

Landslide Susceptibility Assesment in Sikkim using Frequency Ratio and Shannon Entropy

by Ishaan Kumar

Submission date: 08-Jun-2026 02:21PM (UTC+0530)

Submission ID: 2978876226

File name: ISHAAN_major_project.pdf (2.46M)

Word count: 7591

Character count: 38610

CHAPTER 1

INTRODUCTION

1.1. MOTIVATION AND BACKGROUND STUDY

Landslides are normally a common geological hazard which causes loss of life and economic damage. Many landslide catastrophes in the Himalayas have recently greatly impacted civilization. Numerous construction activities, such as road & building construction, are disturbing hill slopes as a result of the growing urbanization. A detailed understanding of the interaction between the inclination and triggering factors that trigger landslides in each location is crucial.

Sikkim is one of the popular tourist destinations. Landslide vulnerability mappings may be utilized as an efficient technique that can offer further efforts in landslide mitigation. A large scale research work is going on with GPS, RS and GIS with various approaches due to upsurge of landslides. There are various types of approaches for creating the vulnerability map such as deterministic approach, heuristic approach, and statistical approach. The heuristic approach is based on opinions from experts that may differ from one person to another, and the map is developed using them. In deterministic approach, a lot of data is required based on field investigation for collecting all the soil data which is not easy to carry for a larger area. In the statistical approach, they incorporate randomness in their approach.

The statistical approach is further of two types, bivariate analysis, and multivariate analysis. A bivariate analysis compares two observations from a single sample or an individual, and is used to explore the relationship between two sets of data. However, none of the samples are mutually dependent. The multivariate analysis examines many factors to determine whether any of these factors predict a particular outcome.

The dependent variable is the outcome and the not dependent variables are the predictive variables. The AUC is mostly used for 2 calculating the model's accuracy. The success rate curve called (SRC) and prediction rate curve generally called (PRC) are plotted with the training and testing points.

1.2. AIM AND SIGNIFICANCE OF THE RESEARCH WORK

1.2.1 AIM

This research work has been carried out to find a relationship between various approaches used in the process of preparing of landslide susceptibility map of Sikkim, India. The dominance of causative factors has also been carried out for the area of research.

1.2.2 SIGNIFICANCE

The findings from this study offer highly practical lessons and applications in various aspects of disaster risk management, infrastructure planning and sustainable development:

In context of Research Community and Academic Institutions: The comparison of FR and Shannon's Entropy methods gives methodologic inputs on merits, demerits and suitable contexts of application of bivariate statistical methods in geologically complex mountain terrain. The variations in performance between the methods can be used to guide model selection for similar environments. This research has established a comprehensive factor database and landslide inventory, which can be used as background conditions for future temporal susceptibility analysis to understand the variation of landslide patterns over time related to the dynamic conditions of urbanization and climate change. The method described in this paper can be modified for other states of the Himalayas, susceptible to landslides.

In context of Urban and Regional Planners: The susceptibility data is used in municipal development plans to identify areas where urban development is not allowed, areas where

urban development is allowed but subject to special approval and engineering measures, and areas where urban development is allowed. Susceptibility-based development control regulations can be a part of master plans for hill towns, with a high density residential or critical facility use being limited in very high susceptibility zones and low intensity uses such as parks, recreational facilities or agriculture allowed. Susceptibility screening is useful during early stages of feasibility for siting of an industrial estate, development of an institutional campus, and planning of a commercial complex.

1.3. OBJECTIVES

The goals and purpose of the research can be listed as:

- To create a unique map for each landslide activating factor ¹³ of the study area.
- To create LVM ⁹ using Frequency Ratio and Shannon's Entropy models.
- Validation of models using the Area under the curve generally called (AUC) ⁹ approach of the Receiver Operating Characteristic called (ROC) ⁹ curve for both ⁹ the training and testing datasets of landslide inventory for both methods.
- Building a relationship between the models prepared on the basis of their Success rate curve and Prediction rate curve.
- Using these two models of Sikkim, calculation of landslide prone areas has been done.

CHAPTER 2

LITERATURE REVIEW

2.1. LITERATURE WORK

1. **Sonker, Irjesh; Tripathi, Jayant Nath; Swarnim (2022).** Sonker et al. have used the Frequency Ratio (FR) model in a GIS to develop the landslide susceptibility map in Sikkim Himalaya. A detailed landslide inventory was developed using satellite images and field data of 688 landslide locations, of which 70% was used in training and 30% in validation. The model demonstrated good predictive power (AUC of ~0.87) and prediction power (AUC of ~0.88) and the FR approach was highly reliable in mountainous areas
2. **Mandal, Biplab; Mandal, Sujit; Maiti, Ramkrishn (2018).** Mandal et al. have investigated the Frequency Ratio (FR) and Information Value (IV) models along the NH-31A corridor in Sikkim Himalaya. There were 287 landslide locations from remote sensing and field verification. The IV model (AUC $\approx$ 0.88) slightly outperformed the FR model (AUC $\approx$ 0.84), which proves the effectiveness of the bivariate statistical approaches in linear infrastructure zones in terms of ROC analysis.
3. **Anbalagan, R.; Chakraborty, D.; Kohli, A.; et al. (20215).** Anbalagan et al. used Frequency Ratio integrated with Fuzzy Logic for landslide hazard zonation in Lachung Valley, North Sikkim. Satellite imagery and extensive field checks were used to prepare the landslide inventory. The validation indicated good spatial correspondence between the high susceptibility zones and the known landslide occurrences without reporting ROC–AUC values.
4. **Sharma, Lalit P.; Patel, Nilanchal; Ghose, M. K.; Debnath, P (2012).** Sharma et

- al. applied Shannon's Entropy method for the evaluation of relative importance of landslide conditioning factors in Sikkim. Twenty-eight mapped landslide events were used in the study and special attention was given to entropy-based weighting for hazard zonation. Spatial agreement analysis was used for validation, with a 90% agreement between high-risk zones and observed landslides.
5. **Kaur, Harjeet; Gupta, Srimanta; Parkash, Surya; Thapa, Raju; Gupta, Arindam; Khanal, G. C.(2019).** Kaur et al. (2019) used a hybrid machine learning and entropy-based approach, such as Random Forest–Index of Entropy (RF-IOE), to evaluate landslide susceptibility in Gangtok, Sikkim. Multiple model performances were assessed by ROC analysis and AUC values were between ~0.67 and ~0.85, demonstrating better prediction accuracy when entropy-based weighting was applied along with machine learning methods.
 6. **Alam, Sk Asraful; Maiti, Ramkrishna; Mandal, Suji (2024).** Alam et al. (2024) have analyzed landslide susceptibility and risk along NH-31A and settlements in East Sikkim using Frequency Ratio and Information Value models, as well as Certainty Factor model. There were 153 landslide locations in the landslide inventory, with 70% in the training dataset and 30% in the validation dataset. The study focused on the susceptibility–risk integration approach.
 7. **Alam, Sk Asraful; et al.** Alam et al. used FR, IV and CF models for a probabilistic landslide risk assessment for the Rorachu watershed in East Sikkim. Model validation with ROC curves revealed good performance with FR having the highest AUC (~0.93), followed by CF (~0.87) and IV (~0.85), indicating the robustness of the bivariate statistical models in the Himalayan terrain.
 8. **Saha, S., Saha, A., Roy, B., Chaudhary, A., & Sarkar, R. (2024).** Implements an Artificial Neural Networks (ANN) and Generalised Linear Models (GLM) ensemble model to develop landslide susceptibility model in the Mirik area of West Bengal, India. A dataset of 373 landslide points and 12 condition factors (LCFs) were analysed, and the selection of the appropriate LCFs was done based on the

multicollinearity tests. A detailed and accurate landslide susceptibility maps (LSMs) have been generated for the region using the ensemble machine learning method.

9. **Pal, Subodh Chandra; Chowdhuri, Indrajit. (2019).** The Frequency Ratio model was used for landslide susceptibility mapping by Pal and Chowdhuri (2019) in the Lachung River basin of North Sikkim. The validation results of the ROC showed that the success rate was ~92%, and the prediction rate was ~89%, which represented the high predictive ability of FR models with the aid of good quality landslide inventories

10. **Mandal, Sujit; Mandal, Kanu (2018).** In Rorachu River Basin, East Sikkim, Mandal and Mandal (2018) employed bivariate model based on Statistical Index (SI) to develop the landslide susceptibility map. Model validation was carried out with the help of ROC analysis, which yielded an accuracy of about 91% indicating the suitability of the statistical index methods in the complex geomorphology of the Himalayas.

11. **Das, Jayanta; Saha, Pritam; Mitra, Rajib; Alam, Asraful; Kamruzaman, Md(2023).** In the Upper Teesta Basin (Darjeeling & Sikkim Himalaya), Das et al. (2023) assessed various bivariate statistical models such as Frequency Ratio, Index of Entropy, etc. A total of 477 landslide locations were used, of which 70% were used for training and 30% for validating the study. Comparative analysis showed that entropy-based models achieved better prediction accuracy, but without giving the exact AUC values in the abstract.

