Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 44 No. 5 (2023)

A Comprehensive Study of Machine Learning and Deep Learning methods for Landslide Susceptibility Mapping

[1]Anil D, [2]Savitha Hiremath, [3]S H Manjula, [4]Venugopal K R

[1]Research Scholar, Bangalore University, Bengaluru
[2][3][4] Department of Computer Science and Engineering
University Visvesvaraya College of Engineering, Bengaluru, India.

Abstract— Landslides are one of the most frequently occurring natural hazards worldwide and cause serious damage to human life, severe social and economic loss. Landslide susceptibility is the likelihood of a landslide occurring in the area determined by local terrain conditions predicting where they are more likely to occur. This paper provides a detailed survey on numerous methods and approaches of determining landslide susceptibility. The key challenges for each method are examined, summarized and discussed with their advantages and disadvantages. Moreover, to improve the accuracy of satellite images several object-based and pixel-based classification methods have been discussed. Based on the detailed survey carried out, our attention is to extend the existing research by considering the root cause and primary triggering factors of landslide to better predict the likelihood in different parts of western ghats region, Karnataka, India. This survey can help the other researchers to know the status on Landslide Susceptibility and provide direction for further research.
Index Terms — Conditioning Factors; Deep Learning; Landslide Susceptibility Mapping; Machine Learning; Remote sensing; Satellite images

1. INTRODUCTION

A disaster refers to an unforeseen and catastrophic event that disrupts the normal functioning of a community, causing extensive harm, significant loss of life, damage to infrastructure, and a breakdown in essential services. These events can encompass a range of natural occurrences such as landslides, hurricanes, earthquakes, floods, wildfires, and tsunamis, as well as human-made crises like industrial accidents, acts of terrorism, and nuclear incidents. Landslides are events where masses of rock, soil, or debris rapidly move down a slope or hillside. Landslides are events that can occur abruptly or progress slowly, and their initiation is often influenced by a range of factors. These factors encompass natural phenomena like intense rainfall, seismic activity, volcanic eruptions, as well as human-related activities like construction and mining. however, these techniques range from qualitative to quantitative approaches. Though numerous models exist, the prediction of landslide susceptibility mapping is always inquisitive [1]. The literature emphasizes various methods to study susceptibility mapping which includes both Qualitative and Quantitative models including geomorphic analysis, statistical analysis namely multivariate and bivariate techniques, multi criteria decision analysis, machine learning techniques and deep learning approaches, object-based and pixel-based classification methods. The previous studies also indicated that there exist different causes of landslides ranging from geological, morphological and human causes. Rainfall duration, lithological material, higher height, steep slope, soil type and depth, slope inclination or aspect, altitude, and drainage networks are the factors that cause landslides.

Problem Statement: Landslides are frequent and hazardous natural events, particularly prevalent in regions characterized by steep terrain, loose soils, and intense rainfall. Machine learning techniques have been widely applied to forecast landslide events by considering the geographical features and landscape types. However, there exists an opportunity for advancing this research by delving into the classification of landslide triggers, which can significantly enhance our ability to predict the probability of future landslides. Our study concentrates on identifying the primary causes or triggers of landslides, thereby improving our capacity to assess the likelihood of recurring or future landslide incidents.

Motivation: The principal goal of this research is to enhance the effectiveness of the prediction models by including historical rainfall data with other conditioning factors used in the previous studies such as soil depth, soil slope, land cover etc. The present study represents indeed a novel approach for predicting landslide susceptibility for Western Ghats and the study area. Landslide susceptibility mapping aims at identifying areas that are most vulnerable to landslides based on conditioning factors that determine the tendency for landslide susceptibility. Remote sensing and geographic information system (GIS) has been used in recent times to act as decision support tool.

