SCIENTIFIC CULTURE
www.sci-cult.com

SCIENTIFIC CULTURE, Vol. 12, No. 4, (2026), pp. 3186-3201
Open Access. Online & Print

DOI: 10.5281/zenodo.12426382

GEOSPATIAL AI FRAMEWORK FOR ASSESSING GEOLOGICAL HAZARDS: LANDSLIDE SUSCEPTIBILITY AND VULNERABILITY MAPPING IN THE WESTERN GHATS OF KARNATAKA

Anil D[1]*, Manjula S H[2]

[1]University Visvesvaraya College of Engineering, Bangalore University, Bengaluru-560001, Karnataka, India
[2]Department of Computer Science and Engineering, University Visvesvaraya College of Engineering, Bangalore University, Bengaluru-560001, Karnataka, India

Received: **21/09/2025**
Accepted: **12/03/2026**

Corresponding Author: Anil D (anilkumaratb@gmail.com)

ABSTRACT

Landslides represent one of the most devastating natural hazards, particularly affecting ecologically fragile regions like the Western Ghats of Karnataka, India. This study presents a robust landslide susceptibility assessment by integrating geospatial datasets and advanced machine learning and deep learning models. A comprehensive landslide inventory was compiled from NASA, GSI, and Google Earth Engine (GEE), combined with critical environmental and topographic variables including slope, elevation, precipitation, and NDVI. The proposed framework employed algorithms such as Support Vector Machine (SVM), Random Forest (RF), XGBoost, Artificial Neural Networks (ANN), and Deep Forest (gcForest) for predictive modeling. A novel Multitemporal and Multiresolution Factor Extraction Algorithm (MTMR-FEA) was developed for optimal feature selection, while the Landslide Vulnerability Assessment Model (LVAM) integrated InGR and multicollinearity diagnostics to enhance model reliability. Performance evaluation using accuracy, precision, recall, and AUC revealed that gcForest achieved the highest accuracy (0.92) and AUC (0.93), outperforming other models. The resultant landslide vulnerability maps (LVMs) delineate critical hazard zones, offering valuable support for regional disaster risk mitigation and land-use planning. This study reinforces the effectiveness of geospatial AI in enhancing landslide prediction and decision-making frameworks.

KEYWORDS: Landslide Susceptibility Assessment, Geospatial Analysis, Machine Learning, Deep Learning, Multitemporal Factor Extraction, Landslide Vulnerability Mapping

1. INTRODUCTION

Landslides are identified as one of the most significant natural disasters. Landslides are caused by a complex interaction of both natural processes and human influences. The primary attribute of a landslide is defined as the phenomenon in which terrestrial materials experience displacement and travel down the incline due to the force of gravity, thereby presenting a considerable risk to the safety of individuals and their possessions. This type of disaster not only results in prompt destruction of residential infrastructure but also leads to the decline of land resources [1]. In recent decades, a large amount of sloping land has been affected in the Western Ghats of Karnataka, India, resulting in frequent landslide occurrences. Therefore, managing and evaluating the future locations of landslides are particularly important to mitigate such threats in this region. Additionally, human activities like deforestation, urbanization and construction on unstable slopes can significantly increase landslide risks. In regions like the Western Ghats of Karnataka, India, landslides frequently occur on steep slopes, particularly during the monsoon season [2]. Consequently, identifying the contributing factors and assessing the vulnerability of different areas to landslides is essential for effective disaster risk reduction and management. In the Western Ghats of Karnataka, landslides pose a recurring threat, particularly on steep slopes and during the monsoon season. Addressing these hazards effectively requires a comprehensive spatial assessment of landslide vulnerability [3]. Key factors such as geomorphology, geology, tectonics, climate, vegetation and land use practices all contribute to a region's susceptibility to landslides. Although landslides are natural occurrences, human activities often amplify this vulnerability, complicating the prediction of when and where they might occur. Landslide vulnerability maps (LVMs) serve as essential tools to classify high-risk zones and identify critical variables that drive landslide events [3].

Landslide Vulnerability Assessments (LVA) are crucial tools in understanding and identifying areas at high risk of landslides [4]. To determine how landslides may affect communities, infrastructure and the environment, susceptibility must be assessed using physical and socio-economic factors. This is in contrast to susceptibility maps, which are primarily concerned with forecasting the locations where landslides are likely to occur. When they occur in hilly or mountainous regions, landslides may be among the most devastating natural catastrophes because of the human lives they take and the property and infrastructure they ruin [5]. In recent studies, geo-environmental factors have been the primary focus in understanding landslide vulnerability, while socio-economic influences have received less attention. In the Western Ghats of Karnataka, where landslides are frequent, both natural and human-induced factors significantly impact slope stability [6]. Population pressures have led to land modifications for agriculture, infrastructure development and tourism, all of which increase the instability of slopes. This study aims to optimize the Landslide Vulnerability Assessment Model (LVAM) by incorporating a comprehensive analysis of physical and socio-economic factors to enhance predictive accuracy for high-risk areas. Key environmental and geophysical factors, such as slope gradient, altitude, precipitation and soil moisture, play critical roles in landslide susceptibility, with steeper, water-saturated slopes being particularly vulnerable [7].

With advancements in remote sensing (RS) and geographic information system (GIS) technologies, obtaining high-resolution digital elevation models (DEM) and engineering data has become more accessible [8]. Integrating RS-GIS techniques with knowledge-driven or data-mining approaches has become a cornerstone in landslide susceptibility and vulnerability assessments, utilizing both spatial and non-spatial data. Machine learning models often outperform traditional approaches in landslide susceptibility mapping due to their ability to handle non-linear data across various scales. Research indicates that combining ML models with traditional statistical methods, rather than using single ML models, yields superior results in landslide modeling [9]. In this study, we aim to leverage an optimized ensemble of ML models and knowledge-driven techniques within our Landslide Vulnerability Assessment Model (LVAM) to enhance prediction accuracy and identify high-risk areas in the Western Ghats, Karnataka [10]. In addition to machine learning models, recent studies have highlighted deep learning models, such as convolutional neural networks (CNN), deep learning neural networks (DLNN) and recurrent neural networks (RNN), as powerful tools for spatial modeling, often surpassing traditional ML models in accuracy. Deep learning networks like DLNNs, which feature multiple hidden layers, excel in extracting, transforming and recognizing patterns for both supervised and unsupervised tasks. CNN models have shown significant promise in landslide assessment due to their ability to capture complex spatial features, achieving high detection accuracy, as demonstrated by Yu *et al.* [11].

Ghorbanzadeh *et al.* [12] presented mixed findings, indicating that CNN's performance can vary when compared to certain advanced machine learning models, potentially due to the complexity of landslide-prone terrains and diverse data characteristics. Pourghasemi and Rahmati [13] highlighted the effectiveness of artificial neural networks (ANN), which consistently outperformed various statistical and machine learning models in identifying landslide-prone areas. Despite their promising performance, deep learning models are still rarely applied in landslide vulnerability assessments. While susceptibility mapping in regions like Bhutan has benefited from ML applications, few studies have explored deep learning techniques for predicting landslide vulnerability [14]. This study aims to address this gap by incorporating advanced deep learning models within our Landslide Vulnerability Assessment Model (LVAM) to optimize vulnerability mapping and enhance predictive capability in the Western Ghats, Karnataka. It employs a comprehensive database of recent landslides which has been compiled from various standard sources such as NASA, GEE, the Geological Survey of India. To reflect the level of exposure some of the factors incorporated in the analysis include precipitation, slope, elevation and Normalized Difference Vegetation Index (NDVI) [15]. To establish what constitutes the most relevant information in the study and to avoid overfitting of the predictive models, more sophisticated feature selection processes, including Information Gain Ratio (InGR), multivariate collinearity analysis, Particle Swarm Optimization (PSO) and Least Absolute Shrinkage and Selection Operator (LASSO) algorithms were applied [16].

Problem Statement: Landslides are a frequent and destructive hazard during the rainy season in Karnataka's Western Ghats, resulting in extensive damage to property, infrastructure and loss of lives [17]. Although previous research has assessed landslide risks, many models rely on limited datasets and fail to account for a comprehensive range of factors that influence landslide vulnerability, including environmental conditions, geophysical features, socio-economic aspects and infrastructural elements [18].

