., & Kanungo, D. P. (2023). A critical review on landslide susceptibility zonation: recent trends, techniques, and practices in Indian Himalaya. *Natural hazards*, *115*(1), 23-72.

Dormann CF, Elith J, Bacher S, Buchman C, Carl G, Carré G, Lautenbac S (2013) Collinearity: A review of methods to deal with it and a simulation study evaluating their performance.Ecography 36(1):27-46.

Fell, R., Corominas, J., Bonnard, C., Cascini, L., Leroi, E., Savage, W. Z., & JTC-1 Joint Technical Committee on Landslides and Engineered Slopes. (2008). Guidelines for landslide susceptibility, hazard and risk zoning for land use planning. *Engineering geology*, *102*(3-4), 85-98.

Gokceoglu, C., & Aksoy, H. (1996). Landslide susceptibility mapping of the slopes in the Gölbaşı area (Turkey) by deterministic stability analyses and image processing techniques. Engineering Geology, 44(3-4), 147-161.

Gorsevski, P. V., Gessler, P. E., Foltz, R. B., & Elliot, W. J. (2006). Spatial prediction of landslide hazard using logistic regression and ROC analysis. *Transactions in GIS*, *10*(3), 395-415.

Gupta, R. P., & Joshi, B. C. (1990). Landslide hazard zoning using the GIS approach—a case study from the Ramganga catchment, Himalayas. *Engineering Geology, 28*(1–2), 119–131.

Guzzetti, F., Mondini, A. C., Cardinali, M., Fiorucci, F., Santangelo, M., & Chang, K. T. (2012). Landslide inventory maps: New tools for an old problem. *Earth-Science Reviews*, *112*(1-2), 42-66.

Guzzetti, F., Reichenbach, P., Ardizzone, F., Cardinali, M., & Galli, M. (2006). Estimating the quality of landslide susceptibility models. *Geomorphology*, *81*(1-2), 166-184.

Himachal Pradesh State Disaster Management Authority. (2023). Cumulative loss report – Monsoon season2023.https://imgs.mongabay.com/wp-content/uploads/sites/30/2023/09/25212410/Himachal-Cumulative-Loss-Report-Monsoon-Season.pdf

Jamwal, A., Kanwar, N., & Kuniyal, J. C. (2020). Use of geographic information system for the vulnerability assessment of landscape in upper Satluj basin of district Kinnaur, Himachal Pradesh, India. *Geology, Ecology, and Landscapes*, *4*(3), 171-186.

Joshi, B. R., Bhandary, N. P., Acharya, I. P., KC, N., & Bhandari, C. (2025). Landslide susceptibility mapping optimization for improved risk assessment using multicollinearity analysis and machine learning technique. *Applied Sciences*, *15*(22), 12152.

Klose, M., Maurischat, P., & Damm, B. (2016). Landslide impacts in Germany: A historical and socioeconomic perspective. *Landslides*, *13*(1), 183-199.

Kumar, V., Gupta, V., Jamir, I., & Chattoraj, S. L. (2019). Evaluation of potential landslide damming: Case study of Urni landslide, Kinnaur, Satluj valley, India. Geoscience Frontiers, 10(2), 753–767.

Kumar, V., Jamir, I., Gupta, V., & Bhasin, R. K. (2021). Inferring potential landslide damming using slope stability, geomorphic constraints, and run-out analysis: a case study from the NW Himalaya. *Earth Surface Dynamics*, *9*(2), 351-377.

Kuniyal, J. C., Jamwal, A., Kanwar, N., Chand, B., Kumar, K., & Dhyani, P. P. (2019). Vulnerability assessment of the Satluj catchment for sustainable development of hydroelectric projects in the northwestern Himalaya. *Journal of Mountain Science*, *16*(12), 2714-2738.

Lee, S. (2005). Application of logistic regression model and its validation for landslide susceptibility mapping using GIS and remote sensing data. *International Journal of Remote Sensing, 26*(7), 1477-1491.

Lee, S., & Evangelista, D. G. (2006). Earthquake-induced landslide-susceptibility mapping using an artificial neural network. *Natural Hazards and Earth System Sciences*, *6*(5), 687-695.

Lee, S., & Pradhan, B. (2007). Landslide hazard mapping at Selangor, Malaysia using frequency ratio and logistic regression models. *Landslides*, *4*(1), 33-41.