Objective: The main aim of this study is to assess and estimate the susceptibility of landslide-prone areas within the Western Ghats regions of Karnataka, India. The assessment is conducted by combining both machine learning and deep learning techniques within Geographic Information System (GIS) software, complemented by remote sensing data analysis. The study encompasses various levels of susceptibility, ranging from very high to low, which are delineated using GIS tools. Importantly, the study aims to provide a framework that can be adapted for similar analyses in different geographic regions. To achieve these objectives, the study employs a range of machine learning techniques, including Logistic Model Tree (LMT), Support Vector Machine (SVC and SVR), Random Forest (RF), Decision Trees (DT), Naive Bayes tree (NBT), as well as ensemble methods like Bagging, Boosting, and stacking.

2. Literature Review

Abeysiriwardana et al., [3] identified the geo-scientific causative factor for the occurrence of landslide is due to weak Shannon Wiener Index (SWI) i.e., intercalation of weathered quartzite and phyllite of Shillong Group of rocks exposed along the moderate to steep (55^0) sloping topography. logistic regression (LR) technique has been used to compute to test for multicollinearity using geographical information systems (GIS) techniques. These methods have been validated using Receiver Operating Characteristics (ROC) curve and Confusion Matrix (CM). Li et al., [4] proposed a method based on SVM called Least Square SVM(LSSVM) which is just a modification of SVM for solving large scale landslides where SVM has to just rely on quadratic programming. LSSVM solves it as it requires to solve only linear equations instead of quadratic programming. They used LSSVM method along with other algorithms to predict large landslides and improved the overall accuracy. LSSVM used for analysing landslides with slope stability and accuracy was better than the neural networks model. LSSVM can be used for predicting shallow landslides and also significantly improved the model by fine tuning parameters.

Pham et al., [5] proposed a hybrid algorithm called Naive Bayes tree model (NBT) which consists of decision trees and naive bayes. It is constructed like a hierarchical tree and probability for determining the classification. They showed how naive bayes can better predict the landslide as it requires less learning time and considered 17 risk factors for classification. It has proved how can we make better predictions among various events considering there influencing factors based on the relationships. The generic approach to increase the accuracy is to make use of ensemble learning algorithms. The main objective for selecting these ensemble methods was due to their ability to train multiple classes while handling the issue of class imbalance.

Jamir et al., [6] used AHP modelling and pair wise comparison matrix of geo factor classes was executed after calculation of Eigen values, fraction rate values and Landslide Occurrence Favorability Score (LOFS) ratings. Once the LOFS ratings for individual geo factor classes for a particular geo factor map is calculated, integer weights of individual maps are also calculated. The integer weights are normalized by minimum weights after that. A raw susceptibility score map was generated using integer weights and LOFS values. The map thus generated, was reclassified using random forest (RF), support vector machine (SVM), and logistic regression (LR). However, results of these methods produced high accuracy and reliable results for the study area. Piralilou et al., [7] identified Object-Based Image Analysis (OBIA) that aims to produce meaningful objects by using geometric and spectral characteristics. Multi scale segmentation technique has been used for analzing high resolution and very high-resolution satellite images. Furthermore, identified ideal segmentation parameters in terms of spectral and spatial quality or the classification is crucial for accurate landslide detection.

3. Study Area

Karnataka, a state in southwestern India, stretches from Dandeli in the north to Mangalore in the south, and extends from the western coastline to include Coorg and Madikeri. Covering an approximate area of 191,791 sq km, it is situated between latitudes 11.5° North and 18.5° North, as well as longitudes 74° East and 78.5° East. Within Karnataka, there are numerous peaks that rise above 1,500 meters in elevation, with

Mullayanagiri standing as one of the tallest at 1,923 meters [8]. The study area is positioned on the eastern side of the Western Ghats and was delineated using ArcGIS mapping software, as illustrated in Figure-1. The climate in this region exhibits characteristics of a subtropical monsoon climate, marked by four distinct seasons: a winter season spanning from January to February, a summer season from March to May, the southwest monsoon from June to September, and a post-monsoon period from October to December.

