IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, November 2025
Impact Factor: 7.67

Integrated Machine Learning Framework for Flood and Landslide Hazard Assessment

Nameet Vyavahare, Darshan Chitale, Rajkumar Vishwakarma, Devang Nimbalkar
Department of AI&ML
ISBM College ofEngineering, Nande, Pune, Maharashtra
vyavaharenameet@gmail.com, rk.1947.2003@gmail.com, nimbalkardevang@gmail.com

Abstract: *Among the most devastating natural disasters in South Asia are floods and landslides, which result in numerous fatalities, damage to infrastructure, and disruption of the local economy. Conventional hazard prediction models are unable to keep up with the rapid changes in climate and the increased variability of rainfall. According to recent research, combining many machine learning (ML) approaches greatly increases the accuracy of hazard prediction. The data inputs, feature engineering techniques, and model performance of current ML-based frameworks for flood and landslide hazard assessment are compared in this foundation study. The study provides an overview of well-known methods utilizing Deep Learning architectures, Random Forest (RF), Support Vector Machines (SVM), and Gradient Boosting. It draws attention to present shortcomings in data accessibility, transferability, geographical resolution, and model interpretability. The conceptual underpinnings of the suggested integrated machine learning framework are presented in this base article.*

Keywords: Machine Learning, Flood Prediction, Landslide Susceptibility, Remote Sensing, GIS, Ensemble Learning, DEM, Hydrological Hazards

I. INTRODUCTION

Landslides and floods are among the most common and destructive natural disasters in the world, especially in areas with complicated topography, unstable geological formations, and high rainfall variability. Every year, South and Southeast Asian nations—such as India, Nepal, Bangladesh, Sri Lanka, Bhutan, Indonesia, and the Philippines—are severely impacted, resulting in billions of dollars' worth of infrastructural damage and thousands of fatalities. The likelihood of hazard events is further increased in sensitive areas by increased urbanization, deforestation, changes in land use, and unplanned development. Traditional hazard assessment and early warning systems are insufficient due to the rapid effects of climate change, which have increased monsoon irregularities, extreme precipitation events, soil saturation levels, and hydrological instability.

Statistical models, hydrological simulations, heuristic techniques, qualitative susceptibility mapping, and expert-driven GIS-based assessments have all been used in the past to anticipate hazards. Although helpful, these traditional methods frequently fall short of capturing the intricate, nonlinear interplay between anthropogenic, geological, meteorological, and climatic elements that lead to landslides and floods. For instance, straightforward linear or threshold-based models are unable to adequately capture the link between rainfall intensity, soil infiltration capacity, slope gradient, land cover, drainage density, and saturation dynamics. Furthermore, many emerging locations lack high-quality temporal data, which is necessary for physical hydrological models.

The subject of hazard prediction has seen a substantial transformation with the advent of machine learning (ML). ML algorithms are excellent for assessing flood and landslide susceptibility because they can learn complex patterns from multi-dimensional, multi-source datasets. Many machine learning models have been used extensively in hazard research over the past 20 years, from sophisticated models like Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM) networks, and hybrid ensemble techniques to more traditional algorithms like Random Forest (RF), Support Vector Machine (SVM), and Logistic Regression. The adoption of data-driven techniques has been

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, November 2025
Impact Factor: 7.67

further expedited by the growing availability of satellite imagery, IoT-based rainfall sensors, high-resolution DEMs, and remote sensing data.

Despite these advancements, the majority of current research concentrates on either landslide mapping or flood prediction separately. However, these risks are intrinsically linked in many areas. In addition to causing flash floods, heavy rainfall frequently damages soil structures and causes landslides. The creation of reliable, practical catastrophe management systems is hampered by the absence of integrated models that take into account both risks at the same time. The operational implementation of ML-based hazard frameworks is further hampered by issues such a lack of training data, regional variability, poor model generalization, and the "black-box" nature of deep learning.

