Rainfall-Triggered Landslide Susceptibility Mapping Using GIS and LightGBM: A Case Study of Uttara Kannada

Babitha Ganesh [1,2] , Shweta Vincent [1,*] , Janardhana Bhat [2,3], Sameena Pathan[4], Ganesh Bhat[5]

[1] *Department of Mechatronics, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, 576104, Karnataka, India*

[2] *Department of Computer Science and Engineering, Canara Engineering College, Bantwal, 574219, Karnataka, India*

[3] *Department of Electronics and Communication Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, 576104, Karnataka, India*

[4] *Department of Information and Communication Technology, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, 576104, Karnataka, India*

[5] *Department of Electronics and Communication Engineering, Canara Engineering College, Bantwal, 574219, Karnataka, India*

Summary: Landslide susceptibility mapping is the crucial process for identifying the areas vulnerable to landslides. Uttara Kannada district, located within the boundaries of Western Ghats of Karnataka, India has often experienced slope failures during the rainfall season. This study aims to ascertain leading triggering factors beyond rainfall and to supplement the high resolution landslide susceptibility map in ArcGIS environment uisng the LightGBM classifier. The dataset comprising information of 867 historical landslides and ten different parameters including slope, aspect, soil type, geomorphology, elevation, LULC, TRI, TWI, lineament density and distance to road were gathered from different remote sensing based sources. The permutation model has been used to discover the most influential attributes. Slope, lineament density and soil type have emerged as the most influential ones. The LightGBM classifier has been chosen to prepare LSM for its superior capability in handling categorical variables and its ensemble based approach which enhances the prediction accuracy by integrating mutiple classifiers. The model's performance was rigorously evaluated using multiple statistical metrics, achieving a precision of 0.86, recall of 0.937, accuracy of 0.91, f1-score of 0.897, Kappa index value of 0.817 and an AUC-ROC value of 0.97 demonstrating its strong predictive effectiveness. The LSM generated shows that three taluks including Kumta, Ankola and Karwar are having more locations that are susceptible to landslides. The findings of this study provide crucial insights for identifying the high-risk zones in Uttara Kannada, enabling the policy makers and urban planners to make well planned decisions on infrastructure development while minimizing concerns about landslide related hazards.

Keywords: Landslide susceptibility mapping(LSM), Landslide conditioning factors (LCF), Random Forest(RF), LightGBM

1. Introduction

Landslides are one of the most devastating natural hazards causing significant threat to human life as well to the infrastructure. They regularly occur at the hilly regions often triggered by hevay precipitation, earthquakes and human activities. While, landslides have been a concern in locations that are having steep slopes and fragile geological formations, their frequency and impact have been increasing due to rising developement activities. Uttara Kannada, a district spanning arround 10,291 km^2 situated within ecologically sensitive Western Ghats had witnessed a surge in landslide occurences during the monsoon season. Historically, this region had no significant record of landslides until the catastrophic event took place in Karwar, which resulted in the death of 19 people. In subsequent years, additional landslides were reported along the Kumta coast and Karwar, where boulders rolling down a steep hill caused the train mishap. The most recent landslide in Shirur,Ankola on July 16, 2024, further highlights the growing susceptibility of hazards, as heavy rains and road expansion activities triggered the slope failure, claiming nine lives and two individuals missing[1].

Identifying the areas that are vulnerable to landslides is the primary motto of the landslide susceptibilty mapping(LSM). Geographic Information systems(GIS) have emerged as a powerful tool for LSM creation by integrating spatial and environmental data to analyze terrain stability. GIS allows overlaying different landslide triggering factors to identify the risk prone areas. In recent years, ML techniques have revolutionized LSM by capturing the complex and non-linear relationship between LCFs and the possibility of slope failure occurence.

Traditional statistical methods, while useful often fail to account for the intricate interactions among the various triggering factors. On the other hand, ML classifiers can learn from historical data, recognize patterns and improve the predictive power. Among the ML methods, ensemble learning techniques have gained prominence due to their ability to enhance classification performance by combining multiple models[2].