2.2. LITERATURE GAPS

Not a lot of research is carried out in the preparation of landslide vulnerability mapping in the Sikkim, India. The use of statistical bivariate methods for landslide susceptibility mapping has been a great deal of interest in recent years, especially since the introduction of new cutting-edge technology and methodology, but there is a significant publication gap in this field.

There is a wide range of research already undertaken on landslide susceptibility mapping with different methods including statistical, machines learning(ML) and hybrid methods. In order to determine landslide susceptibility, these techniques frequently concentrate on multivariate studies, combining several environmental and topographic aspects. But their potential to be used for landslide susceptibility mapping has not been fully realized. Bivariate methods can be used to provide insights into the relationships between some factors and landslide occurrence which can enhance landslide susceptibility assessments and understanding of landslide processes. In this, 2 GIS based statistical bivariate methods i.e., Frequency Ratio and Shannon's Entropy are used for the first time.

2.3. LANDSLIDE DEFINITION AND TYPES

A broad category of processes involving the movement of slope-forming materials (rock, soil, artificial fill or a combination of given materials) outward and downward is termed a landslide. The materials can be conveyed by toppling, sliding, spreading or flowing.

Table 2.1. Classification of Landslides

Type of movement		Type of material	
		Bedrock	Engineering soils
			Predominantly coarse Predominantly fine
Falls		Rock fall	Debris fall Earth fall
Topples		Rock topple	Debris topple Earth topple
Slides	Rotational	Rock slide	Debris slide Earth slide
	Translational		
Lateral Spreads		Rock spread	Debris spread Earth spread
Flows		Rock flow	Debris flow Earth flow
	(deep creep)		Soil creep
Complex		Combination of two or more principal types of movement	

Various landslides types defined by movement of land are shown in Fig. 2.1.

Although landslides are typically found in hilly areas, they can occur in areas with little relief. In low altitude areas, landslides can occur in many different forms such as cut and fill landslides, river slope landslides, lateral growing landslides, failure of mine-waste piles and a range of slope and failures associated with quarries.

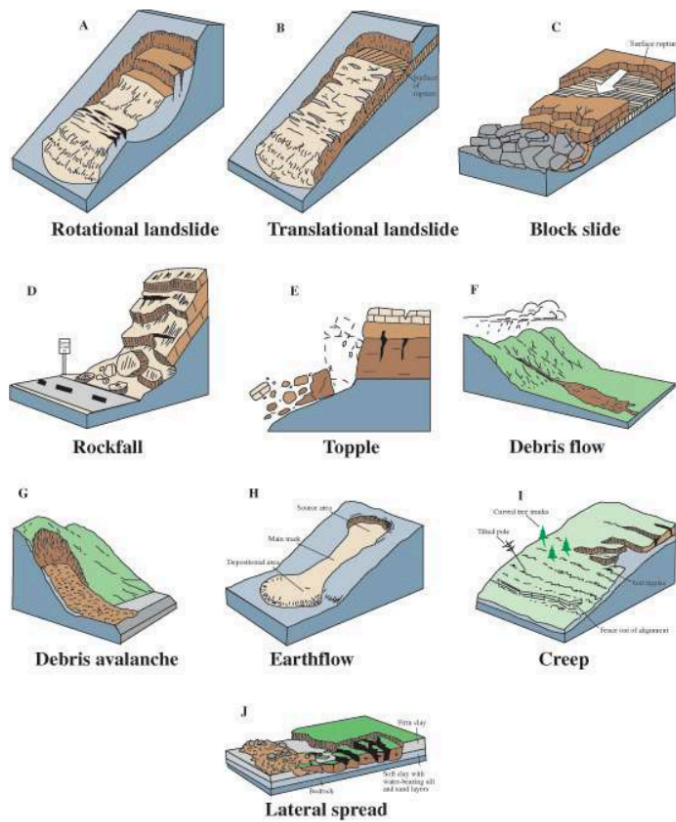


Fig. 2.1. Major types of landslide movements

2.4. REMOTE SENSING (RS) AND GEOGRAPHIC INFORMATION SYSTEM (GIS)

Spatial, spectral, temporal and radiometric resolution of sensor systems and the function of the reconnaissance platform allow us to evaluate the landscape by remote sensing (RS).
The periodicity or return interval of a satellite in its assigned orbit around the Earth is known as temporal resolution. The intensity range used to measure the spectral responses that are measured with the sensor system is called radiometric resolution of the sensor system. The ground area is referred to as spatial resolution. The four remote sensing resolutions work together to characterize the landscape at local, regional, seasonal, annual and global spatial scales, with a range design of radiance intensity to differentiate between landscape features and process factors. In addition, satellites provide a view on earth observation, computer-compatible sensed data, historical view and almost global coverage for landscape studies. Geographic Information System (GIS) technology offers an analytical framework for data synthesis, and integrates a system that can gather, store, administer, retrieve, analyze and display data.

In terms of functionality, GIS methods have the following characteristics:

1. Spatial and non-spatial relation
2. Perspective of landscape representation
3. Information presentation of spatial and temporal pattern
4. Dual occurrence of spatial and non-spatial data
5. Thematic coverage (display): showing the themes of the lesson
6. Modelling of area and response of phenomenon

2.5. LANDSLIDE VULNERABILITY MAPPING

Mapping of landslide susceptibility involves finding the locations that are susceptible to landslides based on a variety of factors, such as topography, rainfall, TWI, land use, and human activity.

The mapping of the landslide susceptibility procedure usually consists of the following steps.

- Collection of data: This involves collating data of the topography, geology, soil characteristics, land use and other factors that could affect the susceptibility of landslides.
- Determination of areas susceptible to landslides, due to the presence of triggering factors such as steep slopes, unstable soils and heavy rains is called hazard assessment.
- Vulnerability Assessment is the analysis of the exposure, sensitivity and adaptability of the population and infrastructure in the indicated danger zones.
- Assessment of Hazard: This process combines vulnerability and hazard analyses to pinpoint the most vulnerable places to landslides.
- The process of creating maps showing the different levels of landslide hazard in the area being mapped is called mapping.
- Communication and planning: This involves informing decision makers and stakeholders of the results of the landslide vulnerability mapping, and developing plans and strategies to reduce the risk of landslides in the identified hazard areas.

2.6. METHODS OF PREPARATION OF VULNERABILITY MAP

The approaches to the creation of the vulnerability map include deterministic approach, heuristic approach and statistical approach

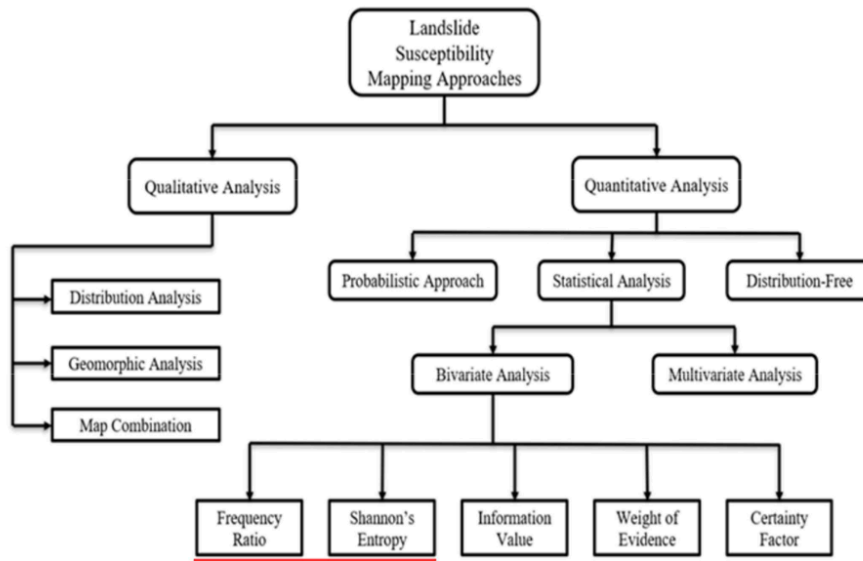


Fig. 2.2. Different approaches for preparation Landslide susceptibility Mapping are shown in a flowchart

CHAPTER 3

STUDY AREA

3.1. SIKKIM

Sikkim is a small but developing state in the eastern Himalayas of India, renowned for its biodiversity, Tibetan influence and Kangchenjunga. Capital Gangtok and area 7,096 km²; rich forests, tea gardens, monasteries, and eco-tourism; borders Nepal, Bhutan, Tibet and West Bengal. Conservation, cultural festivity and sustainable tourism are the main reasons for its popularity with nature lovers from all over the world.

Sikkim has an average of more than 100 landslides every year, which negatively affects communities and infrastructure throughout the state, according to data gathered by the Sikkim State Disaster Management Authority. Scientific data shows that there are approximately 2 landslides per sq.km in North Sikkim and in total, about 2,500 tons of soil losses from the slope failures are estimated to be lost annually per sq.km. The state has been prone to landslide regions of India, with 1561 landslides reported in the last 7 years (2011-2017). These events often occur during the monsoons when heavy rain falls, leading to damage to roads, blockages of key transportation corridors, and disruption to daily life, especially during this time.

Sikkim's National Remote Sensing Center (NRSC) Landslide Atlas of India shows that all 6 districts in Sikkim are susceptible to landslides.