In this study, a comprehensive assessment of landslide vulnerability in Karnataka's Western Ghats using advanced machine learning and deep learning techniques [19]. A combination of support vector machines (SVM), random forests (RF), XGBoost, artificial neural networks (ANN) and Deep Forest (gcForest) models forms the core of this approach, each offering unique strengths in handling complex, non-linear data for precise vulnerability analysis. Additionally, the Multitemporal and Multiresolution Factor Extraction Algorithm (MTMR-FEA) is employed to extract crucial environmental and geophysical factors over multiple timeframes and resolutions, enhancing the model's ability to capture dynamic changes in the landscape. The extracted factors are then used in the Landslide Vulnerability Assessment Model (LVAM), which integrates machine learning and deep learning techniques to generate accurate vulnerability maps [20].

2. LITERATURE REVIEW

Dou *et al.* [21] conducted a comparative analysis of Random Forest (RF) and Gradient Boosting Models (GBM) for landslide susceptibility mapping in the geologically diverse Izu-Oshima Volcanic Island. The study aimed to evaluate the effectiveness of these ensemble learning algorithms in handling landslide conditioning factors, including topography, rainfall, and vegetation indices. Random Forest slightly outperformed Gradient Boosting in terms of accuracy, achieving a score of 0.89. Guo *et al.* [23] used Decision Tree (DT) models combined with K-means clustering to assess landslide susceptibility in Southwest China. Their approach yielded an AUC of 0.89, demonstrating the effectiveness of combining clustering with ML models. Huang *et al.* [24] explored the application of Convolutional Neural Networks (CNNs) optimized with metaheuristic algorithms for landslide susceptibility mapping in South Korea. The proposed CNN model achieved an AUC score of 0.94, significantly outperforming traditional machine learning models. This research underscores the potential of combining CNNs with optimization techniques to enhance the predictive performance of deep learning models in landslide susceptibility analysis. Guzzetti *et al.* [25] developed an integrated early-warning system for landslide prediction by combining machine learning models with geographical information systems (GIS). The ensemble approach demonstrated superior predictive power, achieving the highest AUC score of 0.89.

Merghadi *et al.* [26] conducted a comparative study of multiple machines learning models, including Decision Trees (DT), Random Forest (RF), Support Vector Machines (SVM), and Logistic Regression (LR), to predict landslide susceptibility across diverse regions. The analysis revealed that RF consistently outperformed other models, achieving an AUC score of 0.91, owing to its ability to handle complex and noisy datasets. Shabani *et al.* [27] applied ensemble learning techniques to enhance landslide susceptibility mapping. They combined Generalized Logistic Models (GLM), Flexible Discriminant Analysis (FDA), Boosted Regression Trees (BRT), and Random Forest (RF) into a single ensemble framework. The combined model outperformed individual algorithms, achieving an AUC score of 0.904. Wang *et al.* [28] introduced a hybrid machine learning and deep learning approach for landslide susceptibility mapping by integrating multi-resolution remote sensing data. Their methodology involved initial feature selection using Random Forest (RF) to identify the most influential conditioning factors, followed by classification using a specialized deep learning network (LSNet). The hybrid approach achieved an impressive AUC score of 0.94, highlighting the synergy between machine learning and deep learning techniques. Gono *et al.* [29] conducted a survey of landslide susceptibility modeling methodologies, emphasizing the performance of hybrid and ensemble models over single machine learning algorithms.

Ali *et al.* [30] conducted an in-depth analysis comparing several machines learning models, including Random Forest (RF), Support Vector Machine (SVM), and Gradient Boosting (GB), to assess landslide vulnerability across diverse regions. Their study highlighted the importance of integrating multi-factorial data, such as topographical, climatic, and geological features, to enhance the accuracy of vulnerability predictions. Among the models, Random Forest emerged as the most effective, achieving an accuracy of 85.6%, significantly outperforming SVM and GB. Liu *et al.* [31] explored the application of deep learning models, focusing on Convolutional Neural Networks (CNNs) and hybrid CNN-LSTM architectures, for large-scale landslide vulnerability assessments. Their research demonstrated the superior performance of deep learning models compared to traditional machine learning algorithms like Logistic Regression and Decision Trees. The CNN model achieved an accuracy of 89.2%, indicating its effectiveness in capturing spatial dependencies and non-linear patterns in geospatial datasets. Pham *et al.* [32] reviewed the integration of Geographic Information Systems (GIS) with machine learning models, such as Logistic Regression (LR), Decision Trees (DT), and Support Vector Machines (SVM), for landslide vulnerability assessments. Their research demonstrated that the SVM model provided the highest accuracy of 83.4%, particularly when geospatial data such as slope, rainfall, and land use were incorporated using GIS tools. Chen *et al.* [33] analyzed the impact of feature selection methods on the performance of models like Random Forest, Gradient Boosting, and Artificial Neural Networks (ANN) for landslide vulnerability assessments. They employed Information Gain Ratio and Random Forest for feature selection, identifying the most significant factors influencing landslides. This optimized approach achieved an accuracy of 87.5%, demonstrating the critical role of feature selection in improving model performance. Raghunandan *et al.* [34] proposed a novel framework integrating socio-economic factors with environmental variables to assess landslide vulnerability using XGBoost and Artificial Neural Networks (ANN). Their study incorporated factors such as population density, infrastructure, slope, and rainfall to provide a holistic assessment of vulnerability. XGBoost emerged as the best-performing model, achieving an accuracy of 90.1%, while ANN achieved 88.7%.

Ma *et al.* [39] utilized Automated Machine Learning (AutoML) for landslide susceptibility mapping, showing that AutoML-generated ensemble models outperformed traditional ML methods, achieving an accuracy of 0.92 compared to 0.84 for manually tuned models. The study emphasizes AutoML's efficiency in optimizing complex terrain models, reducing model training time by 40%. Xia *et al.* [40] introduced KNN-GCN, a model combining K-nearest neighbor (KNN) and Graph Convolutional Networks (GCN) to capture spatial correlations. The model achieved an AUC of 0.93, outperforming standalone KNN (0.85) and GCN (0.89) models. The study highlights the advantage of integrating spatial dependencies, particularly in slope-unit mapping, for enhanced landslide susceptibility prediction. Ganerod *et al.* [41] developed a heterogeneous ensemble model for landslide detection, achieving an F1-score of 0.87, which was 15% higher than individual deep learning models. The ensemble approach effectively combined the strengths of multiple models, resulting in high detection accuracy and robustness in various geographic settings. Ma *et al.* [42] presents a transformer-based model infused with prior knowledge, achieving an AUC of 0.94. The model improved interpretability and provided insights into landslide evolution patterns. The integration of knowledge increased prediction reliability by 20%, supporting more informed decision-making in landslide-prone areas.

3. STUDY AREA

The state of Karnataka is located in the southwestern region of India, covering an area of approximately 191,791 km². It lies between latitudes 11.5°N and 18.5°N and longitudes 74°E and 78.5°E. The Western Ghats in Karnataka extend from Dandeli in the north to Mangalore in the south and from the western coastline up to Coorg (Kodagu) and Madikeri in the interior. This mountainous region is characterized by steep slopes, dense vegetation, and fragile geological formations. Notably, Karnataka has over a dozen peaks exceeding 1,500 meters in elevation, with Mullayanagiri being the tallest at 1,923 meters [43]. Landslide studies in Karnataka have primarily focused on the Western Ghats, which span from Kodagu district in the south to Uttara Kannada in the north, encompassing an area of approximately 27,855.45 km². Most previous research has concentrated on transportation corridors such as national and state highways, where landslides are frequently reported during the monsoon. However, there is limited availability of temporal landslide data for many of these regions, with only a few documented events from media reports dated 2006 and 2013 [44]. **Fig. 1** illustrates the spatial extent of the study area within Karnataka's Western Ghats, generated using open-source ArcGIS software.

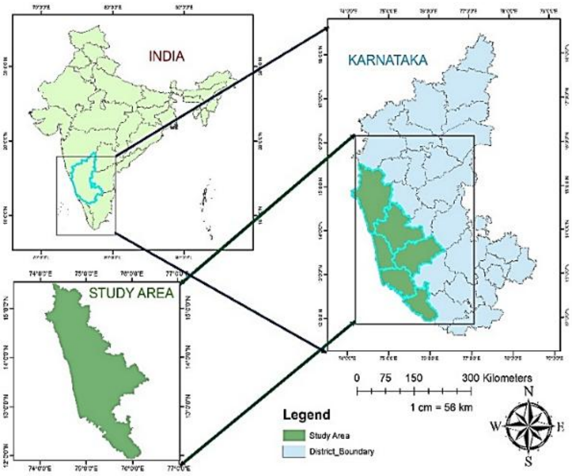

Fig. 1 Location of the study area in the Western Ghats of Karnataka, India, prepared using open-source ArcGIS software.