Liu, M., Xu, B., Li, Z., Mao, W., Zhu, Y., Hou, J., & Liu, W. (2023). Landslide susceptibility zoning in Yunnan Province based on SBAS-InSAR technology and a random forest model. *Remote Sensing*, *15*(11), 2864.

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.

Moore, I. D., Grayson, R. B., & Ladson, A. R. (1991). Digital terrain modeling: A review of hydrological, geomorphological, and biological applications. *Hydrological Processes, 5*(1), 3-30.

National Institute of Disaster Management. (2013). Uttarakhand disaster June 2013.https://nidm.gov.in/pdf/pubs/ukd-p1.pdf

National Remote Sensing Centre. (2023). Landslide atlas of India (Mapping, monitoring and R&D studies using remote sensing data). Indian Space Research Organisation.https://ndem.nrsc.gov.in/documents/downloads/Landslide%20Atlas_2023.pdf

Niyokwiringirwa, P., Gall, T., & Jha, A. K. (2026). *Toward a Standard for Landslide Data: Bridging Gaps in Landslide Susceptibility Modeling and Early Warning Systems* (No. 11324). The World Bank.

O'Brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. *Quality & Quantity*, 41(5), 673–690.

Park, S., & Kim, J. (2019). Landslide susceptibility mapping based on random forest and boosted regression tree models, and a comparison of their performance. *Applied Sciences*, *9*(5), 942.

Petley, D. (2012). Global patterns of loss of life from landslides. *Geology*, *40*(10), 927-930.

Pradhan, B., & Lee, S. (2010). Landslide susceptibility assessment and factor effect analysis: Back-propagation artificial neural networks and their comparison with frequency ratio and bivariate logistic regression modeling. *Environmental Modelling & Software, 25*(6), 747-759.

Pradhan, B., & Lee, S. (2010). Landslide susceptibility assessment and factor effect analysis: Back-propagation artificial neural networks and their comparison with frequency ratio and bivariate logistic regression modeling. *Environmental Modelling & Software, 25*(6), 747-759.

Pradhan, B., & Lee, S. (2010). Regional landslide susceptibility analysis using back-propagation neural network model at Cameron Highland, Malaysia. *Landslides*, 7(1), 13–30.

Saha, A. K., Gupta, R. P., & Arora, M. K. (2002). GIS-based landslide hazard zonation in the Bhagirathi (Ganga) valley, Himalayas. *International journal of remote sensing*, *23*(2), 357-369.

Saha, A. K., Gupta, R. P., Sarkar, I., Arora, M. K., & Csaplovics, E. (2002). An approach for GIS-based statistical landslide susceptibility zonation — with a case study in the Himalayas. *Landslides*, 1(4), 265–273.

Shano, L., Raghuvanshi, T. K., & Meten, M. (2020). Landslide susceptibility evaluation and hazard zonation techniques–a review. *Geoenvironmental Disasters*, *7*(1), 18.

Singh, M., Khajuria, V., Singh, S., & Singh, K. (2024). Landslide susceptibility evaluation in the Beas River Basin of North-Western Himalaya: A geospatial analysis employing the Analytical Hierarchy Process (AHP) method. *Quaternary Science Advances*, *14*, 100180.

Taalab, K., Cheng, T., & Zhang, Y. (2018). Mapping landslide susceptibility and types using Random Forest. *Big Earth Data*, *2*(2), 159-178.

Varnes, D. J. (1984). *Landslide hazard zonation: a review of principles and practice*. Paris: UNESCO.

Wang, Y., Sun, D., Wen, H., Zhang, H., & Zhang, F. (2020). Comparison of random forest model and frequency ratio model for landslide susceptibility mapping (LSM) in Yunyang County (Chongqing, China). *International journal of environmental research and public health*, *17*(12), 4206.

Wieczorek, G., Mandrone, G., & DeCola, L. (1997). The influence of hillslope on debris-flow initiation. In *Debris-Flow hazard mitigation: mechanics, prediction and assessment* 21-31. ASCE, USA.

Yi, Y., Zhang, Z., Zhang, W., Jia, H., & Zhang, J. (2020). Landslide susceptibility mapping using multiscale sampling strategy and convolutional neural network: A case study in Jiuzhaigou region. *Catena*, *195*, 104851