- Gangtok District (East Sikkim) - This is one of the top 10 districts in India with a high risk of landslides
- Namchi District (South Sikkim) - Extremely high risk of landslides

- Mangan District (North Sikkim) - High landslide risk
- Gyalshing District (West Sikkim) - Conditions are moderate to high for landslides in moderate to high terrain
- Pakyong District- Moderate to Low landslide risk
- Soreng District – Low to Moderate landslide risk

Study Area- Sikkim

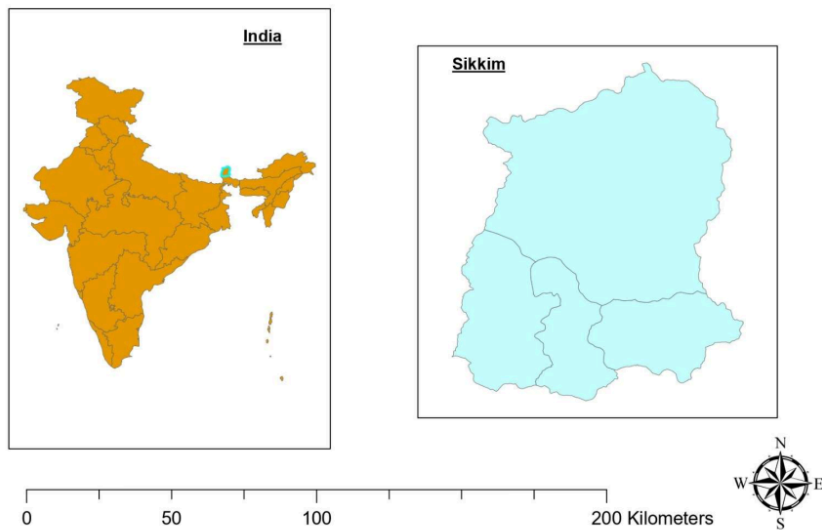


Fig 3.1: Study Area Map - Sikkim

CHAPTER 4

METHODOLOGY

4.1. METHODOLOGY ADOPTED

A landslide vulnerability zonation map will be created using 2 different models and 10 factors. A landslide vulnerability map detects landslide-prone locations and ranks them from very low to very high. The landslide vulnerability map takes into account the location of landslides and the cause of the landslides. We have divided our vulnerability map into 5 classes based on the natural break in ArcGIS which is very low, low, moderate, high and very high. After the preparation of the vulnerability map, it is very essential to validate it. For validation purposes, we are using the AUC method of ROC. The SRC and PRC are computed for each model.

The flow representing the study methodology is shown in Fig. 4.1.

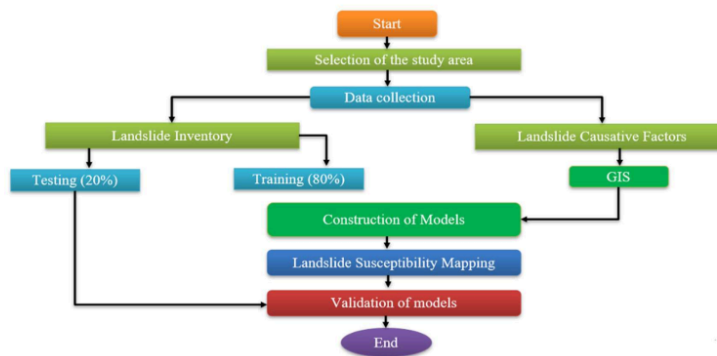


Fig. 4.1. Adopted method is shown in the following flow chart

CHAPTER 5

DATA SOURCE

TABLE 5.1 DATA SOURCE

MAPS	DATA SOURCE
MAP OF INDIA	“https://onlinemaps.surveyofindia.gov.in/Digital_Product_Show.aspx”
LANDSLIDE INVENTORY POINTS	“https://bhukosh.gsi.gov.in/Bhukosh/Public” “https://svs.gsfc.nasa.gov/4710”
STATE MAP OF SIKKIM	“https://esriindia1.maps.arcgis.com/home/item.html?id=b89de19cafb94ea38552a55eb5b2d13d”
DEM	“https://portal.opentopography.org/datasets”
CURVATURE, HILLSHADE, ELEVATION	“https://earthexplorer.usgs.gov/”
SLOPE, ASPECT, ROUGHNESS, TWI	“https://opentopography.org/”

CHAPTER 6

LANDSLIDE INVENTORY MAP

Landslide Inventory Map (LIM) documents previous landslides and pertinent topographic, geologic and meteorological information. It is an important tool for determining the relationship of landslide occurrence and conditioning factors (CFs) and for susceptibility mapping and prediction. Sixty seven hundred and seventy (677) landslide points were extracted for the Sikkim. Following common research practices, 80% of the data were used for model training and 20% for testing. These inventory points are used to develop landslide susceptibility models and are the basis for identifying high-risk areas by relating historic landslides to environmental variables.

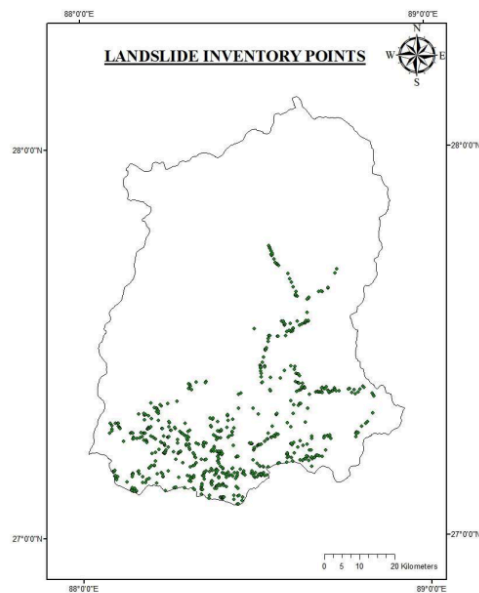


Figure 6.1 Landslide Inventory Points

CHAPTER 7

LANDSLIDE CONDITIONING FACTORS (LCF)

7.1 SLOPE:

Slope is generally one of the most crucial landslide conditioning factors (LCF), and it is the steepness or inclination of a surface which is important in determining the stability of the slope. Slopes tend to be more susceptible to landslides on steeper slopes, because the gravitational force on the surface materials is greater. The slope map was created in Sikkim from Digital Elevation Model (DEM) data using spatial analyst tool in ArcGIS. The map was divided into five slope classes, which allowed for a detailed analysis of the effect of different slopes on the susceptibility to landslides. It is important to understand the variation of slopes because it can help identify areas of greater risk for slope failure, which can help with effective landslide risk assessment and management.

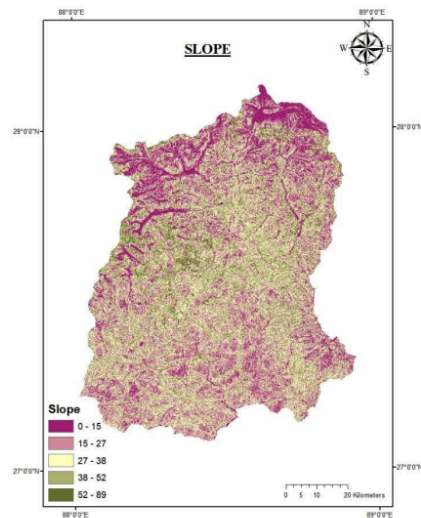


Figure 7.1 SLOPE

7.2 ASPECT:

Aspect is a major key Landslide Conditioning Factors (LCF) that is the direction in which the slope faces, usually in relation to the cardinal directions. It has a significant impact on the environmental factors like sunlight, wind, and moisture which influence the stability of slopes. During the study of Sikkim, aspect mapping was done with the help of Digital Elevation Model (DEM) data and processed using ArcGIS. Changes in aspect may affect the sensitivity of slopes to landsliding; some aspects may have more moisture or be subject to climatic influences that make the slope less stable. This analysis contributes to a more accurate landslide susceptibility mapping and disaster management strategy, as it helps identify areas more susceptible to landslides, depending on the slope orientation.

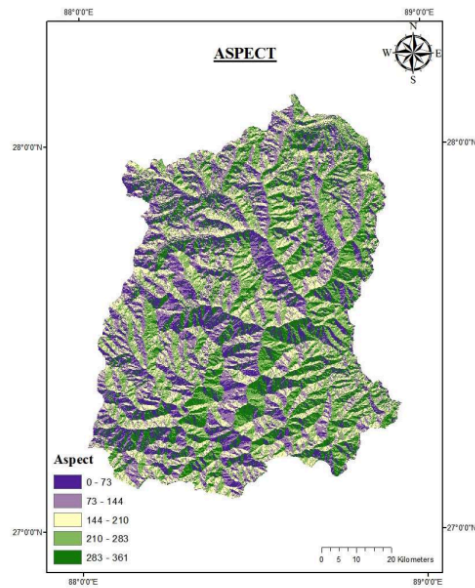


Figure 7.2 ASPECT

7.3 ROUGHNESS:

Roughness is a significant Landslide Conditioning Factor (LCF) that describes the irregularity or unevenness of the terrain surface, and affecting the stability of a slope and the potential for landslides. It is a very important factor in determining water retention, soil movement and mechanical stability of slopes. Surface roughness also tends to increase where it is higher, which can mean that the surface is not even, resulting in concentrated runoff, decreased vegetation stability, and higher risk of rock fractures, and thus higher risk for landsliding. For Sikkim state, the roughness was estimated using Digital Elevation Model (DEM) data in ArcGIS. Roughness analysis can be used to identify areas more prone to slope failures, which can be very useful for the production of accurate landslide susceptibility maps and the development of effective mitigation measures.