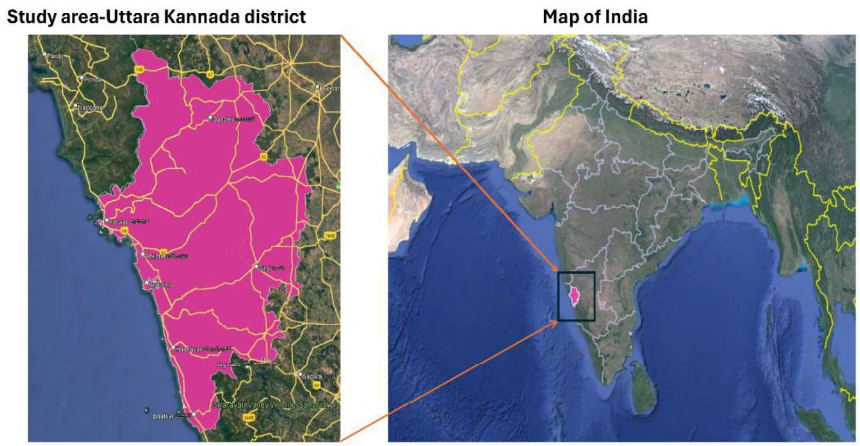


Fig. 1 Study area mapping

LightGBM(Light Gradient Boosting machine) is the gradient boosting framework developed by Microsoft, has been proven to be highly effective in generating LSMs. This algorithm is well suited to manage large datasets with high dimensions. Unlike conventional tree based models, lightGBM efficiently handles categorical variables, reduces overfitting and often faster training speed[3]. Moreover, its ability to assign greater importance to the most influential features enhances the landslide prediction accuracy. These benefits make LightGBM a superior choice for developing reliable LSM which is essential for disaster preparedness and land-use planning.

Although some research studies have been conducted in different parts of Uttara Kannada[4] [5]to the best of our knowledge, no study has been focused on creating the LSM for the entire district. Additionally, despite its strong capabilities in LSM generation, LightGBM remains largely unexplored in the context. Therefore, the primary objective of this study is to develop an LSM for the entire Uttara Kannada district by integrating GIS and LightGBM classification algorithm. The study area mapping is shown in Fig. 1.

2. Methodology

The overall methodology followed in this work has three main steps as depicted in the Fig. 2. These include data collection, modelling and evaluation. The description of these steps is presented in this section.

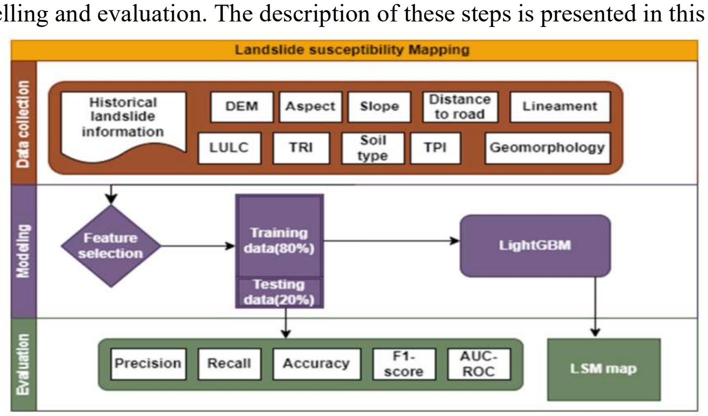
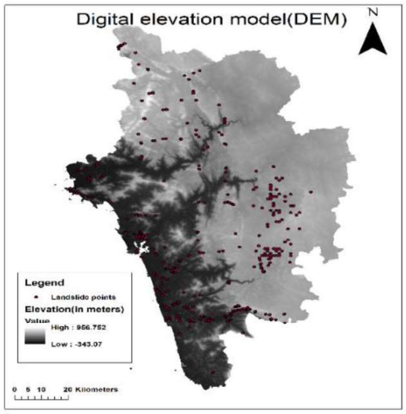
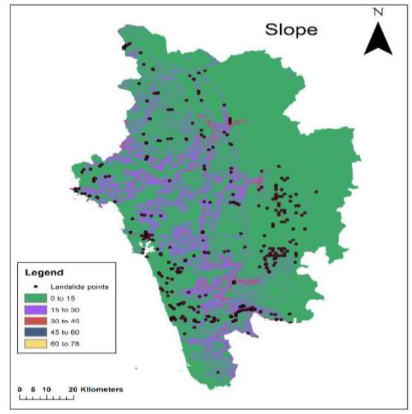