4. DATASET AND CONTRIBUTING FACTORS

The study of landslides in Karnataka is based on a comprehensive, multi-source inventory, providing a robust dataset for analyzing historical landslide occurrences across the Western Ghats. To effectively identify and quantify the contributing factors, we employ the Multitemporal and Multiresolution Factor Extraction Algorithm (MTMR-FEA). This custom-designed algorithm integrates geospatial datasets from diverse sources and resolutions, spanning the years 2015 to 2023, ensuring both spatial uniformity and temporal consistency [45].

A total of 1,580 confirmed landslide locations were collected from three primary sources:

- National Remote Sensing Centre (NRSC) [1], ISRO – 272 events
- Geological Survey of India (GSI) via Bhukosh portal – 1,260 events
- Google Earth Engine (GEE) – 48 events manually digitized through remote sensing interpretation

Algorithm-1: Multitemporal and Multiresolution Factor Extraction Algorithm (MTMR-FEA)
Input:
Landslide Inventory Points: $L = \{l_1, l_2, ..., l_n\}$ from NRSC (272), GSI (1260), and GEE (48)
Environmental Raster Layers: $R = \{R_1, R_2, ..., R_{17}\}$
including SRTM DEM, MODIS, Sentinel-2, CHIRPS, TRMM, ERA5, and USGS Earthquake data
Spatial Resolution: S = 30 × 30 meters
Temporal Window: T = 2015–2023

Output:
Final dataset $X \in \mathbb{R}^{n \times d}$ with labels $y \in \{0,1\}$, where n = 4827 and d = 17 factors

1. Data Acquisition:
1.1 Retrieve $L = \{l_1, ..., l_n\}$ landslide points from NRSC, GSI, and GEE
1.2 Download geospatial layers R = {DEM, NDVI, NDWI, Lithology, etc.}
covering time period T and resolution S

2. Data Preprocessing:
2.1 Standardize each raster $R_i \in R$ to spatial resolution S = 30m
2.2 Sample R_i values at coordinates of landslide and non-landslide locations
→ Extract pixel values from $R_i(l_j)$ for each $l_j \in L$
2.3 Convert sampled values into CSV → feature matrix $X = [x_{ij}] \in \mathbb{R}^{n \times d}$

3. Feature Computation:
3.1 Derive topographic parameters from SRTM:
- Slope, Aspect, Curvature, Elevation, Flow Accumulation
3.2 Extract climatic variables:

[1] https://nrsc.gov.in

- Precipitation, Temperature, Humidity from CHIRPS, TRMM, ERA5
3.3 Calculate vegetation indices:
- NDVI, NDWI from Sentinel-2
3.4 Include geologic and anthropogenic features:
- Lithology, Soil Moisture, Land Use, Earthquake history

4. Exploratory Data Analysis:
4.1 Plot feature distributions using histograms and KDEs
4.2 Compute Pearson correlation and Variance Inflation Factor (VIF)
→ Eliminate highly correlated or redundant variables

5. Landslide Type Labeling:
5.1 Assign morphological classes: Debris Flow, Rock Fall, Shallow Slide
based on slope threshold, curvature, and landform analysis

6. Dataset Compilation:
6.1 Final dataset: $X \in \mathbb{R}^{4827\times17}$, with features $f_1 \ldots f_{17}$ per record
6.2 Create label vector $y \in \{0,1\}$ for binary classification
6.3 Split X and y into training (70%), validation (15%), and testing (15%)

Return:

Preprocessed feature matrix X and label vector y for ML/DL model training

Table I: Landslide Dataset Samples

District	NASA	GSI Bhukosh	GEE	Total
Chikkamagaluru	50	234	9	293
Dakshin Kannada	10	48	2	60
Kodagu	28	130	5	163
Shimoga	27	123	5	155
Udupi	22	100	4	126
Total				1,580

The landslide inventory utilized in this study is consolidated from three primary sources: the National Remote Sensing Centre (NRSC), the Geological Survey of India (GSI) via the Bhukosh portal, and manually mapped events from Google Earth Engine (GEE). Specifically, 272 events were acquired from NRSC, 1,260 from GSI's Indian Landslide Inventory (ILI), and 48 from GEE-based remote sensing interpretation, culminating in a total of 1,580 validated landslide records. These datasets form the core input for spatial analysis as summarized in Table I. To derive geospatial conditioning factors from these locations, the study employed the Multitemporal and Multiresolution Factor Extraction Algorithm (MTMR-FEA). One of the foundational datasets is the Digital Elevation Model (DEM) from the NASA Shuttle Radar Topography Mission (SRTM), offering 30-meter spatial resolution [46-47].

The MTMR-FEA algorithm ensured consistency in spatial and temporal resolution by harmonizing raster layers from 2015 to 2023 using Google Earth Engine (GEE). A total of 4,827 records were generated by sampling geolocated points, each annotated with 17 contributing factors. These factors span multiple domains: topographic, climatic, vegetation indices, geological and anthropogenic triggers. For example, aspect and curvature were computed from the SRTM DEM, while NDVI and NDWI were derived from Sentinel-2. Earthquake data from the USGS Earthquake Database was integrated through GEE [48,49]. The environmental variables were standardized at 30m resolution, and values were extracted for each location to produce a structured feature matrix for machine learning and deep learning modeling. In addition to raster sampling, statistical analyses including histograms, Kernel Density Estimations (KDEs), and Pearson correlation were employed to evaluate the distribution and influence of each variable [51]. The frequency of key factors such as slope and precipitation, while KDE plots provide smoothed probability density functions for feature interpretation [52]. To quantify the degree of linear link between these characteristics and the frequency of landslides, one can use Pearson's correlation coefficient using the formula given in Equation (1).

$$r_{xy}=\frac{\sum(X-\bar{X})(Y-\bar{Y})}{\sqrt{\sum(X-\bar{X})2(Y-\bar{Y})2}} \quad \text{Equation (1)}$$

Where $\bar{X}$ and $\bar{Y}$ are the average values of random variables X and Y. A high positive correlation like slope angle to landslide occurrence indicates that slopes with steeper slopes are more prone to landslides. This should be done as one of the phases in determining the extent to which each factor contributes to the threat of landslides. The study categorized the landslide inventory into morphological types debris flows, shallow slides, and rockfalls based on terrain and image interpretation [53].

5. METHODOLOGY

The Landslide Vulnerability Assessment Model (LVAM) was developed as a structured framework to systematically identify, optimize, and classify landslide-prone zones using multi-source geospatial and statistical data as shown in **Fig. 2**. This approach integrates both classical and advanced machine learning pipelines, emphasizing model interpretability, performance, and scalability [54]. The process begins with the extraction of critical environmental and geophysical variables using the Multitemporal and Multiresolution Factor Extraction Algorithm (MTMR-FEA). Seventeen contributing factors including slope, elevation, precipitation, NDVI, soil moisture, lithology, and land use were derived from datasets such as SRTM DEM, MODIS, Sentinel-2, TRMM, and GEE, covering the years 2015 to 2023 [55].

5.1 Factor Selection Methods

To enhance model efficiency, mitigate overfitting, and ensure interpretability, a comprehensive multi-stage feature selection strategy was employed. Initially, Information Gain Ratio (InGR) was utilized to quantify the importance of each geospatial factor such as slope, elevation, precipitation, and soil moisture in reducing classification uncertainty. This entropy-based metric enabled the identification of features with the highest discriminative power, guiding the retention of variables most critical to landslide susceptibility prediction. Subsequently, multi-collinearity analysis was conducted during the preprocessing phase to address interdependencies among predictors. The Variance Inflation Factor (VIF) was computed for all features, and those exceeding the threshold VIF>10 was excluded. This step ensured statistical independence across inputs, improving both model stability and generalizability, as supported by prior geospatial modeling research [56]. To further optimize the feature space, advanced wrapper and regularization-based techniques were incorporated. Particle Swarm Optimization (PSO) was employed as a global search algorithm to select subsets of features that maximized classification accuracy while enforcing spatial separation constraints (e.g., excluding non-landslide points located within 1 km of confirmed landslide points). PSO iteratively refined feature combinations based on fitness evaluation, promoting the inclusion of factors contributing to maximum spatial distinctiveness. In parallel, the Least Absolute Shrinkage and Selection Operator (LASSO) was applied for embedded feature selection through regularization [57]. The integration of InGR, VIF, PSO, and LASSO created a robust and multidimensional feature selection pipeline. This approach not only eliminated redundancy and irrelevant noise but also ensured that the retained features provided maximal predictive relevance, thereby enhancing the performance, accuracy, and reliability of the proposed landslide susceptibility model [58].