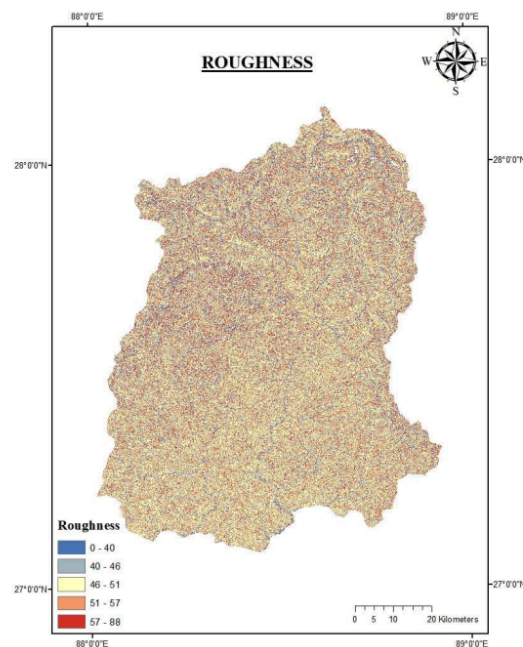


Figure 7.3 ROUGHNESS

7.4 CURVATURE:

Curvature is an important Landslides Conditioning Factor (LCF) that indicates the curvature of the topography and their effect on the flow of water and soil movement. It is based on Digital Elevation Model (DEM) data and divided into concave, convex and planar surfaces. Concave areas will retain water, making the soil more saturated and thus more susceptible to slope failure, whereas convex areas will encourage water to pass over and away from the area, decreasing the likelihood of water accumulating. The curvature function was used in ArcGIS to evaluate the curvature of the terrain in the context of Sikkim state, which would give some idea of the influence of terrain morphology on landslide susceptibility. This is one of the factors used to determine the areas that are susceptible to water concentration or erosion, which helps to improve the estimation of landslide susceptibility and the planning of mitigation measures.

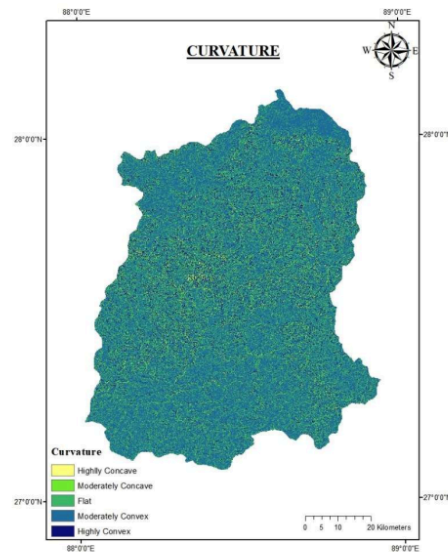


Figure 7.4 CURVATURE

7.5 ELEVATION:

Elevation is one of the most basic Landslide Conditioning Factors (LCFs) that greatly affects the stability of terrain and the susceptibility of the terrain to landslides. It affects various environmental and geological processes such as the slope steepness, vegetation distribution, weathering rates and patterns of water flow. Slopes on higher elevations tend to be steeper and are subject to more erosion, thus are more likely to fail, while slopes on lower elevations may collect debris and retain more water, causing slope failure as well. The elevation range in Sikkim is from 250 m to 7500 m above sea level. The variation was analysed with the help of Digital Elevation Model (DEM) data to help identify areas of different landslide susceptibility. Accurate landslide mapping and effective landslide risk mitigation strategies rely on understanding elevation patterns.

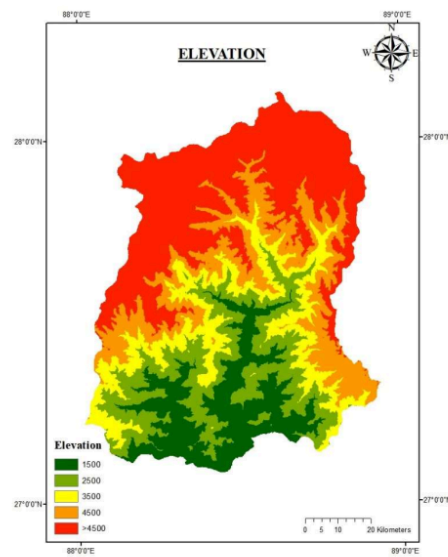


Figure 7.5 ELEVATION

7.6 HILLSHADE:

Hillshade is one of the ³⁴Landslide Conditioning Factors (LCF) that ¹⁹can be used to simulate the interaction between the sun and the terrain, allowing the features on the terrain to be highlighted by the light and shadow. It offers a thorough representation of the orientation and gradient of slopes, a key component in comprehending terrain stability. The shaded areas tend to be moister with higher soil saturation and may contribute to loss of slope stability, whereas the sunlit areas are typically drier and will not favor landsliding. Digital Elevation Model (DEM) data was used to implement the hill shade analysis in ArcGIS to delineate areas with different moisture holding capacity and exposure in the study of Sikkim. Introducing hillshade into landslide susceptibility mapping can help to better understand the influence of solar radiation, moisture distribution and slope stability, which can lead to better landslide risk analyzing.

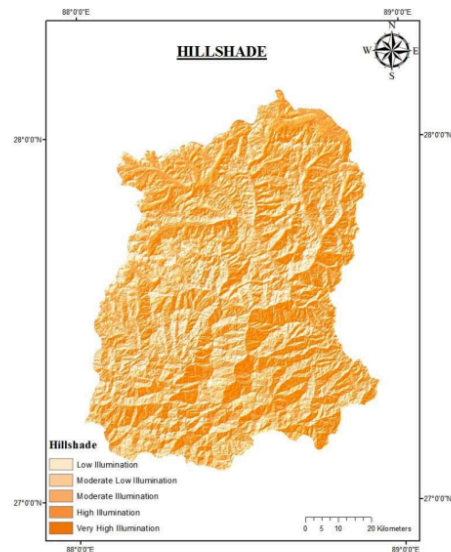


Figure 7.6 HILLSHADE

7.7 TOPOGRAPHIC WETNESS INDEX (TWI):

Topographic Wetness Index (TWI) generally is a spatial index to quantify the relative wetness of a landscape in relation to its topographic characteristics, especially in terms of water accumulation in an area. It is a product of the integration of a number of topographic parameters, such as elevation, slope, and contributing area (or drainage area), which are used to describe the flow and accumulation of water over the landscape. The areas with higher TWI values are generally more susceptible to water saturation and will tend to have more moisture available for landslides because of their topography. Conversely, areas with low TWI values have improved drainage, which means that there is less water retention and a lower risk of landslide hazards arising from water saturation. TWI is useful to gain an understanding and map the vulnerability of landscapes to landslides and other water related hazards.

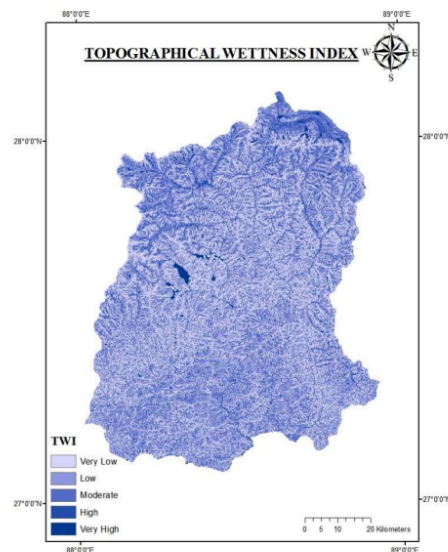


Figure 7.7 TOPOGRAPHIC WETNESS INDEX

7.8 NORMALIZED DIFFERENCE VEGETATION INDEX (NDVI); ¹⁶

Normalized Difference Vegetation Index (NDVI) is a spatial index that reflects relative density and health of vegetation cover on a landscape based on the spectral reflectance of plant leaves. Satellite imagery was used for ArcGIS to calculate NDVI for Sikkim and classify it into five categories: -0.277 to 0.010 (snow/water), 0.010 – 0.092 (barren land/Built-up), 0.092 – 0.183 (light vegetation), 0.183 – 0.280 (moderate vegetation), and 0.280 – 0.522 (dense vegetation). The statistical analysis showed that the barren and lightly vegetated classes (NDVI 0.092 – 0.183) have the highest positive weights for both Frequency Ratio and Shannon's Entropy models, thus indicating that vegetation loss is a main factor of the slope failure in the state.