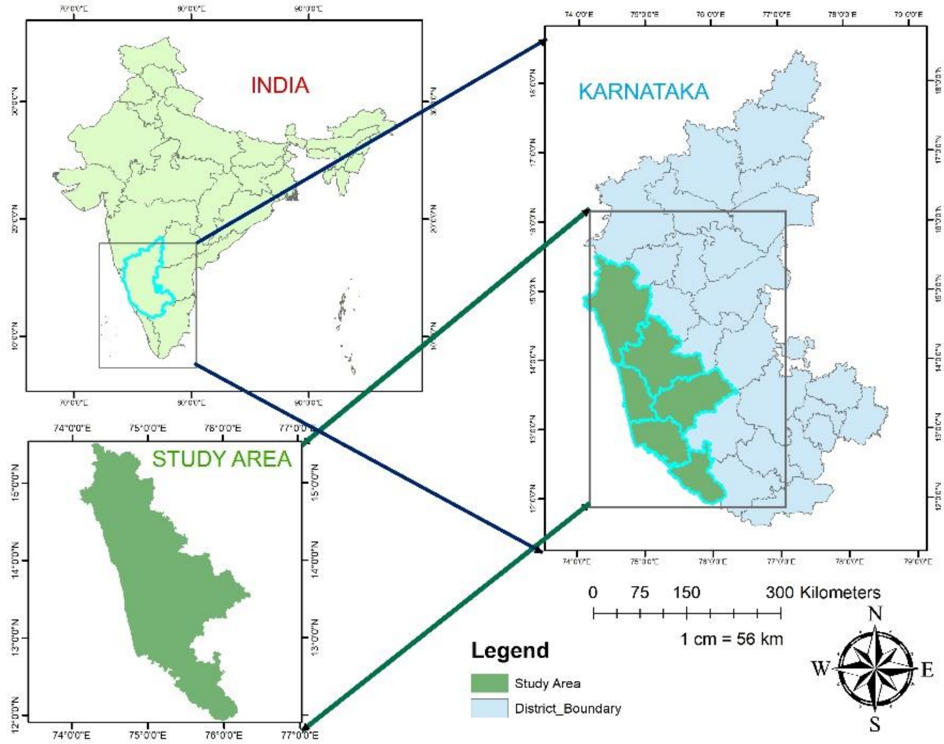

Fig 1: Study Area Map

4. Data Collection

The Data Collection is the first step, Topographic map is obtained from Survey of India, Nakshe portal. Sample toposheet for Kodagu region has been obtained from Survey of India. The Satellite image is obtained from Sentinel-2 satellite in earth explorer i.e USGS portal. Geo-referencing of topographic map in QGIS (Quantum Geographic Information System). Creation of different layers in topography map, the layers include Elevation, Slope, Aspect etc… Satellite images which is collected is geo-referenced and the different layers are then created and projected from that image using interpretation keys [9]. Based on previous studies and local site conditions. Digital Elevation Model (DEM) has been obtained from the Earth explorer of the United States Geological Survey (USGS) website [10]. The spatial data, DEM data and Survey of India toposheet data was used for GIS processing and various maps for geology, slopes, land cover, drainage was obtained using QGIS and Google Earth Engine (GEE). The maps like lithology, lineaments, landforms and slopes prepared from satellite image and other data sources like DEM will help to assess the landslide.