The goal of this review paper is to offer a thorough synthesis of current ML-based methods for assessing the risk of landslides and floods. It assesses popular datasets, ML algorithms, feature engineering methods, ensemble tactics, and deep learning architectures. Alongside a discussion of new developments including explainable AI, multi-sensor data fusion, transfer learning, and real-time hazard predictions, a comparative comparison of accuracy, performance, and constraints is provided. Through this study, we draw attention to the crucial research gaps that must be filled in order to create an integrated, multi-hazard machine learning framework that can assist the construction of climate-resilient infrastructure, policy planning, and early warning systems.

II. LITERATURE REVIEW

Over the past 20 years, research on machine learning-based flood and landslide hazard assessment has greatly increased due to the growing availability of remote sensing data, high-resolution terrain models, and sophisticated computational methods. Early research mostly used statistical and heuristic models, but these methods frequently fell short of capturing the nonlinear interactions between geological, hydrological, and environmental elements. Consequently, machine learning (ML) became a more potent and adaptable substitute for multi-factor hazard prediction.

ML techniques for flood vulnerability and forecasting have been the subject of a significant amount of research. Among the most popular classical machine learning models in flood mapping are Random Forest (RF) and Support Vector Machine (SVM). RF and Gradient Boosting may successfully model interactions between rainfall, soil moisture, drainage density, and land-use/land-cover (LULC), obtaining AUC values above 0.90, according to studies like Yaseen et al. (2022). While Artificial Neural Networks (ANN) often perform better with big time-series datasets, SVM has been demonstrated to perform reliably when training data is scarce.

By incorporating long-range temporal correlations, deep learning models—in particular, Long Short-Term Memory (LSTM) networks—have outperformed conventional hydrological models in rainfall–runoff predictions in recent years. Additionally, Sentinel-1 SAR photos have been used to map the extent of floods using convolutional neural networks (CNN) and UNet architectures, allowing for high-resolution segmentation even in foggy conditions.

Much study has been done on landslide risk assessment using machine learning approaches in parallel with flood research. Because RF can handle nonlinear interactions between topographical parameters, soil characteristics, and geological factors, it constantly shows up as a top-performing model. RF is superior to logistic regression, decision trees, and SVM in terms of accuracy and spatial generalization, according to studies by Aditian et al. (2020) and Tien Bui et al. (2019). Slope, aspect, curvature, lithology, proximity to streams, and vegetation indices are frequently used as important factors for landslide mapping in GIS-integrated machine learning systems. CNN-based models have drawn interest for their capacity to more accurately classify spatial patterns and identify landslide scars as high-resolution remote sensing data becomes more readily available. Boundary detection is further improved by UNet segmentation models, which show significant gains over pixel-based classical classifiers.

Additionally, remote sensing has been essential to studies on landslides and floods. Important spectral indices like NDVI and NDWI are provided by optical satellite datasets (Sentinel-2, Landsat), whereas DEM-derived terrain parameters like slope, topographic wetness index (TWI), and stream power index (SPI) are commonly employed as predictors. Because Sentinel-1 SAR data can pierce cloud cover, it has become a primary source for flood detection. Despite these developments, a lot of research still uses a small number of predictors, such DEM, rainfall, and LULC, which shows the need for more thorough multi-source data fusion techniques.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, November 2025
Impact Factor: 7.67

Researchers are actively investigating ensemble and hybrid machine learning models to increase forecasting accuracy. When compared to individual algorithms, ensemble techniques like XGBoost, LightGBM, bagging, and stacking have shown better stability and performance. CNN-RF and LSTM-XGBoost, two hybrid models that combine deep learning and traditional machine learning, have demonstrated encouraging outcomes in capturing both spatial and temporal hazard patterns. The majority of multi-hazard interactions, however, are still poorly understood. Despite the substantial connection between heavy rainfall, soil saturation, and slope failure, very few contemporary research have attempted to jointly predict flood and landslide triggers.