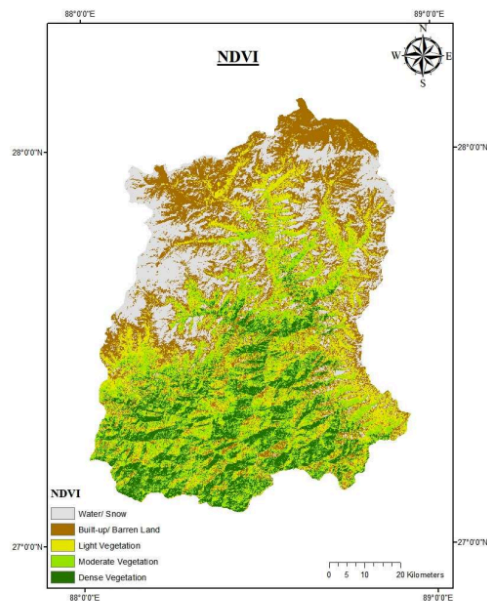


Figure 7.8: NDVI
25

7.9 DISTANCE TO STREAM:

¹⁷Distance to stream is one of the most important Landslide Conditioning Factors (LCF) which is the horizontal length from a point on a slope to the closest drainage channel or water body. For this study, the distance to stream map for Sikkim was prepared using Euclidean's Distances tool in ArcGIS from the hydroSHEDS derived stream network. The factor was classified into five buffer zones: 0 to 250 m, 250 to 500 m, 500 750 m, 750 to 1000 m, and greater than 1000 m. The statistical analysis revealed that the highest frequency of landslides is associated with the zone 0-250 m from stream channels and according to both the FR & Shannon's Entropy models, the proximity to the stream channels had the highest positive weight, indicating that this is a major influencing factor of slope instability in Sikkim.

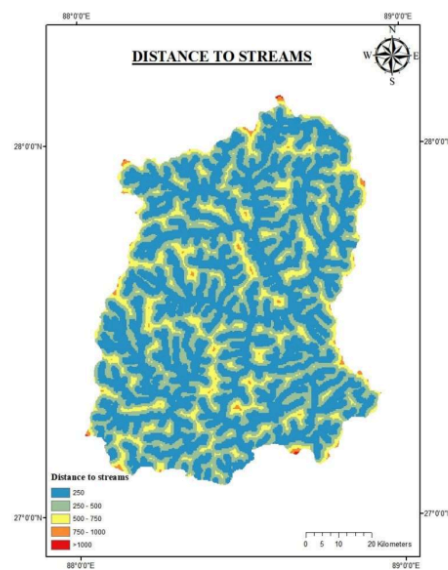


Figure 7.8: Distance to stream

7.10 RAINFALL:

Rainfall is one of the important conditioning factors used to assess the susceptibility of Sikkim to landslides, which quantifies the amount and intensity of rainfall. It enters the soil, increases pore water pressure, and reduces slopes' shear strength, which greatly helps the initiation of landslides. Due to increased soil saturation and decreased slope stability, areas that receive more rainfall or have strong rainstorm events are more likely to experience landslides. It is possible to identify places with higher rainfall volumes or intensities that are more susceptible to the occurrence of landslides by including rainfall data into landslide susceptibility mapping. This will aid in hazard assessment and mitigation of landslides.

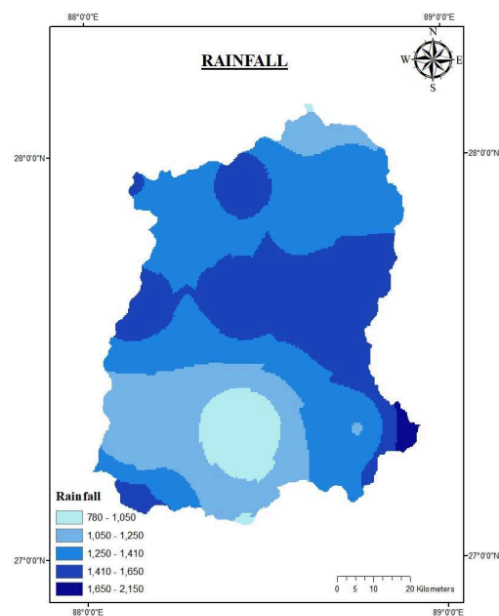


Figure 7.10: Rainfall

CHAPTER 8

SHANNON ENTROPY

8.1 Introduction

Shannon Entropy is a basic principle in information theory which is used to quantify the uncertainty or randomness of a system. It is a measure of the information needed to describe a random variable, and is one of the fundamental instruments for understanding the uncertainty of a system. Shannon Entropy is used in spatial analysis and landslide susceptibility mapping (LSM) to assess the relative importance of different factors like slope, elevation and lithology to landslide occurrence. It can be used to quantify the entropy of these factors to determine the contribution of each factor to the spatial distribution of landslides and make more accurate and reliable predictions of landslide prone areas. This method helps to minimize uncertainty and enhance landslide risk knowledge.

8.2 Formula

$$\%FR P_{ij} = \frac{\% \text{Landslide pixels}}{\% \text{Class pixels}}$$

Where P_{ij} from equation represents frequency ratio and (P_{ij}) from equation gives probabilistic density values of every class

$$(P_{ij}) = \frac{(P_{ij})}{(\sum_1^n P_{ij})}$$

E_{ij} and $E_{ij\max}$ from equation and shows the entropy values for every class whereas n_{ij} is the numbers of class in every factor.

$$E_{ij} = \sum_{i=1}^{n_j} (P_{ij}) * \log(P_{ij})$$

$$E_{ij\max} = \log_2(n_{ij})$$

The information coefficient known as I_j , and the final weights index, W_j is determined from the equations shown below.

$$I_j = \frac{H_{j\max} - H_j}{H_{j\max}}$$

$$W_j = I_j \times P_j$$

8.3 Outcome

The results of making use of Shannon Entropy to landslide susceptibility mapping are normalized entropy values that represent weights that reflect the influence of each landscape classification factor (LCF) on landslide susceptibility. The normalized entropy values represent the degree to which each factor (e.g., slope, aspect, or curvature) is associated with the probability of landslides in a particular area. These weights, together with the various LCFs, form a more detailed and precise landslide susceptibility map. This map shows the areas of different risk levels, which is a useful tool to help identify areas at higher risk for landslides and to help with risk management and mitigation.

8.4 Advantages

Shannon Entropy has been used to benefit landslide susceptibility mapping (LSM) in several ways. It offers an objective and data-based way of quantifying the effect of each landslide conditioning factor, which decreases subjectivity in the weighting process. The

method is especially effective in dealing with uncertainty and variability in spatial data, and is appropriate for complex terrain such as the Sikkim. It also has good compatibility with GIS software, enabling spatial analysis and mapping capabilities. Furthermore, Shannon Entropy can be used even if there is not much ground truth data available, because it is based on the statistical relations between the occurrence of landslides and the environmental factors. It creates normalized weights, making susceptibility maps more accurate and reliable, which aids in risk assessment and management.

8.5 Applications

Landslide susceptibility mapping uses a landslide inventory and a set of landslide conditioning factors (LCFs) including slope, elevation, aspect and lithology to calculate Shannon Entropy. The landslide inventory includes historical information on the landslide occurrences, and the conditioning factors are the environmental and topographical conditions that affect landslide risks. These elements are combined to quantify the uncertainty and relative importance of each of the factors that determines landslide susceptibility in a study area, Shannon Entropy. This method is also applied to the evaluation of other natural hazards like flood, soil erosion and environmental hazards other than landslides. In these applications, Shannon Entropy is used to help develop risk maps to identify areas at risk from these hazards, thereby assisting in effective disaster management and mitigation planning.

TABLE 8.1: SHANNON ENTROPY CALCULATION

Factor (LCF)	Class / Range	FR Value (Pij)	Pij / ΣPij (Prob. Density)	ln(Pij/ΣPij)	(Pij/ΣPij) × ln(Pij/ΣPij)	Ej (Entropy)	Ej_max = ln(nij)	Ij = 1 - (Ej/Ej_max) (Info. Coeff.)	Wj = Ij × ΣPij (Weight Index)
SLOPE (°)	< 10°	0.199	0.0302	-3.5000	-0.1057	1.3494	1.6094	0.1616	1.0649
	10–20°	0.498	0.0756	-2.5827	-0.1952				
	20–30°	1.253	0.1901	-1.6600	-0.3156				
	30–40°	2.503	0.3798	-0.9681	-0.3677				
	> 40°	2.137	0.3243	-1.1262	-0.3652				
ASPECT	Flat	0.920	0.0920	-2.3855	-0.2196	2.1668	2.1972	0.0138	0.1384
	North	0.702	0.0702	-2.6559	-0.1865				
	NE	0.906	0.0906	-2.4008	-0.2176				
	East	1.246	0.1247	-2.0821	-0.2596				
	SE	1.628	0.1629	-1.8147	-0.2956				
	South	1.444	0.1445	-1.9347	-0.2795				
	SW	1.200	0.1201	-2.1198	-0.2545				
	West	0.991	0.0991	-2.3111	-0.2291				
	NW	0.958	0.0958	-2.3450	-0.2248				
ROUGHNESS	Very Low	0.328	0.0640	-2.7493	-0.1759	1.4921	1.6094	0.0729	0.3736
	Low	0.679	0.1324	-2.0217	-0.2677				
	Moderate	1.073	0.2093	-1.5641	-0.3273				
	High	1.516	0.2957	-1.2184	-0.3603				
	Very High	1.531	0.2986	-1.2086	-0.3609				
CURVATURE	Concave	1.543	0.4710	-0.7529	-0.3546	1.0567	1.0986	0.0381	0.1249
	Planar	0.803	0.2451	-1.4060	-0.3446				
	Convex	0.930	0.2839	-1.2592	-0.3575				
ELEVATION (m)	250–1000	0.904	0.1915	-1.6529	-0.3165	1.5017	1.6094	0.0670	0.3161
	1000–2000	1.321	0.2798	-1.2736	-0.3564				
	2000–3000	1.385	0.2934	-1.2263	-0.3598				
	3000–4500	0.849	0.1798	-1.7157	-0.3085				
	4500–7500	0.262	0.0555	-2.8914	-0.1605				