5. Multi Criteria Decision Analysis

Analytic Hierarchy Process (AHP) is a suitable multi criteria decision making (MCDM) method that can be used to solve decision making problems. Analytic Hierarchy Process, was actually developed by Thomas L. Saaty during 1977-80’s. AHP is used to solve ranking decision problems and it mainly has four steps to perform decision making. The first one is problem structuring. In this step, we structure our problem in the sense of the goal, the criteria to be considered and the alternatives to be evaluated and the framing, formulation of the problem and will come to know about decision makers. The second step is on priority calculation. Where, once we have done our problem structuring and the decision makers, we understand the decision problem, goal criteria and alternatives. These responses are taken in using pairwise comparisons and this step is based on pairwise comparisons, and instead of using a numerical scale, we typically use a verbal scale. So, a verbal scale is used and then it is mapped to a numeric scale. Third step is a consistency check to find out whether our pairwise comparisons are consistent with decision makers. Fourth step is sensitivity analysis which checks the robustness of results [11]. Weighted Linear Combination Model is broadly divided into three phases namely data processing, weighting, and mapping then finally model comparison. It assigns weights according to importance of each factor map layer and total score can be calculated by multiplying weight assigned to attributes and finally performing summation operation of all attributes. Final Sustainable map will have factors with higher weight that will be considered as influencing factors [12]. Ordered weighted averaging method was first introduced by Ronald R. Yager in 1988, It is a technique for ranking criteria and addressing the uncertainty from their interaction and has mainly two steps importance weighting and order weighting.

6. Conventional Machine Learning And Deep Learning Methods

Machine learning models have indeed been extended for use in various Earth modeling and geospatial applications, including earthquake prediction and landslide susceptibility mapping. Supervised learning involves learning the model parameters from a given labelled training data set and labelled training data set consisting of input-output pairs, and some examples of supervised learning methods are regression and classification. In Classification problems output are a small number of discrete labels which are also known as classes, whereas in the case of regression problems find a target value based on independent predictors and the output is continuous variable. In clustering grouping of samples based on similarity. Dimensionality reduction is a process of minimizing the dimensions of your feature set that aim at reducing the variance in the given dataset and removing any outliers present in a dataset. A frequently employed technique for reducing dimensionality is principal component analysis (PCA) [12]. Landslides can be identified on satellite images by observing various characteristic surface features which are associated with ground deformations or ground movement. However, identifying these landslides is a very challenging task, this is because the surface features which help us identify landslide valleys vary widely. They can also be very different in different sources of data region or type of event. The object-based and the pixel-based methods have extensively used to automatically map landslides [13]. Deep learning method which has worked well for mapping landslides from digital terrain models and it would be the new future for mapping landslides from satellite and Earth Observation. In order to harness the extensive Earth observation data available, efforts have been made to develop data-intensive machine learning methods. While traditional pixel-based and object-based approaches have been utilized for landslide mapping, the integration of deep learning methods has brought about a revolution in image interpretation tasks within this context.

Table 1: Summary of the Survey of Conventional Machine Learning and Deep Learning Methods

Author	Study	Advantages	Disadvantages
Jiang et al., (2021)	Integrated method for landslide identification using machine learning and deep learning	multiple landslide databases used to validate the proposed method	inconsistencies in the terrain data and the landslide data
Ng et al., (2021)	Rainfall induced landslides using antecedent rainfall factors	different rainstorm characteristics in terms of distinct rainfall spatial distribution	haven't considered feature selection criteria and other machine learning algorithms
Zhu et al., (2020)	Deep Learning approach with cascade-parallel integrated with LSTM for concrete features	LSTM-CRF can extract features and they are then further smoothed and optimized	applying the LSTM-CRF model along with other algorithms for prediction of the landslide on different study areas
Tehrani et al., (2022)	Various forms on prediction have been carried out like detection, spatial forecasting and temporal forecasting	critically evaluated the use of ML methods to model landslide processes	processes of prediction and decision making is not reasonable
Zhang et al., (2022)	Gradient boosting framework-LightGBM and Random Forest used to carry out spatial detection of landslides.	Results shown that weighted models exhibit better than unweighted models.	A high-resolution data is not freely available, which may be result in poor performance of weighted models.
Wang et al., (2021)	XGBoost is used for geospatial data with high spatial and temporal resolutions.	XGBoost algorithm and the proposed hyperparameters were successful for the regional LS mapping and can be recommended for future studies.	Used lower spatial resolution data for the prediction power of the XGBoost method.