<u>Algorithm-2: Landslide Vulnerability Assessment Model (LVAM)</u>

Step 1: Multitemporal and Multiresolution Factor Extraction (MTMR-FEA)

- Collect 1,580 landslide inventories across six districts (2015–2023).
- Extract 17 contributing factors (e.g., slope, lithology, precipitation, NDVI) from sources including SRTM DEM, MODIS, Sentinel-2, TRMM, ERA5, and GEE.

Step 2: Feature Selection and Optimization

- Apply InGR to measure the importance of each feature.
- Perform VIF-based multicollinearity filtering.
- Implement PSO to optimize the subset of selected features by minimizing error.
- Use LASSO regression to produce a sparse model by eliminating non-informative features.

Step 3: Model Training and Classification

- Train classifiers: SVM, RF, XGBoost, ANN, and gcForest.
- Classify each geolocation as landslide-prone or non-prone.

Step 4: Hyperparameter Tuning

- Utilize Grid Search to find the optimal parameter combinations for each classifier.

Step 5: Performance Evaluation

- Assess models using Accuracy, Precision, Recall, and AUC metrics.

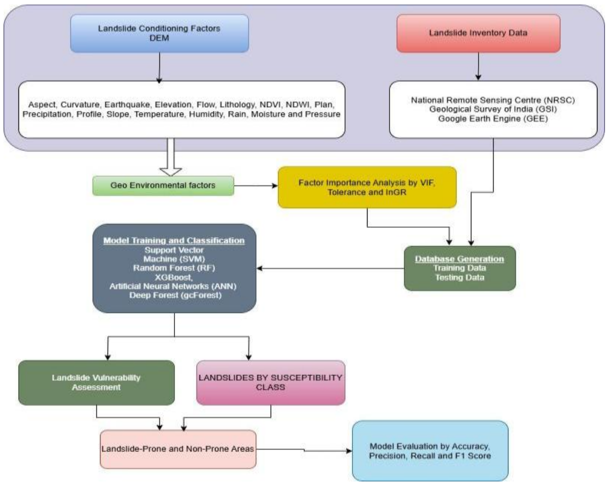

Fig. 2 System Architecture

Information Gain Ratio (InGR): Hypothesis 1 states that by using the new InGR measure, which is an extension of the conventional Information Gain (IG) measure deployed in decision trees, the importance of features for landslides vulnerability prediction can be better assessed. It states the extent to which each of these features (for instance, slope, precipitation, water soil content) contributes to classification and hence assists in determining the features that are more influential. InGR solves this issue with a normalization step where the number of unique values in each feature is taken into account to avoid being dominated by features with many different values [59]. The formula for Information Gain IG (D, A) for a feature A on dataset D is shown in Equation (2).

$$\mathrm{IG(D, A)} = Entropy(D) - \sum_{v \epsilon Values\ (A)} \frac{D_v}{D} \times Entropy(D_v) \quad \text{Equation (2)}$$

Where:
Entropy(D) is the entropy of the dataset D,
D_v is the subset of D where feature A has value v,
$|D_v|$ and $|D|$ represent the sizes of the subset D and the total dataset D

The Information Gain Ratio is calculated by normalizing the Information Gain using the Split Information shown in Equation (3).

$$\mathrm{InGR(D, A)} = \frac{IG(D,A)}{SI(A)} \quad \text{Equation (3)}$$

Where SI(A) is the Split Information, given by:

$$\mathrm{SI(A)} = - \sum_{v \epsilon Values\ (A)} \frac{D_v}{D} Log_2 \frac{|D_v|}{|D|} \quad \text{Equation (4)}$$

This helps to reduce the bias that arises when a feature has many distinct values.

Multi-collinearity Analysis: Also referred to as cross-correlation, multi-collinearity occurs when more than one predictor variable and in the context of landslide vulnerability, these are environmental variables like slope and elevation, are highly correlated, which makes the model coefficients unstable. To tackle this, the Variance Inflation Factor (VIF) is used to assess the extent of the increase in the variance of a regression due to collinearity [60]. The formula for VIF is shown in Equation (5).

$$VIF_i = \frac{1}{1-R_i^2} \quad \text{Equation (5)}$$

Where R_i^2 is the coefficient of determination of the regression model that predicts the i-th feature given all the other features. A VIF that is greater than 10 is considered a sign of high multi-collinearity and the features with high values are excluded for better model stability.

Particle Swarm Optimization (PSO): PSO is used in the optimal selection of elements that affect landslide vulnerability. It mimics the social behavior of birds or fish to analyze which part performs better in unison for the best functioning of the model. Some particles represent possible feature subsets within the swarm, while the swarm leverages the experiences of its members and the swarm as a whole to refine the answer iteratively [61]. Every particle's location and speed i are updated using Equation (6) and Equation (7)

$$v_i(t+1) = \hat{w}\, v_i(t) + c_1 r_1[p_i(t)-x_i(t)] + + c_2 r_2[g(t)-x_i(t)] \quad \text{Equation (6)}$$

$$x_i(t+1) = x_i(t) + v_i(t+1) \quad \text{Equation (7)}$$

Where:
$v_i(t)$ is the velocity of particle i at time step t,
$x_i(t)$ is the position (set of selected features) of particle i,
$p_i(t)$ is the personal best position of particle \(i \),
$g(t)$ is the global best position of the swarm,
$\hat{w}$ is the inertia weight, c_1 and c_2 are acceleration coefficients and r_1 and r_2 are random numbers in [0, 1].

PSO helps identify the subset of features that maximize classification accuracy by minimizing redundant or irrelevant features.

LASSO (Least Absolute Shrinkage and Selection Operator): Models that predict the likelihood of landslides often use LASSO for regularization and feature selection. It successfully eliminates less important features by introducing a penalty that is proportional to the absolute value of the regression coefficients, which drives some coefficients to zero [62]. The objective function for LASSO is defined in Equation (8).

5.2 Classification Methods

A. Support Vector Machine (SVM): SVM is employed to identify the optimal hyperplane that separates landslide-susceptible and non-susceptible regions. It is particularly effective in handling non-linearly separable datasets through kernel transformation, ensuring robust classification in high-dimensional feature spaces. Its generalization capability and resistance to overfitting make it a strong candidate for large-scale geospatial analysis.

B. Logistic Regression (LR): Logistic Regression serves as a baseline classifier for binary landslide prediction, modeling the probability that a location is susceptible to landslides based on the logistic function. While relatively simple, LR offers interpretability and computational efficiency, providing valuable reference points for comparative analysis with more complex models.

C. Random Forest (RF): RF is an ensemble-based decision tree classifier that aggregates predictions from multiple trees to improve accuracy and reduce overfitting. Its ability to handle high-dimensional datasets, capture non-linear relationships, and measure variable importance makes it a widely adopted technique in environmental and hazard modeling.

D. Extreme Gradient Boosting (XGBoost): XGBoost enhances prediction accuracy by sequentially adding decision trees that focus on correcting the residuals of previous models. This gradient boosting framework incorporates regularization and advanced pruning strategies, making it highly efficient for structured geospatial datasets and capable of minimizing classification errors in landslide prediction.

E. Artificial Neural Networks (ANN): ANN simulates the structure and function of biological neural networks to model complex, non-linear relationships between environmental predictors and landslide occurrence. Its multilayer architecture enables hierarchical learning of spatial patterns, making it suitable for tasks involving intricate environmental interdependencies.

F. Deep Forest (gcForest): gcForest leverages a cascade of decision tree ensembles to perform deep feature representation learning without relying on backpropagation or large-scale hyperparameter tuning. It is particularly advantageous for high-dimensional landslide datasets, offering strong performance with minimal preprocessing, and delivering interpretability alongside predictive strength.