HILLSHADE	Very Low (0–50)	0.693	0.1448	-1.9324	-0.2798	1.5743	1.6094	0.0218	0.1046
	Low (50–100)	1.212	0.2532	-1.3734	-0.3478				
	Moderate (100–150)	1.192	0.2491	-1.3901	-0.3462				
	High (150–200)	1.061	0.2217	-1.5065	-0.3340				
	Very High (>200)	0.628	0.1312	-2.0309	-0.2665				
TWI	< 4	0.383	0.0657	-2.7226	-0.1789	1.4691	1.6094	0.0872	0.5084
	4–6	0.639	0.1096	-2.2107	-0.2423				
	6–8	1.317	0.2259	-1.4875	-0.3361				
	8–10	2.069	0.3549	-1.0358	-0.3676				
	> 10	1.421	0.2438	-1.4115	-0.3441				
NDVI	Snow/Water	0.184	0.0274	-3.5979	-0.0985	1.2709	1.6094	0.2104	1.4136
	Barren/Built-up	2.629	0.3912	-0.9385	-0.3672				
	Light Veg.	2.491	0.3707	-0.9924	-0.3679				
	Mod. Veg.	1.114	0.1658	-1.7971	-0.2979				
	Dense Veg.	0.302	0.0449	-3.1024	-0.1394				
DIST. TO STREAM (m)	0–250	2.371	0.4861	-0.7214	-0.3507	1.2992	1.6094	0.1927	0.9402
	250–500	1.317	0.2700	-1.3094	-0.3535				
	500–750	0.545	0.1117	-2.1917	-0.2449				
	750–1000	0.440	0.0902	-2.4057	-0.2170				
	> 1000	0.205	0.0420	-3.1695	-0.1332				
RAINFALL (mm)	< 1500	0.498	0.1100	-2.2072	-0.2428	1.5458	1.6094	0.0396	0.1791
	1500–2000	0.726	0.1604	-1.8303	-0.2935				
	2000–2500	1.114	0.2461	-1.4021	-0.3450				
	2500–3000	1.423	0.3143	-1.1573	-0.3638				
	> 3000	0.766	0.1692	-1.7766	-0.3006				

CHAPTER 9

FREQUENCY RATIO

9.1 Introduction

Frequency Ratio is a well known empirical method to evaluate the relationship between occurrence of landslides and conditioning factors.

Estimates a probability of landsliding given the wide distribution of each given factor class.

9.2 Formula

The ratio of landslide pixels in the Factor Class to the total pixels in the Factor Class divided by the ratio of the total pixels in the Factor Class to the total pixels in the Study Area is the FR. The ratio obtained is the strength of the correlation between the factor class and the occurrence of landslides

$$FR = (\text{Landslide Pixels in Factor Class} / \text{Total Landslide Pixels}) \div (\text{Total Pixels in Factor Class} / \text{Total Pixels in Study Area})$$

The resulting ratio represents the correlation strength between the factor class and landslide occurrence.

$$FR = \left(\frac{P_L}{\sum_{i=1}^n P_L} \right) / \left(\frac{P_C}{\sum_{i=1}^n P_C} \right)$$

where PC = Pixels of a factor class, PL = Landslide pixels contained in a factor class. (8.1)
Categorize landslide conditioning factors (e.g., slope, TWI, elevation) into classes. Find the FR value for each class of the conditioning factors. Determine the weights of the classes according to FR values.

9.3 Outcome

Produces weighted maps for each conditioning factor.

Assists in the development of a combined landslide susceptibility map by adding up weighted factors.

9.4 Applications

The Frequency Ratio (FR) method is one of the most popular methods for making landslide susceptibility maps because it is data driven. It compares the frequency of landslides with Landslide Conditioning Factors (LCFs) by determining the likelihood of landslide occurrence in each of the LCF classes. This approach can be used to measure the contribution of each factor and give it a weight. In geospatial analysis, GIS technology is often employed to create landslide susceptibility maps, with FR being a commonly used method. The method can be used to zone landslide-prone areas with good accuracy by combining historical landslide data with environmental and topographical parameters. These susceptibility maps help to prepare for disaster, plan infrastructure and assess disasters; they help policy makers to make informed decisions about disaster mitigation strategies.

9.5 Advantages of Frequency Ratio

The Frequency Ratio method is one of the major benefits because of its simplicity and efficiency in analysing landslide susceptibility. It is based on historical data, so it does not depend on subjective expert opinions for results or conclusions, but is objective and quantifiable. Furthermore, it can be easily applied using GIS software, so it is preferred for regional scale susceptibility assessment.

One more major benefit is its versatility, which can be utilized on different surfaces and in various scenarios. The method also enables it to be compared with other models to assess the effectiveness of various susceptibility assessment techniques used by researchers and planners.

TABLE 9.1: FREQUENCY RATIO CALCULATION

Factor (LCF)	Class / Range	No. of Pixels in Class (Np_c)	% of Total Pixels (Tp_c%)	No. of Landslide Pixels in Class (Np_l)	% of Total Landslide Pixels (Tp_l%)	FR Value
SLOPE (°)	< 10	12,15,000	24.99	27	4.98	0.20
	10 – 20	14,58,000	29.99	81	14.94	0.50
	20 – 30	11,66,400	23.99	163	30.07	1.25
	30 – 40	6,80,400	13.99	190	35.05	2.51
	> 40	3,40,200	6.99	81	14.94	2.14
ASPECT	Flat (–1)	48,600	1.00	5	0.92	0.92
	North (0–22.5, 337.5–360)	4,85,946	9.99	38	7.01	0.70
	NE (22.5–67.5)	5,34,549	10.99	54	9.96	0.91
	East (67.5–112.5)	5,83,254	11.99	81	14.94	1.25
	SE (112.5–157.5)	5,34,549	10.99	97	17.90	1.63
	South (157.5–202.5)	5,34,549	10.99	86	15.87	1.44
	SW (202.5–247.5)	4,85,946	9.99	65	11.99	1.20
	West (247.5–292.5)	5,34,549	10.99	59	10.89	0.99
	NW (292.5–337.5)	5,34,549	10.99	57	10.52	0.96
ROUGHNESS	Very Low (1–3)	8,74,800	18.00	32	5.90	0.33
	Low (3–6)	10,69,200	22.00	81	14.94	0.68
	Moderate (6–10)	13,12,200	27.00	157	28.97	1.07
	High (10–15)	11,17,800	23.00	189	34.87	1.52
	Very High (>15)	4,86,000	10.00	83	15.31	1.53
CURVATURE	Concave (<–0.05)	10,69,200	22.00	184	33.95	1.54
	Planar (–0.05–0.05)	24,78,600	51.00	222	40.96	0.80
	Convex (>0.05)	13,12,200	27.00	136	25.09	0.93
ELEVATION (m)	250 – 1000	4,86,000	10.00	49	9.04	0.90
	1000 – 2000	12,14,856	24.99	179	33.03	1.32
	2000 – 3000	12,63,456	25.99	195	35.98	1.38
	3000 – 4500	9,71,928	19.99	92	16.97	0.85

	4500 – 7500	9,23,760	19.00	27	4.98	0.26
HILLSHADE	Very Low (0–50)	3,88,800	8.00	30	5.54	0.69
	Low (50–100)	6,80,400	14.00	92	16.97	1.21
	Moderate (100–150)	12,63,456	26.00	168	30.99	1.19
	High (150–200)	15,55,056	32.00	184	33.95	1.06
	Very High (>200)	9,72,288	20.00	68	12.55	0.63
TWI	< 4 (Low)	12,63,456	26.00	54	9.96	0.38
	4 – 6 (Mod-Low)	13,60,800	28.00	97	17.90	0.64
	6 – 8 (Moderate)	10,69,200	22.00	157	28.97	1.32
	8 – 10 (High)	6,80,400	14.00	157	28.97	2.07
	> 10 (Very High)	4,86,144	10.00	77	14.21	1.42
NDVI	–0.277 to 0.010 (Snow/Water)	2,43,000	5.00	5	0.92	0.18
	0.010–0.092 (Barren/Built-up)	3,88,800	8.00	114	21.03	2.63
	0.092–0.183 (Light Veg.)	6,80,400	14.00	189	34.87	2.49
	0.183–0.280 (Mod. Veg.)	12,63,456	26.00	157	28.97	1.11
	0.280–0.522 (Dense Veg.)	22,84,344	47.00	77	14.21	0.30
DISTANCE TO STREAM (m)	0 – 250	9,72,288	20.00	257	47.42	2.37
	250 – 500	10,69,200	22.00	157	28.97	1.32
	500 – 750	10,69,200	22.00	65	11.99	0.55
	750 – 1000	8,74,800	18.00	43	7.93	0.44
	> 1000	8,74,512	18.00	20	3.69	0.21
RAINFALL (mm)	< 1500	4,86,144	10.00	27	4.98	0.50
	1500 – 2000	7,29,000	15.00	59	10.89	0.73
	2000 – 2500	12,63,456	26.00	157	28.97	1.11
	2500 – 3000	13,60,800	28.00	216	39.85	1.42
	> 3000	9,71,600	20.00	83	15.31	0.77

CHAPTER 10

RESULTS

10.1 DISCUSSIONS

² Landslide Susceptibility Maps were generated using two different methods:

Frequency Ratio (FR), Shannon Entropy (SE). They are used to ²⁸ delineate landslide prone areas based on the analysis of environmental factors like slope, NDVI, aspect and Topographical Wetness Index.