They tend to outperform the traditional pixel based and object-based methods. In the traditional machine learning workflow, we take an existing landslide inventory and split it into two non-overlapping regions. It is very necessary to use feature selection techniques in selecting proper factors for effectively predicting landslides susceptibility modeling. Hence different feature selection techniques were introduced for conditioning factors selection in this literature that includes information gain ratio, chi square statistics and others. So, factors selection is a very useful technique that helps to avoid noisy factors and also reduces inappropriate and useful input datasets. Thus, using qualitative datasets increases the accuracy and performance of predictive fitness of the models. Deep learning methods have revolutionized the image interpretation tasks when adopted for landslide mapping. They tend to outperform the traditional pixel based and object-based methods. In the traditional machine learning workflow, we take an existing landslide inventory and split it into two non-overlapping regions. However, this approach cannot be used for a truly automatic mapping of landslides which are triggered by future Events. They Developed a combined learning approach which is effective when mapping landslides or events which are not yet seen by our convolutional neural network

Artificial neural networks are inspired by neural networks found in the human brain where neurons are interconnected in a complex network. With exponential increase in computational power of processors with development of GPUs and CPUs and this led to increase in depth of artificial neural networks. The result of this was Big Bang of deep learning several new architectures and algorithms were developed like Convolutional Neural Networks, Recurrent Neural Networks, Transformers, Encoder and Decoder etc., Segmentation technique can be used for dividing an image into meaningful objects or regions and object size depends on several factors such as scale, shape etc... Satellite imagery plays a pivotal role in landslide detection and the maintenance of inventory maps. The process of extracting information from satellite imagery primarily involves two approaches: pixel-based analysis and object-based image analysis (OBIA) [15]. While OBIA has addressed certain challenges associated with per-pixel image classification, achieving enhanced accuracy in complex tasks through machine learning (ML) models presents several challenges. Several Deep Learning techniques such as U-Net, PSP-Net, Deeplabv3+ and Mask R-CNN are used for predicting landslides using the remote sensing images [16]. To effectively increase the accuracy, backbone feature extraction network method is combined with the deep residual structure. To increase the overall performance pre-defined filter values are used to extract objects in complex satellite images. CNN algorithm and multilayer perceptron (MLP) layer has been used for classifying multi-temporal satellite sensor images.

One of the key challenges is determining the optimal scale parameter for defining objects, particularly for geographic features like landslides. Landslides can vary in size and characteristics, and they may be present at multiple scales within the extent of satellite imagery. This variation in scale can pose difficulties when applying OBIA and ML models. Selecting an appropriate scale parameter is critical because it impacts how objects are defined and classified in the imagery. Furthermore, landslides can exhibit diverse textures, shapes, and spectral signatures, making their accurate identification a complex task. ML models need to be trained on diverse datasets to effectively recognize these variations. Additionally, ensuring that the ML models are robust to changes in lighting, atmospheric conditions, and sensor characteristics is essential for reliable landslide detection.

7. Conclusions

This paper addresses the mapping of landslide susceptibility in various regions of the Western Ghats in Karnataka, India. The selection of methods and factors influencing landslide occurrence is contingent upon available data. The study employs Remote Sensing technology and Geographic Information System (GIS) to investigate landslides and assess the potential for future occurrences. QGIS and ArcGIS software are utilized for landslide mapping and identification. Landslide displacement data is crucial for addressing nonlinear and complex problems. Various conventional machine learning approaches, including supervised and unsupervised methods, are evaluated. The research demonstrates that ensemble models exhibit superior predictive performance compared to other methods. The integration of Remote Sensing and GIS proves to be a valuable computational tool for mapping and analysing Earth's surface features using satellite data. Additionally, the study finds that deep learning techniques, specifically recurrent neural networks, outperform alternative methods in landslide prediction.

References

[1] Das, Suvam, Shantanu Sarkar, and Debi Prasanna Kanungo. "A critical review on landslide susceptibility zonation: recent trends, techniques, and practices in Indian Himalaya." *Natural Hazards* 115.1 (2023): 23-72.

