CHAPTER 5

RESULT AND DISCUSSIONS

5.1 KEY FINDINGS

- The Random Forest model provides a probability value for each pixel in the study area, representing the likelihood that a given location is prone to a landslide ranging from **0 to 1**. By incorporating all the parameters, the Random Forest model assesses the likelihood of landslides across different areas, considering both physical terrain features (like slope and aspect, etc) and human-made influences (such as roads and fault lines).
- After obtaining the probability for each pixel, the classification is done by applying these predefined thresholds and classified into 5 classes.
- The probability values are compared against the thresholds, and each pixel is assigned a class based on where its probability falls:
- If the probability is less than or equal to **0.15**, the pixel is classified as Very Low (1) susceptibility.
- If the probability falls between **0.15 and 0.35**, it is classified as Low (2).
- If the probability is between **0.35 and 0.60**, it is classified as Moderate (3).
- If the probability is between **0.60 and 0.80**, the pixel is classified as High (4).
- If the probability exceeds **0.80**, the pixel is classified as Very High (5) susceptibility.
- The classification uses **range-based interval method** to categorize the probability values, where each class is assigned based on a fixed range of probability values.
- After the threshold-based classification, the model assigns a susceptibility rank **(1 to 5)** to each pixel in the study area. The resulting classification maps the entire area with the corresponding susceptibility classes, which can then be visualized in GIS systems or used for further analysis.

- The range-based interval method is used to classify landslide susceptibility because it provides a clear, standardized way to interpret the probability values. It simplifies the classification process into easily understandable categories, such as Very Low to Very High, making it accessible to a wider audience. The method is based on common practices in landslide studies and allows for consistent comparisons across different regions. By dividing probability values into predefined ranges, it effectively segments the landscape according to varying levels of risk.

Table 5.1.1 Susceptibility Grading

Range of Probability value of each pixel	**Classified value**	**Susceptibility Grading**
Less than 0.15	1	VERY LOW
0.15 to 0.35	2	LOW
0.35 to 0.60	3	MODERATE
0.60 to 0.80	4	HIGH
Higher than 0.80	5	VERY HIGH

5.2 ACCURACY ASSESMENT

5.2.1 CONFUSION MATRIX

A confusion matrix is a performance measurement tool for classification problems, providing insight into how well a machine learning model performs. It shows the number of correct and incorrect predictions, broken down by each class. The matrix is typically represented in a 2x2 format for binary classification (positive and negative), but can be extended for multi-class problems. It includes:

- True Positives (TP): Correctly predicted positive cases.
- False Positives (FP): Incorrectly predicted positive cases (false alarms).
- True Negatives (TN): Correctly predicted negative cases.
- False Negatives (FN): Incorrectly predicted negative cases.

The confusion matrix helps calculate various evaluation metrics like accuracy, precision, recall, and F1-score.

- Confusion Matrix generated in this model:

[[82, 17],

[32, 56]]

- True Positives (TP) = 82: These are the correctly classified landslide points (class 1).
- False Positives (FP) = 32: These are the locations predicted as landslide-prone but are actually non-landslide areas.
- False Negatives (FN) = 17: These are the landslide-prone areas that were incorrectly classified as non-landslide.
- True Negatives (TN) = 56: These are the correctly classified non-landslide areas.

Table 5.2.1.1 Confusion matrix

Predicted

Actual		Positive	Negative
	Positive	82(TP)	17(FN)
	Negative	32(FP)	56(TN)

5.2.2 Overall accuracy

Overall accuracy is a measure of the classification model's performance, indicating the proportion of correct predictions (both positive and negative) out of all predictions made by the model. It is one of the simplest and most commonly used metrics to evaluate a model's performance.

- Overall Accuracy = (TP + TN) / Total predictions
- Total Predictions = TP+TN+FP+FN
- Overall accuracy of this model = (82+56) / (82+56+17+32)

  = 138/187

  Overall accuracy is 0.73797 (73.8%)

5.2.3 Kappa coefficient

The Kappa coefficient (also known as Cohen's Kappa) is a statistical measure that evaluates the agreement between two raters or classification models, correcting for agreement occurring by chance. In the context of classification, it helps to understand how well the model's predictions align with the actual values, considering random chance.

Kappa values range from -1 to 1:

- 1 indicates perfect agreement.
- 0 indicates no agreement better than random chance.
- Negative values indicate worse than random agreement

Kappa coefficient formula:

$$\kappa = \frac{P_o - P_e}{1 - P_e}$$

Po is Overall accuracy = 0.73797

$$P_e = \left(\frac{(TP + FP)}{N} \times \frac{(TP + FN)}{N} \right) + \left(\frac{(TN + FP)}{N} \times \frac{(TN + FN)}{N} \right)$$

N is total no of samples = 187

Kappa Coefficient value of this model is 0.49

5.2.4 Testing and Training Data

- Randomly selecting 70% of the data for training and the remaining 30% for testing helps ensure unbiased, reproducible results.
- The use of a fixed random seed (1234) ensures that the split is reproducible. This means that every time you run the model, you will get the same training and testing sets, which is essential for ensuring consistent results, especially during iterative model improvement or evaluation.
- This balance of training and testing allows for proper model tuning and ensures that the final model has learned general patterns in the data, not just noise or specificities from the training set.

5.2.5 Feature Importance

The Random Forest model assigns varying levels of importance to each feature based on how much they contribute to reducing classification error. In this model, the following features received significant importance:

- **Faults Distance (77.15%):** The proximity to fault lines is the most important feature, as faults are closely linked to tectonic activity, which can trigger landslides.
- **Slope (71.42%):** Steeper slopes are highly influential in landslide susceptibility, making slope one of the most important factors in the model.
- **NDVI (74.68%):** Vegetation index is crucial because vegetation helps stabilize the soil, and areas with less vegetation are more prone to landslides.
- **Aspect (70.80%):** The direction of the slope affects exposure to sunlight and rainfall, both of which influence landslide occurrences.
- **Roads Distance (62.35%):** Proximity to roads is important, as roads can destabilize the land and contribute to increased landslide risk.
- **Streams Distance (64.47%):** Streams can erode soil and weaken terrain, making their proximity significant in landslide prediction.
- **Rainfall (37.19%):** The amount of rainfall directly impacts soil saturation, increasing the likelihood of landslides during wet conditions.
- **Soil (25.58%):** Soil type is also important, though it has a slightly lesser impact compared to other factors like slope and rainfall.
- **LULC (21.94%):** Land use and land cover types affect the terrain's stability, but have a lower contribution compared to the more direct physical factors.
- **Geomorphology (2.60%):** While geomorphology influences the terrain, it has the least contribution in this model for predicting landslide susceptibility.
- **Lithology (3.00%):** Lithology, representing the type and structure of rocks, showed minimal influence in the model compared to other parameters, indicating that rock type alone may not strongly govern landslide susceptibility in the Nilgiris region.
- **Landslide inventory locations** are used for training the model

Table 5.2.5.1 Feature Importance

*Features are arranged in input order

Parameter	Feature name according to random forest model	Importance (%)
Slope	b1	71.42
Aspect	b1_1	70.80
NDVI	b1_1_1	74.68
LULC	b1_2	21.94
Rainfall	b1_3	37.19
Geomorphology	b1_4	2.60
Lithology	b1_5	3.00
Soil	b1_6	25.58
Fault	faults_distance	77.15
Road	roads_distance	62.35
Streams	streams_distance	64.47