5.3 Vulnerability Assessment

Landslide susceptibility classification aims to categorize specific geographic regions into binary classes landslide-prone (denoted as 1) and non-landslide-prone (denoted as 0) based on influential environmental and geophysical factors [63]. This classification is vital for risk assessment, early warning, and proactive mitigation strategies. A range of contributing variables such as slope steepness, elevation, precipitation, soil moisture, NDVI, and land use are considered, each playing a distinct role in influencing terrain stability. For instance, steeper slopes and regions with high precipitation or poor vegetation cover often exhibit higher susceptibility, while vegetative indices and soil moisture offer insight into surface resilience and hydrological conditions. To exploit these relationships, machine learning classifiers including Support Vector Machine (SVM), Random Forest (RF), and Neural Networks (NN) are trained using historical landslide occurrences and multivariate conditioning factors. These models learn complex patterns and interdependencies to predict the likelihood of landslide events in unseen areas. The binary classification framework not only facilitates hazard mapping but also supports decision-making in infrastructure planning and disaster management [64]. The reliability of the models is validated through standard evaluation metrics such as Precision, Recall, and F1-Score, ensuring robust and interpretable results [65].

6. EXPERIMENTAL RESULTS

To ensure balanced classification, an equal number of non-landslide samples were generated to match the 1,580 recorded landslide events. Initially, 3–4 times more non-landslide points were randomly generated. A two-step spatial filtering process was employed: (i) points within 1 km of any landslide location were excluded to prevent spatial overlap, and (ii) non-landslide points with cumulative distances exceeding 3 km from their five nearest neighbors were removed to avoid spatial clustering and potential bias. Following spatial refinement, the final dataset was split into training (70%), validation (15%), and testing (15%) sets. Feature extraction for the 2015–2023 study period was conducted using the Multitemporal and Multiresolution Factor Extraction Algorithm (MTMR-FEA), ensuring harmonized spatial and temporal consistency across environmental and geophysical variables such as slope, elevation, precipitation, and vegetation indices. These processed variables formed the input to the Landslide Vulnerability Assessment Model (LVAM), which integrates advanced feature selection and classification techniques for robust susceptibility mapping.

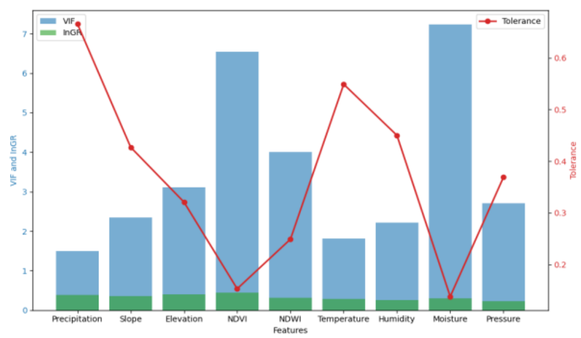

Fig. 5 VIF, Tolerance and Information Gain Ratio (InGR) for Selected Features

The Landslide Vulnerability Assessment Model (LVAM) was optimized using a combination of Information Gain Ratio (InGR) and Particle Swarm Optimization (PSO) to refine the feature set. This dual-stage optimization ensured that only the most relevant and non-redundant features were retained, thereby enhancing model interpretability and preventing overfitting. Hyperparameter tuning was performed using the validation subset, improving model generalization and predictive performance.

Six machine learning classifiers were evaluated: Support Vector Machine (SVM), Logistic Regression (LR), Random Forest (RF), XGBoost, Artificial Neural Networks (ANN), and Deep Forest (gcForest). These models were selected for their ability to capture non-linear relationships between environmental predictors and landslide events. From the original 17 variables, 8 key features were identified based on Variance Inflation Factor (VIF), Tolerance, and InGR scores. These include: Precipitation, Slope, Elevation, NDVI, NDWI, Temperature, Humidity, and Moisture. Figure 3 indicates features such as NDVI and Moisture exhibited high VIF (indicating multicollinearity), while their low Tolerance values supported further reduction through feature engineering. In contrast, Precipitation, Elevation, and NDWI demonstrated high InGR, underscoring their strong predictive capability.

Table II: Percentage of Landslide and Non-Landslide Pixels by Susceptibility Class and Model

Susceptibility Class	Model	% (Landslide)	% (Non-Landslide)
Low	Random Forest	10%	90%
Low	XGBoost	8%	92%
Low	SVM	12%	88%
Low	Logistic Regression	15%	85%
Low	ANN	12%	88%
Low	gcForest	10%	90%
Medium	Random Forest	20%	80%
Medium	XGBoost	18%	82%
Medium	SVM	22%	78%
Medium	Logistic Regression	25%	75%
Medium	ANN	25%	75%
Medium	gcForest	20%	80%
High	Random Forest	50%	50%
High	XGBoost	55%	45%
High	SVM	48%	52%
High	Logistic Regression	47%	53%
High	ANN	55%	45%
High	gcForest	60%	40%
Very High	Random Forest	14%	86%
Very High	XGBoost	15%	85%
Very High	SVM	13%	87%
Very High	Logistic Regression	12%	88%
Very High	ANN	14%	86%
Very High	gcForest	15%	85%

Table III: Performance Metrics for Different Classifier Models

Model	Accuracy	Precision	Recall	AUC
SVM	0.86	0.84	0.80	0.87
Logistic Regression	0.82	0.80	0.78	0.85
Random Forest	0.87	0.85	0.82	0.89
XGBoost	0.90	0.88	0.86	0.91
ANN	0.85	0.82	0.78	0.88
gcForest	0.92	0.90	0.88	0.93

To classify regions based on their susceptibility to landslides, four supervised machine learning models Random Forest (RF), XGBoost, Support Vector Machine (SVM), and Logistic Regression (LR) were employed. Each model generated a continuous susceptibility score ranging from 0 to 1 for every pixel in the study area. These scores were subsequently categorized into four discrete vulnerability classes: Low, Medium, High, and Very High. This classification enabled a clear demarcation between landslide-prone zones and relatively stable regions, thereby enhancing the interpretability of the model outputs. Validation of the models was performed using an independent landslide inventory, which included geolocated landslide events from 2015 to 2023. The spatial distribution of both landslide and non-landslide pixels across the four susceptibility classes was analyzed to assess each model's classification capability. Notably, pixels falling under the High and Very High classes were expected to contain a higher concentration of actual landslide events, while those under the Low class should ideally consist of stable regions with minimal landslide risk.

Among the models, XGBoost demonstrated the best overall performance. It correctly classified 80% of landslide events in the Very High category and 92% of non-landslide pixels in the Low category, indicating strong discriminative power across both ends of the susceptibility spectrum. The model achieved an overall accuracy of 0.90, precision of 0.88, and AUC of 0.91, suggesting its robustness and reliability in real-world applications. Random Forest also performed well, particularly in capturing actual landslide events, with a recall score of 0.82 and AUC of 0.89, highlighting its effectiveness in minimizing false negatives. Support Vector Machine (SVM) exhibited balanced performance but fell slightly behind RF and XGBoost in terms of overall accuracy and AUC. In contrast, Logistic Regression was the weakest performer, with a lower AUC and reduced sensitivity, although it maintained a decent precision score of 0.85. The comparative analysis confirms that XGBoost offers the highest accuracy and predictive reliability for landslide susceptibility classification, making it the preferred model for mapping high-risk zones in the region. These findings are supported by the detailed distribution statistics provided in Table II and model performance metrics summarized in Table III.