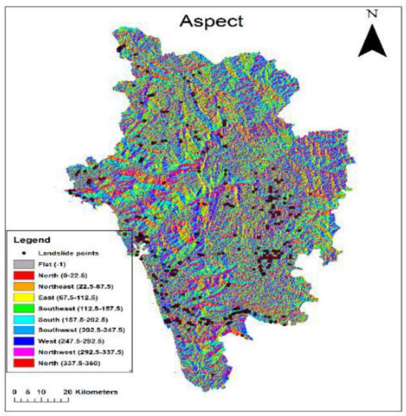
Fig. 2 Methodology



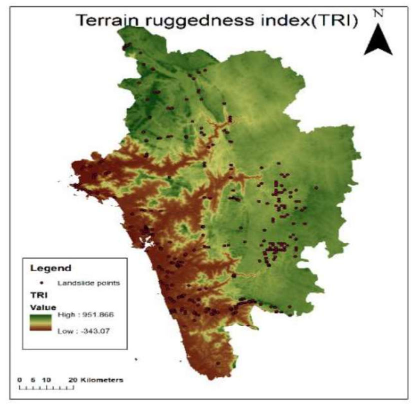
a) DEM



b) Slope



c) Aspect



d) TRI

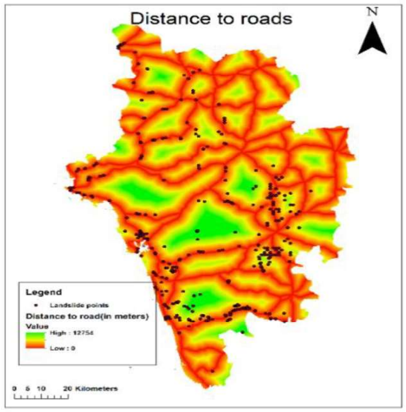
Fig. 3 LCFs(DEM, Slope, Aspect, TRI)

2.1 Data collection:

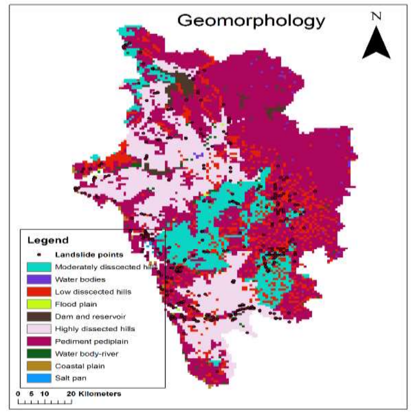
The preparation of landslide susceptibility map using ML classifier mainly requires the information of historical landslide locations and the different triggering factors that influence the occurence of slope failures. The information of landslides has been collected from past records maintained by GSI in Bhukosh portal[6]. This inventory includes 437 landslide locations that has witnessed landslides during the period of 2014 to 2020. The dataset analyis revealed that all of these landlsides were triggered by heavy rainfall. It is also essential to consider other secondary factors that lead to the occurrence of slope failures. The LCF maps generated for the study area is depicted in the Fig. 3 and Fig. **4**. These maps have been generated in ArcGIS environment by using the remote sensing data available in public sources. The 30 m resolution DEM that represent the terrain of the study area has been downloaded from Bhukosh portal. Factors such as TRI, TWI and aspect have been derived from DEM. Lithology and soil maps were collected from Bhukosh portal. The LULC map has been clipped from the global LULC map available as ESRI land cover data[7]. DIVA-GIS provides free geographic data, including road networks, which can be used to create a distance to road map[8]. The soil map shown in the figure has been created using data provided by FAO-UNESCO[9]. The lineament map of the study area was created by digitizing the lineament data obtained from Bhuvan[10] at a 1:50,000 scale.