Each technique uses a different method to balance the factors, and therefore yields different susceptibility classifications.

Frequency Ratio (FR): An empirical approach based on data that determines the ⁶ probability of a landslide based on the frequency of landslides in each factor class.

Shannon Entropy (SE): A statistical measure of the randomness of factor distribution and weights based on information gain.

This study gives the comparative analysis of two ⁷ landslide susceptibility mapping techniques: Frequency Ratio (FR), Shannon Entropy (SE). The predictive accuracy and the effectiveness of classification ¹ of landslide prone areas were used to evaluate each model. The Frequency Ratio model had the best accuracy (AUC = 0.760), followed closely by the Shannon Entropy model (AUC = 0.724). The predictive power of both models was found to be good, indicating the effectiveness of data-driven and statistical approaches.

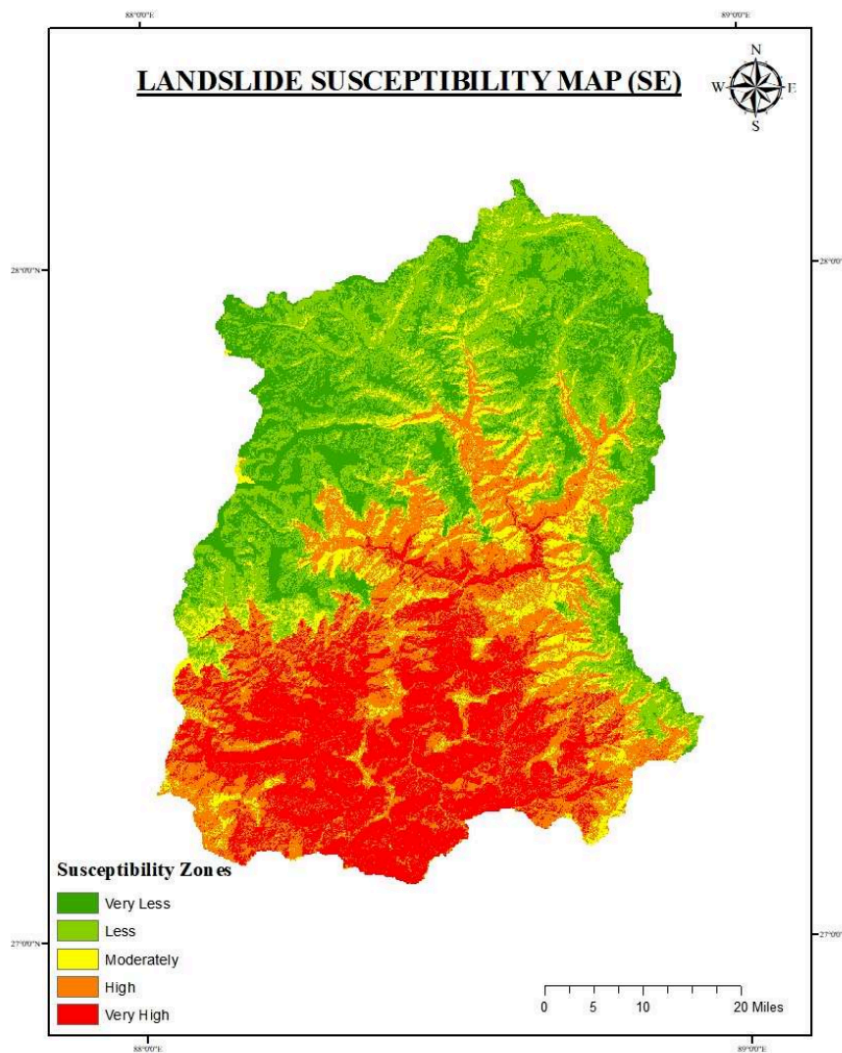


Figure 10.1: Susceptibility Map using Shannon Entropy

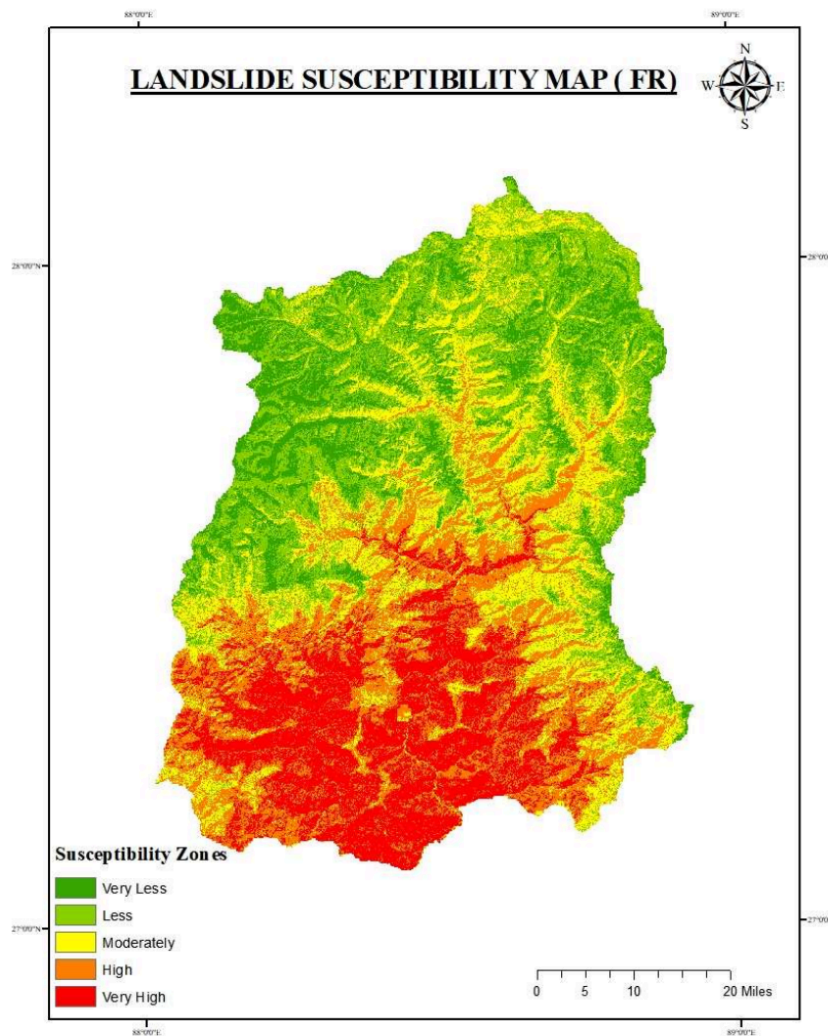


Figure 10.2: Susceptibility Map using Frequency Ratio

10.2 VALIDATION

Validation is an important process to evaluate the performance and reliability of a predictive model. It guarantees the model to be effective on training data and perform well on unseen data. Validation is used to assess the accuracy of the model in landslide susceptibility mapping and other machine learning applications to determine the ability of the model to classify areas as susceptible or non-susceptible. Model effectiveness is evaluated using various validation techniques including k-fold cross validation, hold-out validation and statistical metrics.

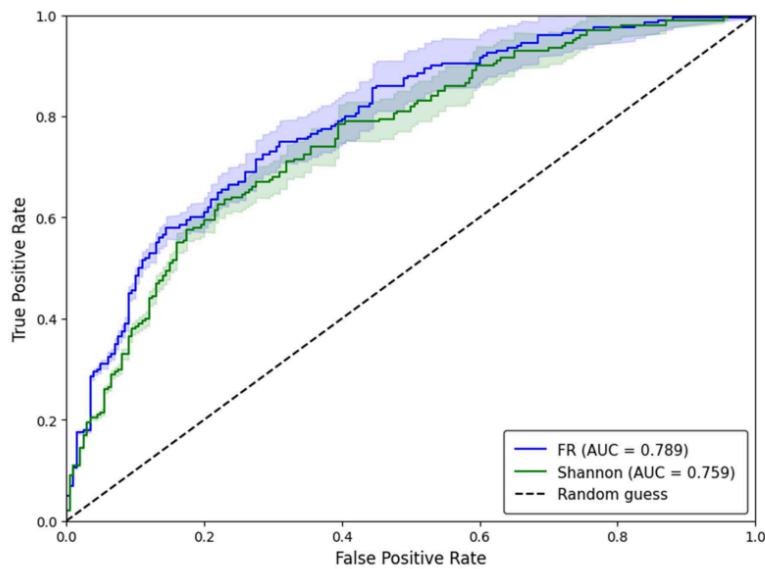


Fig 10.3: Success Rate Curve

A key element of validation is the success rate curve, which compares the accuracy of the model's prediction in terms of landslide prone areas using the training set. It is produced by plotting the cumulative percentage of the correctly predicted landslide areas against the total study area. The higher the success rate, the more the model is able to reflect the pattern of the training data and the better it is calibrated to map susceptibility.