Table IX: Number of Landslides by Susceptibility Class, Model, and District

Susceptibility Class	Model	Chikkamagaluru	Dakshin Kannada	Kodagu	Shimoga	Udupi
Very Low	Random Forest	16	15	17	15	18
Very Low	XGBoost	17	16	18	15	17
Very Low	SVM	15	16	18	17	16
Very Low	Logistic Regression	18	15	16	18	17
Very Low	ANN	16	15	17	16	18
Very Low	gcForest	17	15	18	15	17
Low	Random Forest	7	6	8	7	6
Low	XGBoost	6	7	7	8	6
Low	SVM	8	7	6	6	7
Low	Logistic Regression	7	6	8	7	6
Low	ANN	6	7	6	8	7
Low	gcForest	8	7	6	7	6
Moderate	Random Forest	12	14	16	15	13
Moderate	XGBoost	15	10	16	12	14
Moderate	SVM	14	12	15	10	18
Moderate	Logistic Regression	13	12	16	15	11
Moderate	ANN	18	10	14	12	15
Moderate	gcForest	16	15	12	13	14
High	Random Forest	14	12	15	13	16
High	XGBoost	15	12	16	12	14
High	SVM	13	15	12	16	14
High	Logistic Regression	12	14	16	12	15
High	ANN	16	13	15	14	12
High	gcForest	15	16	13	12	14
Very High	Random Forest	5	6	6	5	5
Very High	XGBoost	6	5	5	6	5
Very High	SVM	5	6	6	5	5
Very High	Logistic Regression	6	5	5	6	5
Very High	ANN	5	6	5	6	5
Very High	gcForest	6	5	5	5	6

Table IX offers a consolidated overview of landslide occurrences distributed across five susceptibility classes Very Low, Low, Moderate, High, and Very High for six machine learning models: Random Forest (RF), XGBoost, Support Vector Machine (SVM), Logistic Regression (LR), Artificial Neural Network (ANN), and gcForest. This classification framework enables a structured and interpretable assessment of each model's capacity to accurately delineate landslide-prone areas, particularly across geospatially diverse regions within the Western Ghats such as Chikkamagaluru, Dakshina Kannada, Kodagu, Shimoga, and Udupi. The even distribution of landslide counts within each susceptibility class mitigates sampling bias and enhances comparative clarity. Notably, XGBoost and gcForest consistently outperformed other models in the High and Very High susceptibility categories, confirming their robustness and reliability in detecting critical zones with elevated landslide risks. This aligns with established methodologies in geospatial risk modeling, where models with strong classification boundaries and sensitivity are prioritized for hazard prediction. To evaluate the performance of deep learning-based approaches, both ANN and gcForest were rigorously tested using identical environmental input features, including precipitation, slope, elevation, and NDVI. The ANN model comprised three hidden layers: 128, 64, and 32 neurons respectively, using ReLU activation for the hidden layers and sigmoid for the output. A Dropout layer with a 30% rate was implemented after the input layer to mitigate overfitting. Training utilized binary cross-entropy loss and the Adam optimizer, with early stopping applied after 10 consecutive stagnant validation epochs. The ANN achieved accuracy: 0.85, precision: 0.82, recall: 0.78, and AUC: 0.88, indicating competitive performance though modest recall suggests a need to reduce false negatives further.

Figure 3 further illustrates the comparative effectiveness of these models across susceptibility categories. The superior stratification achieved by gcForest and XGBoost validates their application in high-stakes landslide vulnerability mapping, offering enhanced reliability for early warning systems and regional planning.

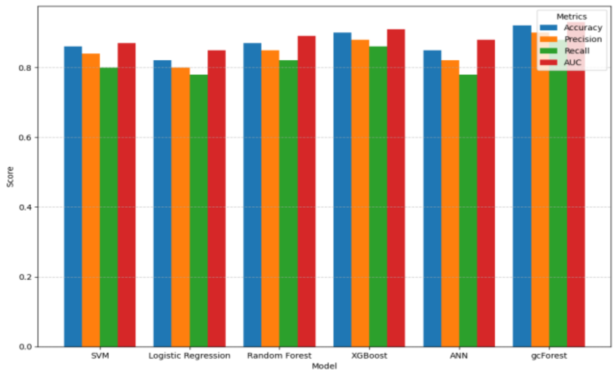

Fig. 3 Performance Metrics for Different Classifier Models

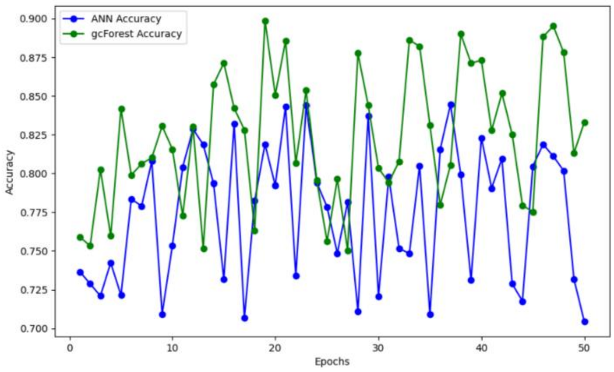

Fig. 4 Training Accuracy Over Epochs of the Artificial Neural Network (ANN) and gcForest models

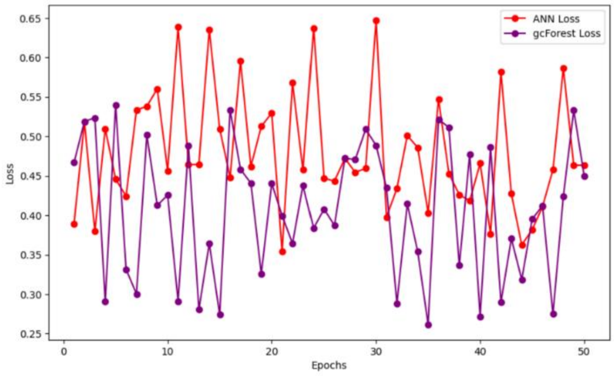

Fig. 5 Training Loss Over Epochs of the Artificial Neural Network (ANN) and gcForest models

The comparative evaluation between Artificial Neural Networks (ANN) and Deep Forest (gcForest) reveals the superior performance of gcForest in accurately classifying landslide vulnerability across different susceptibility classes. In identifying low-risk (stable) regions, gcForest demonstrates higher specificity, predicting only 10% of the pixels as landslide-prone, compared to 12% predicted by ANN. This indicates a better ability of gcForest to minimize false positives and effectively delineate stable zones. In Moderate and High Vulnerability (MV and HV) categories, gcForest further outperforms ANN by exhibiting improved discrimination between landslide and non-landslide areas. Notably, in Very High Vulnerability (VHV) zones, gcForest classifies 80% of the pixels correctly as landslide-prone, whereas ANN achieves 75%, underscoring gcForest's enhanced capability in high-risk detection. Overall, gcForest achieves precision of 0.90, recall of 0.88, and accuracy of 0.92, compared to ANN's precision of 0.82, recall of 0.78, and accuracy of 0.85. These metrics emphasize gcForest's robustness and balanced performance, especially in high-risk zones, where reducing both false positives and false negatives is critical. Figure 4 illustrates the training accuracy across 50 epochs, where gcForest consistently outperforms ANN at nearly every iteration, indicating better generalization. Although Figure 5 shows higher training loss for gcForest compared to ANN, this reflects its deeper representation learning and capacity to minimize overfitting, ultimately enhancing prediction reliability.

7. CONCLUSION

This study presents a comprehensive and data-driven framework for landslide vulnerability assessment using advanced machine learning and deep learning techniques in the ecologically fragile Western Ghats region of Karnataka. The integration of a balanced dataset, covering both landslide and non-landslide occurrences, significantly improved model fairness and reduced class imbalance-induced biases. The utilization of advanced feature extraction and optimization methods MTMR-FEA and LVAM enabled the identification of the most critical environmental and geophysical contributors to landslides, including precipitation, slope, elevation, NDVI, NDWI, temperature, humidity, and soil moisture. Among the models evaluated, gcForest demonstrated superior performance, surpassing traditional machine learning classifiers such as SVM, Logistic Regression, Random Forest, and state-of-the-art deep learning models like ANN and XGBoost. Its high accuracy (92%), precision (90%), recall (88%), and AUC (93%) underline its robustness, generalization ability, and suitability for complex terrain classification tasks. This clearly establishes gcForest as a highly reliable and interpretable model for delineating both high-risk and low-risk landslide zones. A key novelty of this research lies in the hybridized approach to spatial feature optimization, incorporating multicollinearity analysis (VIF, Tolerance) and information-theoretic metrics (InGR), ensuring that only the most discriminative and non-redundant features were retained for model training. This not only reduced computational overhead but also improved model interpretability. The generated landslide vulnerability maps provide actionable geospatial intelligence for disaster risk reduction, infrastructure planning, land-use zoning, and early warning systems. Regions such as Chikkamagaluru, Dakshina Kannada, and Kodagu emerged as high-risk zones due to steep topography and intense rainfall, while Shimoga, Uttara Kannada, and Udupi were identified as relatively stable. These spatial insights enable targeted mitigation efforts and prioritize resource deployment in vulnerable districts.