2.2 Modeling:

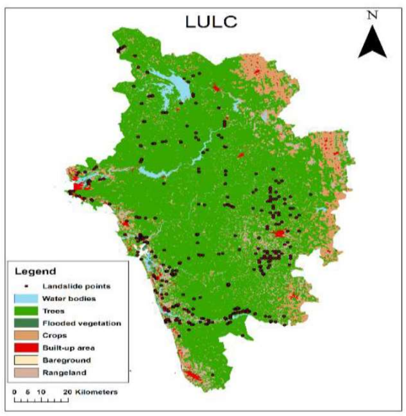
Although we have prepared 10 LCFs, all of them need not be used as some may not significantly influence slope failure prediction for the study area considered. To determine the most relevant features, Permutation feature importance model has been employed. The Fig. 5 illustrates the feature importance values, revealing that the slope, lineament density, soil type and geomorphology are most influential features. In contrast, LULC, aspect and TWI showed minimal contribution to the results. Therefore these factors have been excluded from further analysis.



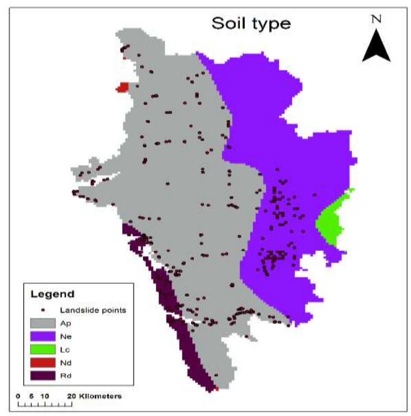
a) Distance to road



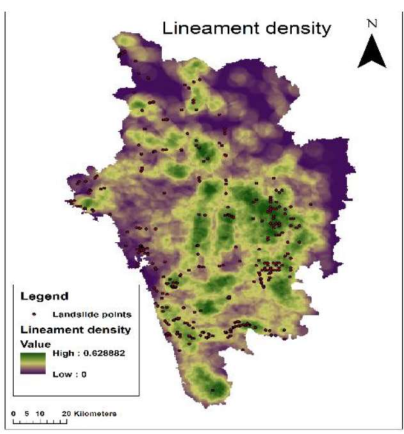
b) Geomoprohology



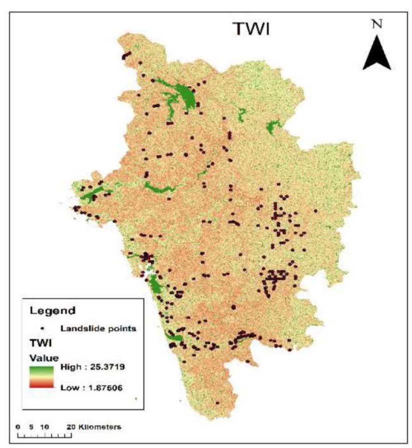
c) LULC



d) Soil type



Lineament density



f) TWI

Fig. 4 LCFs(Distance to road, Geomorphology, LULC,Soil type, Lineament denisty and TWI)

After selecting the optimal set of LCFs, these maps were overlaid on the 437 landslide amd randomly seelcted 505 non-landslide locations. By using ”**extract multi values to points**” function of the ArcMap, the CSV file reqired to train and test the model has been generated. Since the dataset included soil type and geomorphology as categorical variables, it was aimed to select the classifier that can handle these without requiring explicit encoding. Additionally, we needed a model that could automatically assign the appropriate weights to the features based on their significance. Considering these criteria, LightGBM(Light Gradient Boosting Machine) has been selected for generating the LSM due to its efficiency, ability to handle categorical variables, and superior performance in feature selection.