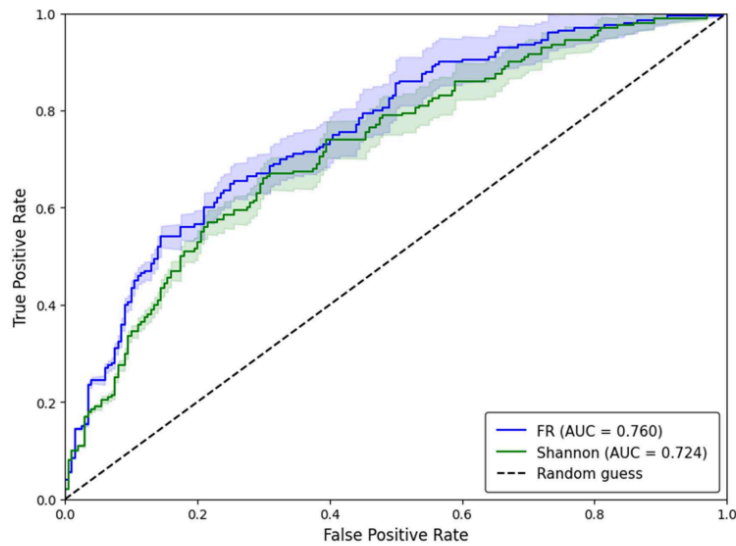


Fig 10.4: Prediction Rate Curve

Likewise, the prediction rate curve is used to measure the model's prediction performance using test data. It comes from independent data and is used to gauge the model's performance on new data. If the prediction rate is high, it means that the model is robust and reliable for practical use, and it has a good generalization ability.

Besides these curves, true positives and true negatives are very important in validation. True positives (TP) are regions that are correctly classified as landslide hazard and true negatives (TN) are regions that are correctly classified as non-landslide. These values are important in determining various performance measures like accuracy, sensitivity, specificity, and area under the curve (AUC). The accuracy of susceptibility mapping will be best achieved by a well-validated model that has a high TP rate and a low false positive and false negative rate.

CHAPTER 11

CONCLUSION AND FUTURE FOCUS

11.1 CONCLUSION

In Sikkim, India, landslide susceptibility mapping was done by two methods, Frequency Ratio (FR), Shannon Entropy (SE). The generated susceptibility maps were assessed by Area under the Curve (AUC) values with the FR model showing the highest predictive accuracy (AUC = 0.760) and the SE model closely following with a value of 0.724, suggesting good predictive ability. In order to further assess the performance of the model, Success Rate and Prediction Rate curves were created, which not only helped in determining the model fit with training data but also in testing the prediction capability with test data.

The results have shown that the statistical-based models (FR and SE) are suitable for predicting landslide prone areas and can be used for hazard assessment. The produced landslide susceptibility maps are used in the disaster risk management, in the planning of infrastructure in high risk areas and in the early warning system, in the field of sustainable land use planning. This study will give scientific basis for decision-making, which will help to minimize risk due to landslides in the mountainous region like Sikkim.

11.2 FUTURE FOCUS

The results indicate that statistical based models (FR and SE) are reliable for landslide hazard assessment with good prediction power in delineating landslide prone areas. The produced landslide susceptibility maps are of great value in disaster risk management, in the planning of infrastructure in high-risk areas and in early warning systems, for sustainable land-use planning. This study gives a scientific basis for decision making, which will help to mitigate the risk associated with landslide hazard in the mountainous region like Sikkim.

Hybrid models (a mix of statistical, analytical, and machine learning models) for more reliable susceptibility mapping are additional developments that can be made. The temporal analysis of landslides can be used to identify temporal changes, and field validation will ensure that models are applicable to the real world. These enhancements will result in more adaptive and data-driven models, which will boost the preparedness and risk mitigation measures for landslide-prone areas.

Landslide Susceptibility Assessment in Sikkim using Frequency Ratio and Shannon Entropy

ORIGINALITY REPORT

9%

SIMILARITY INDEX

3%

INTERNET SOURCES

8%

PUBLICATIONS

0%

STUDENT PAPERS

PRIMARY SOURCES

- | | | |
|---|---|-----|
| 1 | "Geomorphic Risk Reduction Using Geospatial Methods and Tools", Springer Science and Business Media LLC, 2024
Publication | 1% |
| 2 | Pravinkumar, Pathak Yashodhar. "Machine Learning & GIS Based Landslide Susceptibility Modelling for the Part of Himachal Pradesh, India.", Gujarat Technological University
Publication | 1% |
| 3 | collections.mun.ca
Internet Source | 1% |
| 4 | "Impact of Climate Change, Land Use and Land Cover, and Socio-economic Dynamics on Landslides", Springer Science and Business Media LLC, 2022
Publication | 1% |
| 5 | "GIS Landslide", Springer Science and Business Media LLC, 2017
Publication | 1% |
| 6 | Jayanta Das, Pritam Saha, Rajib Mitra, Asraful Alam, Md Kamruzzaman. "GIS-based data-driven bivariate statistical models for landslide susceptibility prediction in Upper Tista Basin, India", Heliyon, 2023
Publication | 1% |
| 7 | www.mdpi.com
Internet Source | <1% |

Landslide Science and Practice, 2013.

8

Publication

<1 %

9

"Trends and Advancements of Image Processing and Its Applications", Springer Science and Business Media LLC, 2022

Publication

<1 %

10

Pooja Rana, Jeganathan Chockalingam, Arvind Chandra Pandey. "chapter 7 Space-Time Integrated Landslide Hazard Zonation near Tehri Dam in Uttarakhand, India", IGI Global, 2017

Publication

<1 %

11

www.researchgate.net

Internet Source

<1 %

12

irc.kdu.ac.lk

Internet Source

<1 %

13

Matebie Meten, Netra Prakash Bhandary, Ryuichi Yatabe. "GIS-based frequency ratio and logistic regression modelling for landslide susceptibility mapping of Debre Sina area in central Ethiopia", Journal of Mountain Science, 2015

Publication

<1 %

14

www.frontiersin.org

Internet Source

<1 %

15

"Applied Geomorphology and Contemporary Issues", Springer Science and Business Media LLC, 2022

Publication

<1 %

16

Asif Iqbal Shah, Jimi Basumatary, Nibedita Das Pan. "GIS and exploratory regression-based flood susceptibility mapping in the Muhuri River Basin, India", Discover Geoscience, 2026

Publication

<1 %

assets.researchsquare.com

17	Internet Source	<1 %
18	edepot.wur.nl Internet Source	<1 %
19	ijaerd.com Internet Source	<1 %
20	Stephen J. Walsh, David R. Butler, George P. Malanson. "An overview of scale, pattern, process relationships in geomorphology: a remote sensing and GIS perspective", Geomorphology, 1998 Publication	<1 %
21	upcommons.upc.edu Internet Source	<1 %
22	Ankit Singh, Adaphro Ashuli, Niraj K C, Nitesh Dhiman, Chandra Shekhar Dubey, Dericks Praise Shukla. "Evaluating causative factors for landslide susceptibility along the Imphal-Jiribam railway corridor in the North-Eastern part of India using a GIS-based statistical approach", Environmental Science and Pollution Research, 2023 Publication	<1 %
23	Francisco Parra, Jaime González, Max Chacón, Mauricio Marín. "Modeling and Evaluation of the Susceptibility to Landslide Events Using Machine Learning Algorithms in the Province of Chañaral, Atacama Region, Chile", Sustainability, 2023 Publication	<1 %
24	Johnny Vega, Cesar Hidalgo. "Comparison study of a landslide-event hazard mapping using a multi-approach of fuzzy logic, TRIGRS model, and support vector machine in a data-	<1 %

scarce Andes Mountain region", Arabian Journal of Geosciences, 2023

Publication

Exclude quotes Off

Exclude matches

< 14 words

Exclude bibliography Off

Landslide Susceptibility Assesment in Sikkim using Frequency Ratio and Shannon Entropy

GRADEMARK REPORT

FINAL GRADE

GENERAL COMMENTS

/0

PAGE 1

PAGE 2

PAGE 3

PAGE 4

PAGE 5

PAGE 6

PAGE 7

PAGE 8

PAGE 9

PAGE 10

PAGE 11

PAGE 12

PAGE 13

PAGE 14

PAGE 15

PAGE 16

PAGE 17

PAGE 18

PAGE 19

PAGE 20

PAGE 21

PAGE 22

PAGE 23

PAGE 24

PAGE 25

PAGE 26

PAGE 27

PAGE 28

PAGE 29

PAGE 30

PAGE 31

PAGE 32

PAGE 33

PAGE 34

PAGE 35

PAGE 36

PAGE 37

PAGE 38

PAGE 39

PAGE 40

PAGE 41

PAGE 42

PAGE 43

PAGE 44