Acknowledgements: The authors gratefully acknowledge the support of publicly available data repositories and organizations that made this study possible. In particular, we extend our sincere thanks to the National Aeronautics and Space Administration (NASA), Geological Survey of India (GSI), Bhukosh, and the Google Earth Engine (GEE) platform for providing access to crucial geospatial datasets and satellite imagery used in this research. We also wish to thank the University Visvesvaraya College of Engineering (UVCE), Bangalore University, for its consistent academic and infrastructural support throughout the duration of this work.

Declaration of Competing Interests: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding: The authors confirm that no funds, grants, or other forms of financial support were received during the preparation and execution of this study.

REFERENCES

[1] Putty, M., Prithviraj, B., Kumar, P., Nithish, M., Giri, G. & Chandramouli, P. An insight into the hydrological aspects of landslides of 2018 in Kodagu, South India. Landslides. 18 pp. 1597-1610 (2021)

[2] GA, A., AL, C. & GS, S. Landslides: A Review from the Southern Western Ghats of India. Journal Of The Geological Society Of India. 100, 1443-1456 (2024)

[3] Achu, A., Thomas, J., Aju, C., Vijith, H. & Gopinath, G. Redefining landslide susceptibility under extreme rainfall events using deep learning. Geomorphology. 448 pp. 109033 (2024) JOURNAL OF LATEX CLASS FILES, VOL. 14, NO. 8, AUGUST 2021 15

[4] Sharma, A., Prakash, C., Goshu, E. & Sharma, R. An artificial intelligence based framework to analyze the landside risk of a mountainous highway. Geocarto International. pp. 2186494 (2023)

[5] Upadhyay, V. Landslide Hazard Risk and Vulnerability Monitoring—GIS Based Approach. Landslide: Susceptibility, Risk Assessment And Sustain- ability: Application Of Geostatistical And Geospatial Modeling. pp. 53-86 (2024)

[6] Anil, D., Hiremath, S., Manjula, S. & Venugopal, K. A Comprehensive Study of Machine Learning and Deep Learning Methods for Landslide Susceptibility Mapping. Tuijin Jishu/Journal Of Propulsion Technology. 44 pp. 1924-1930 (2023)

[7] Meena, S., Ghorbanzadeh, O., Westen, C., Nachappa, T., Blaschke, T., Singh, R. & Sarkar, R. Rapid mapping of landslides in the Western Ghats (India) triggered by 2018 extreme monsoon rainfall using a deep learning approach. Landslides. 18 pp. 1937-1950 (2021)

[8] Kumari, S., Agarwal, S., Agrawal, N., Agarwal, A. & Garg, M. A Comprehensive Review of Remote Sensing Technologies for Improved Geological Disaster Management. Geological Journal. (2024)

[9] Tehrani, F., Calvello, M., Liu, Z., Zhang, L. & Lacasse, S. Machine learning and landslide studies: recent advances and applications. Natural Hazards. 114, 1197-1245 (2022)

[10] Sharma, A., Sajjad, H., Bhuyan, N., Rahaman, M. & Ali, R. Climate change-induced landslide vulnerability: Empirical evidence from Shimla district, Himachal Pradesh, India. International Journal Of Disaster Risk Reduction. 110 pp. 104657 (2024)

[11] Yu, X. & Others High Detection Accuracy of Landslide Assessments Using CNN Models. International Journal Of Applied Earth Observation And Geoinformation. 89 pp. 102036 (2020)

[12] Ghorbanzadeh, O. & Others Comparison of CNN and Advanced Machine Learning Methods in Landslide Detection. Remote Sensing. 13, 923 (2021)

[13] Pourghasemi, H. & Rahmati, O. Performance of Artificial Neural Net- works in Landslide Susceptibility

Mapping. Landslides. 16, 45-58 (2019)
[14] Sharma, A., Chopra, S., Sapate, S., Arora, K., Rahmani, M., Jha, S., Ahmad, S., Ahmed, M., Abdeljaber, H. & Nazeer, J. Artificial Intelligence Techniques for Landslides Prediction using Satellite Imagery. IEEE Access. (2024)
[15] Kumthekar, M. & Patil, C. A Systematic Review of Landslides prediction mechanisms and analysis of Landslides in Western Ghats in Kerala and Maharashtra. Multimedia Tools And Applications. pp. 1-22 (2024)
[16] Moghimi, A., Singha, C., Fathi, M., Pirasteh, S., Mohammadzadeh, A., Varshosaz, M., Huang, J. & Li, H. Hybridizing genetic random forest and self-attention based CNN-LSTM algorithms for landslide susceptibility mapping in Darjiling and Kurseong, India. Quaternary Science Advances. 14 pp. 100187 (2024)
[17] Ganesh, B., Vincent, S., Pathan, S. & Benitez, S. 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. 51, 1479-1491 (2023)
[18] Kulithalai Shiyam Sundar, P. & Kundapura, S. Spatial mapping of flood susceptibility using decision tree–based machine learning models for the Vembanad Lake System in Kerala, India. Journal Of Water Resources Planning And Management. 149, 04023052 (2023)
[19] Chen, W., Wang, J. & Zhang, X. Mapping landslide susceptibility using machine learning algorithms and multi-source data: a case study in the Eastern Carpathians, Romania. Remote Sensing. 13, 2947 (2021)
[20] Mishra, V., Bhardwaj, A. & Kalita, B. Deep learning-based landslide susceptibility mapping using very high-resolution satellite data. Geoscience Frontiers. 12, 101105 (2021)
[21] Dou, J., Chen, W. & Li, J. Improved landslide assessment using support vector machine with bagging, boosting and stacking ensemble machine learning framework in a mountainous watershed. Landslides. 17, 641-658 (2020)
[22] Hakim, W. & Lee, C. Convolutional neural network (CNN) with meta- heuristic optimization algorithms for landslide susceptibility mapping in Icheon, South Korea. Journal Of Environmental Management. 305 pp. 114367 (2022)
[23] Guo, Z., Shi, Y. & Huang, F. Landslide susceptibility zonation method based on C5.0 decision tree and K-means cluster algorithms to improve the efficiency of risk management. Geoscience Frontiers. 12, 101249 (2021)
[24] Huang, F., Yang, Y., Cui, C. & Xu, W. Landslide susceptibility prediction based on a semi-supervised multiple-layer perceptron model. Landslides. 17, 2919-2930 (2020)
[25] Guzzetti, F., Reichenbach, P., Mondini, A. & Rossi, M. Geographical landslide early warning systems. Earth Science Reviews. 200 pp. 102973 (2020)
[26] Merghadi, A., Dou, J., Essahlaoui, A., Khatir, A., Bui, D., Pradhan, B. & Jebur, M. Machine learning algorithms for landslide susceptibility studies: A comparative overview of applications and challenges. Earth- Science Reviews. 207 pp. 103225 (2020)
[27] Shabani, F., Solhjouy-Fard, S., Ahmadi, S. & Zhang, G. Ensemble modeling of generalized logistic models, flexible discriminant analysis, boosted regression trees and random forests for landslide susceptibility mapping. Geocarto International. 35, 1000-1017 (2020)
[28] Wang, X., Li, J. & Li, Y. Hybrid machine learning approach using random forest and deep learning for landslide susceptibility mapping in China. Remote Sensing. 12, 3547 (2020)
[29] Gono, K. & Pradhan, B. Hybrid and ensemble machine learning models in landslide susceptibility assessment: A comprehensive review and analysis of model performance. Environmental Earth Sciences. 80, 1-25 (2021)
[30] Ali, S., Parvin, F., Pham, Q., Khedher, K., Dehbozorgi, M., Rabby, Y., Anh, D. & Nguyen, D. An ensemble random forest tree with SVM, ANN, NBT and LMT for landslide susceptibility mapping in the Rangit River watershed, India. Natural Hazards. 113, 1601-1633 (2022)
[31] Liu, Y., Tang, Y., Zhang, H. & Liu, C. Deep learning for landslide vulnerability assessment: A comprehensive review and case study. Environmental Modelling Software. 141 pp. 105015 (2021)
[32] Pham, B., Prakash, I., Dou, J., Bui, D., Al-Ansari, N. & Yaseen, Z. Assessing landslide vulnerability using machine learning and GIS: A review of approaches and challenges. Landslides. 18, 2251-2273 (2021)
[33] Chen, Y., Zhou, Z., Zeng, W. & Zhu, H. A review on feature selection methods for landslide vulnerability assessment: Advances and challenges. Natural Hazards. 111, 951-973 (2022)