2.3 Evaluation:

The model that has been trained by using the training dataset generated has to be tested to check how well the model is able to distinguish between the locations that are susceptible to landslides or not. For this six different performance metrics including precision, recall,accuracy, F1-score,kappa index and AUC-ROC.

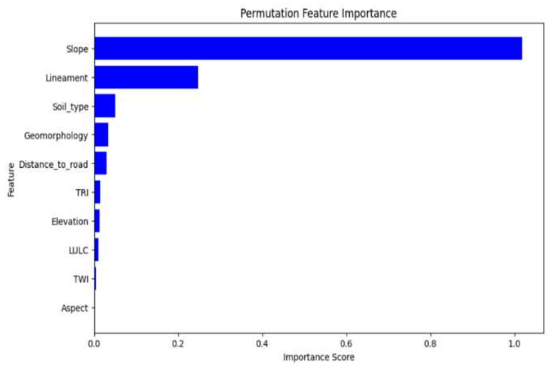


Fig. 5 Importance of features

.

3. Results and Discussions

Tabel 1 shows the values of evaluation metrics. The precision reveals that 86% of the locations identified as landslide-prone are actually correct. The recall value indicates that the model successfully detects 93.7% of the actual landslide prone locations, which is the strong indicator of very low false negative rate. The accuracy suggests that 91% of all locations were correctly classified. The F1-score reveals the well-balanced model, effectively handling both false positives and false negatives. The Kappa index is an indication of strong agreement between the model's predictions and actual values beyond random chance. The AUC-ROC shown in the Fig. 6 AUC-ROC shows that the model has an excellent ability to differentiate between landslide-prone and non-prone areas.

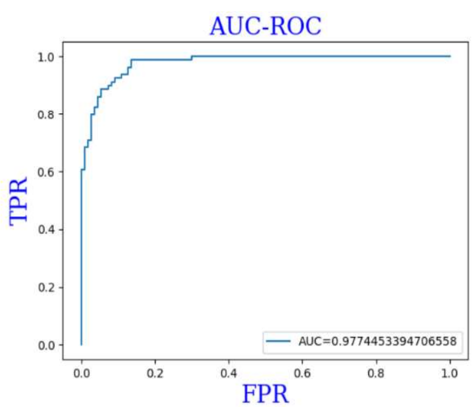


Fig. 6 AUC-ROC

After evaluating the model, approximately 9000 points were randomly selected from the study area, and the corresponding prediction probability values for landslide occurence(1: susceptible, 0: non-susceptible) were generated. The values indicating the confidence of the model in predicting it as susceptible were then spatially distributed using the Inverse Distance Weighting(IDW) method.These values were then reclassified into 5 different classes for the better interpretation. The areas shown in purple color are highly susceptible to landslides with the prediction probability values ranging from 0.6 to 0.8. The prediction probability values less than 0.6 indicates that they are safe places without the risk of landslides. The resultant LSM is shown in the Fig. 7.

Tabel 1 Values of performance metrics

Performance metric	Values
Precision	0.86
Recall	0.937
Accuracy	0.91
F1-score	0.897
Kappa index	0.817

This LSM reveals that around 29% of the total study area is at high risk of landslides. Additionally, around 29% of the total area falls under the moderate risk, indicating the potential for landslides under extreme conditions such as heavy rainfall or land disturbances. The remaining 47% of the area is free is from the landslide risk, making it relatively stable locations. Among the 12 taluks of Uttara Kannada, the Western Coastal taluks of Karwar, Kumta and Ankola exhibit the higher denisty of landslide prone areas. Because of its steep terrain and closeness to the Arabian Sea, Karwar, which is the district seat and is situated in the northwest coastal region, frequently encounters landslides. Significant landslide vulnerability is also seen in Ankola, which is located south of Karwar, and Kumta, which is located farther south along the coast. These areas are probably affected by deforestation, infrastructural development, and the effects of the seasonal monsoon. Meanwhile, areas categorized under moderate risk require precautionary measures, especially in Siddapur, Sirsi, and Yellapu**r**, where hilly terrain and land-use changes contribute to occasional landslides.