[2] Saha, Abhik, Vasanta Govind Kumar Villuri, and Ashutosh Bhardwaj. "Development and assessment of a novel hybrid machine learning-based landslide susceptibility mapping model in the Darjeeling Himalayas." *Stochastic Environmental Research and Risk Assessment* (2023): 1-24.

[3] Abeysiriwardana, Himasha D., and Pattiyage IA Gomes. "Integrating vegetation indices and geo-environmental factors in GIS-based landslide-susceptibility mapping: using logistic regression." *Journal of Mountain Science* 19.2 (2022): 477-492.

[4] Li, S. H., L. Z. Wu, and Jinsong Huang. "A novel mathematical model for predicting landslide displacement." *Soft Computing* 25 (2021): 2453-2466.

[5] Pham, Binh Thai, et al. "Ensemble modeling of landslide susceptibility using random subspace learner and different decision tree classifiers." *Geocarto International* 37.3 (2022): 735-757.

[6] Jamir, Mademshila, et al. "Landslide susceptibility mapping of Noklak Town, Nagaland, North-east India using bivariate statistical method." *Geological Journal* 57.12 (2022): 5250-5264

[7] Tavakkoli Piralilou, Sepideh, et al. "Landslide detection using multi-scale image segmentation and different machine learning models in the higher himalayas." *Remote Sensing* 11.21 (2019): 2575.

[8] Ganesh, Babitha, et al. "Integration of GIS and Machine Learning Techniques for Mapping the Landslide-Prone Areas in the State of Goa, India." *Journal of the Indian Society of Remote Sensing* (2023): 1-13.

[9] Sowjanya, A., and M. Geetha Priya. "Crop Monitoring of Agricultural Land in Chikkaballapura District of Karnataka Using HSR Data." *Futuristic Communication and Network Technologies: Select Proceedings of VICFCNT 2021, Volume 1*. Singapore: Springer Nature Singapore, 2023. 437-449.

[10] Bhakar, Prashant, Ajit Pratap Singh, and Ravi Kant Mittal. "Assessment of groundwater suitability using remote sensing and GIS: a case study of Western Rajasthan, India." *Arabian Journal of Geosciences* 15 (2022): 1-18.

[11] Saha, Abhik, Vasanta Govind Kumar Villuri, and Ashutosh Bhardwaj. "Development and assessment of GIS-based landslide susceptibility mapping models using ANN, Fuzzy-AHP, and MCDA in Darjeeling Himalayas, West Bengal, India." *Land* 11.10 (2022): 1711.

[12] Li, Huan, et al. "Landslide susceptibility mapping using weighted linear combination: A case of gucheng town in ningxia, China." *Geotechnical and Geological Engineering* 41.2 (2023): 1247-1273.

[13] Wang, Xin, et al. "Change detection-based co-seismic landslide mapping through extended morphological profiles and ensemble strategy." *ISPRS Journal of Photogrammetry and Remote Sensing* 187 (2022): 225-239.

[14] Zhang, Tingyu, et al. "Improved tree-based machine learning algorithms combining with bagging strategy for landslide susceptibility modeling." *Arabian Journal of Geosciences* 15.2 (2022): 183.

[15] Ghorbanzadeh, Omid, et al. "Landslide detection using deep learning and object-based image analysis." *Landslides* 19.4 (2022): 929-939.

[16] Liu, Jie, et al. "Comparison of Deep Learning Methods for Landslide Semantic Segmentation Based on Remote Sensing Images." *2022 5th International Conference on Pattern Recognition and Artificial Intelligence (PRAI)*. IEEE, 2022.

[17] Saha, Sunil, et al. "Prediction of spatial landslide susceptibility applying the novel ensembles of CNN, GLM and random forest in the Indian Himalayan region." *Stochastic Environmental Research and Risk Assessment* 36.10 (2022): 3597-3616.

[18] Achu, A. L., et al. "Landslide susceptibility modelling using hybrid bivariate statistical-based machine-learning method in a highland segment of Southern Western Ghats, India." *Environmental Earth Sciences* 81.13 (2022): 360.