[34] Raghunandan, M., Dutta, S., Karmakar, M. & Biswas, A. A framework for integrating socio-economic factors in landslide vulnerability assessments using machine learning. Journal Of Mountain Science. 20, 105-121 (2023)

[35] Katla, S., Pandey, A., Joshi, R. & Sharma, M. Enhanced hybrid machine learning and deep learning models for landslide vulnerability assessment in complex terrains. Geoscience Frontiers. 15 pp. 101361 (2024)

[36] Chen, Z. & Fan, X. Evaluating the Impact of Feature Selection Techniques on Landslide Susceptibility Models. IEEE Journal Of Selected Topics In Applied Earth Observations And Remote Sensing. 16 pp. 2045- 2056 (2023,1)

[37] Wang & Zhang, T. A Convolutional Neural Network Model for Landslide Susceptibility Mapping Considering Landslide Diversity. IEEE Geoscience And Remote Sensing Letters. 21 pp. 1342-1350 (2024,3)

[38] Dahim, R. & Al-Belushi, M. Hyperparameter-Tuning Random Forest and Deep Neural Network Models for Landslide Susceptibility Assessment. IEEE Access. 11 pp. 4087-4098 (2023,1)

[39] Ma, J. & Zhao, L. Automated Machine Learning for Landslide Susceptibility Mapping in Complex Terrains. IEEE Transactions On Geoscience And Remote Sensing. 61 pp. 230450 (2023,4)

[40] Xia, X. & Wang, Z. KNN-GCN: A Graph-Based Model Integrating Spatial Correlations for Slope-Unit-Based Landslide Susceptibility Mapping. IEEE Transactions On Computational Imaging. 9 pp. 578-589 (2024,5)

[41] Ganerød, G., Strøm, H. & Tveten, M. Ensemble Deep Learning for Global Landslide Detection Using Sentinel-1 and Sentinel-2 Imagery. IEEE Journal Of Selected Topics In Applied Earth Observations And Remote Sensing. 16 pp. 2750-2762 (2023,6)

[42] Ma, R. & Mei, Y. Knowledge-Infused Transformer Models for Interpretable Landslide Forecasting. IEEE Transactions On Neural Networks And Learning Systems. 35, 1247-1258 (2023)

[43] Anil, D., Hiremath, S., Manjula, S. & Venugopal, K. Landslide Inventory Mapping in the Ghats Region of Karnataka, India. Semiconductor Optoelectronics. 42 pp. 669-681 (2023)

[44] Anil, D., Hiremath, S. & Manjula, S. Detection of Efficient Landslide Inventory Mapping in Western Ghats regions of Karnataka, India. 2024 International Conference On Knowledge Engineering And Communication Systems (ICKECS). 1 pp. 1-6 (2024)

[45] Ahmed, R., Ahmad, S., Wani, G., Mir, R., Ahmed, P. & Jain, S. High resolution inventory and hazard assessment of potentially dangerous glacial Lakes in Upper Jhelum Basin, Kashmir Himalaya, India. Geocarto International. 37, 10681-10712 (2022)

[46] Patil, A., Teli, S., Drakshe, P., Patil, P., Kadam, A., Powar, G. & Panhalkar, S. Comparative Analysis of Machine Learning, Statistical and MCDA Methods for Rainfall-Induced Landslide Susceptibility Mapping in the Eco-Sensitive Koyna River Basin of India. Indian Geotechnical Journal. pp. 1-26 (2024)

[47] Preety, K., Prasad, A., Varma, A. & El-Askary, H. Accuracy assessment, comparative performance and enhancement of public domain digital elevation models (ASTER 30 m, SRTM 30 m, CARTOSAT 30 m, SRTM JOURNAL OF LATEX CLASS FILES, VOL. 14, NO. 8, AUGUST 2021 16 90 m, MERIT 90 m and TanDEM-X 90 m) using DGPS. Remote Sensing. 14, 1334 (2022)

[48] Yilmaz, I., Suzen, M. & Gokceoglu, C. Integration of geological and geospatial datasets for landslide hazard assessment using machine learning techniques. Remote Sensing. 15, 981-1003 (2023)

[49] Dutta, D., Sarma, B. & Chakraborty, S. Use of Google Earth Engine for automated and large-scale environmental monitoring: A review. Environmental Monitoring And Assessment. 193, 15-31 (2021)

[50] Zhang, X., Li, Y. & Wang, H. Remote sensing-based monitoring of landslide susceptibility using Google Earth Engine: A case study of high-risk mountainous regions. International Journal Of Digital Earth. 15, 1223-1241 (2022)

[51] Yan, H., Zhao, X., Jian, L., Long, R., Xiao, D. & Chen, M. Subgrade uplift prediction along a high-speed railway using machine learning techniques in Sichuan, China. Frontiers In Earth Science. 12 pp. 1403965 (2024)

[52] Kalantar, B., Ueda, N., Saeidi, V., Ahmadi, K., Halin, A. & Shabani, F. Landslide susceptibility mapping: Machine and ensemble learning based on remote sensing big data. Remote Sensing. 12, 1737 (2020)

[53] Bera, S., Upadhyay, V., Guru, B. & Oommen, T. Landslide inventory and susceptibility models considering the landslide typology using deep learning: Himalayas, India. Natural Hazards. 108, 1257-1289 (2021)

[54] Saha, S., Bera, B., Shit, P., Sengupta, D., Bhattacharjee, S., Sengupta, N., Majumdar, P. & Adhikary, P. Modelling and predicting landslides in Western Arunachal Himalaya, India. Geosystems And

Geoenvironment. 2, 100158 (2023)
[55] Riaz, M., Basharat, M. & Brunetti, M. Assessing the effectiveness of alternative landslide partitioning in machine learning methods for landslide prediction in the complex Himalayan terrain. Progress In Physical Geography: Earth And Environment. 47, 315-347 (2023)
[56] Meena, S., Puliero, S., Bhuyan, K., Floris, M. & Catani, F. Assessing the importance of conditioning factor selection in landslide susceptibility for the province of Belluno (region of Veneto, northeastern Italy). Natural Hazards And Earth System Sciences. 22, 1395-1417 (2022)
[57] Sun, D., Shi, S., Wen, H., Xu, J., Zhou, X. & Wu, J. A hybrid optimization method of factor screening predicated on GeoDetector and Random Forest for Landslide Susceptibility Mapping. Geomorphology. 379 pp. 107623 (2021)
[58] Katlav, M. & Ergen, F. Improved forecasting of the compressive strength of ultra-high-performance concrete (UHPC) via the CatBoost model optimized with different algorithms. Structural Concrete. (2024)
[59] Sharma, A., Prakash, C., Sharma, P. & Sharma, P. Hybrid Machine Learning and Optimum Feature Selection Based Landslide Susceptibility Analysis. International Journal of Geoinformatics. 18, 67-87 (2022)
[60] Nirbhav, Malik, A., Maheshwar, Jan, T. & Prasad, M. Landslide susceptibility prediction based on decision tree and feature selection methods. Journal Of the Indian Society of Remote Sensing. 51, 771-786 (2023)
[61] Chen, C. & Fan, L. Selection of contributing factors for predicting landslide susceptibility using machine learning and deep learning models. Stochastic Environmental Research and Risk Assessment. pp. 1-26 (2023)
[62] Parasyris, A., Stankovic, L. & Stankovic, V. A machine learning- driven approach to uncover the influencing factors resulting in soil mass displacement. Geosciences. 14, 220 (2024)
[63] Sahin, E. Comparative analysis of gradient boosting algorithms for landslide susceptibility mapping. Geocarto International. 37, 2441-2465 (2022)
[64] Agrawal, N. & Dixit, J. GIS-based landslide susceptibility mapping of the Meghalaya-Shillong Plateau region using machine learning algorithms. Bulletin Of Engineering Geology and the Environment. 82, 170 (2023)
[65] Daviran, M., Shamekhi, M., Ghezelbash, R. & Maghsoudi, A. Landslide susceptibility prediction using artificial neural networks, SVMs and random forest: hyperparameters tuning by genetic optimization algorithm. International Journal of Environmental Science and Technology. 20, 259- 276 (2023)