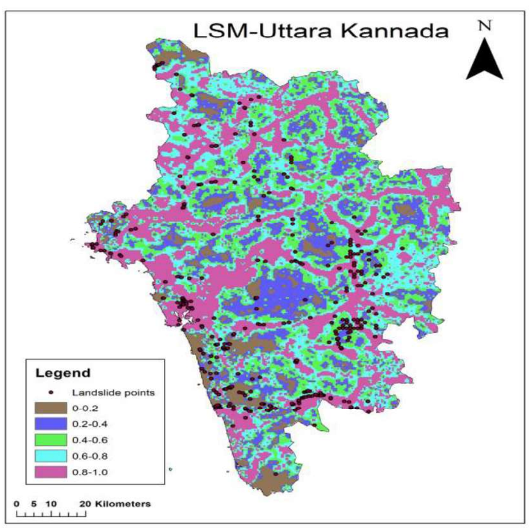


Fig. 7 Landslide susceptibility map of Uttara Kannada

4. Conclusion

This study successfully developed the high resolution LSM for the entire Uttara Kannada district using LightGB ensemble classification algorithm in the ArcGIS environment. The analysis identified slope, lineament density and soil type as the most influential factors along with heavy rainfall in the study area. The model developed exhibited good performance with an AUC-ROC of 0.97 and other strong evaluation metrics, confirming its reliability. The LSM highlights Kumta, Ankola, and Karwar as the most landslide-prone taluks. These findings provide valuable insights for policymakers and urban planners, aiding in risk mitigation and sustainable infrastructure development in landslide-prone areas.

References

[1] P. Kundu, V. Menon, S. Kolathayar, and P. U, "The Shirur landslide of July 2024 triggered by intense rainfall and unchecked development," *Natural Hazards Research*, Jan. 2025, doi: 10.1016/j.nhres.2025.01.005.

[2] S. Narayana and K. Jayaraman, "Landslide Hazard Zonation Mapping Using ML Techniques – A Study Over Kodagu District, Karnataka," in *2024 IEEE 6th PhD Colloquium on Emerging Domain Innovation and Technology for Society (PhD EDITS)*, IEEE, Nov. 2024, pp. 1–2. doi: 10.1109/PhDEDITS64207.2024.10838639.

[3] G. Ke *et al.*, "LightGBM: A Highly Efficient Gradient Boosting Decision Tree." [Online]. Available: https://github.com/Microsoft/LightGBM.

[4] Avinash KG and Ashamanjari KG, "A GIS and frequency ratio based landslide susceptibility mapping: Aghnashini river catchment, Uttara Kannada, India," 2010. [Online]. Available: www.IndianJournals.com

[5] R. T. V Subash Chandran *et al.*, "Landslide Susceptible Zone Mapping in Uttara Kannada, Central Western Ghats Sahyadri Conservation Series 7," 2012.

[6] "Bhukosh." Accessed: Feb. 09, 2025. [Online]. Available: https://bhukosh.gsi.gov.in/Bhukosh/Public

[7] "Esri Land Cover." Accessed: Feb. 09, 2025. [Online]. Available: https://livingatlas.arcgis.com/landcover/

[8] "Data — DIVA-GIS." Accessed: Feb. 09, 2025. [Online]. Available: https://diva-gis.org/data.html

[9] "FAO/UNESCO Soil Map of the World | FAO SOILS PORTAL | Food and Agriculture Organization of the United Nations." Accessed: Feb. 09, 2025. [Online]. Available: https://www.fao.org/soils-portal/data-hub/soil-maps-and-databases/faounesco-soil-map-of-the-world/en/

[10] "Thematic Data - Bhuvan Wiki." Accessed: Feb. 09, 2025. [Online]. Available: https://bhuvan.nrsc.gov.in/wiki/index.php/Thematic_Data