III. METHODOLOGY

The methodology of this review paper focuses on systematically identifying, selecting, analyzing, and synthesizing relevant research on machine learning–based flood and landslide hazard assessment. Since this is a review study and not an experimental research paper, the methodology emphasizes the review protocol, literature sourcing strategy, evaluation criteria, and comparative analysisframework.This section explains how the reviewed studies were collected, what criteria were used for inclusion, which parameters were analyzed, and how findings were compared.

A. Dataset

A critical component of this review involved identifying the datasets most commonly used in previous research. Machine learning models for geohazard prediction rely on several heterogeneous datasets, each contributing unique predictive features. Datasets used :

1. Topographic & Terrain Data (DEM and derivatives)

- Typical sources: SRTM (30 m), ALOS PALSAR (12.5 m), ASTER GDEM, and national high-resolution DEMs or LiDAR where available.
- Why used: DEM-derived products (slope, aspect, curvature, roughness) are primary predictors of both landslides and overland flow patterns for floods.
- Typical resolutions: 10 m to 30 m for regional studies; sub-meter to 5 m for urban/local LiDAR-based studies.
- Common derivatives: slope (% or degrees), aspect, plan/ profile curvature, terrain ruggedness index (TRI), topographic wetness index (TWI), stream power index (SPI), elevation variance.

2. Hydrometeorological Data (rainfall, discharge, soil moisture)

- Typical sources: national meteorological services (e.g., IMD in India), gridded products (TRMM, GPM), reanalysis datasets (ERA5), and local gauge records.
- Temporal scale: hourly/daily for forecasting; monthly/seasonal for susceptibility analyses.
- Variables: accumulated rainfall over multiple antecedent windows (e.g., 1, 3, 7 days), peak intensity, daily totals, river discharge, antecedent soil moisture.

3. Remote Sensing & Spectral Data

- Optical imagery: Sentinel-2, Landsat (multispectral) — used for NDVI, NDWI, LULC classification.
- SAR imagery: Sentinel-1 — preferred for flood mapping because of all-weather capability and backscatter response to water.
- High-resolution commercial imagery: Planet, WorldView for scar detection when fine spatial detail is required.

4. Geological & Soil Data

- Attributes: lithology (rock types), soil texture/class, depth to bedrock, cohesion/frictional parameters where available.
- Sources: national geological surveys, FAO/USDA soil maps, academic field surveys.

IJARSCT

International Journal of Advanced Research in Science, Communication and Technology

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

ISSN: 2581-9429

Volume 5, Issue 2, November 2025

Impact Factor: 7.67

5. Human/Anthropogenic Data

- Examples: land-use/land-cover (LULC), impervious surface fraction, drainage network modifications, roads and cut-and-fill slopes.
- Why: urbanization and land-use change modulate both runoff generation and slope stability.

6. Hazard Inventories and Ground Truth

- Flood inventories: delineated flood extents from SAR or historical records (post-event satellite mapping).
- Landslide inventories: point or polygon locations derived from field surveys, historical records, or manual interpretation of imagery.
- Role: these labeled datasets are the ground truth for supervised learning and for calculating evaluation metrics.

7. Sensor/IoT Data (where available)

- Examples: in-situ soil moisture probes, river-stage sensors, tipping-bucket rain gauges — used in high-resolution, near-real-time systems.

B. Data Preprocessing

Why preprocessing is important?

Before being fed to machine learning models, heterogeneous data sources need to be cleaned (noise, missing values, outliers) and harmonized (projection, resolution, extent). Model repeatability and stability are directly impacted by preprocessing.

A thorough preparation pipeline that is frequently employed in reviewed research

1. Coordinate Systems & Projection

- Reproject all vector/raster datasets to a common coordinate reference system (commonly UTM or WGS84/UTM zone of study area).
- Ensure consistency of units (meters for distances/elevations).

2. Spatial Extent & Masking

- Clip all layers to the study-area bounding polygon (watershed or administrative unit).
- Apply masks (e.g., exclude water bodies for landslide models, or restrict to floodplains for flood susceptibility).

3. Resampling & Harmonization

- Resample rasters to a common grid resolution appropriate for the study (e.g., 10 m for detailed landslide mapping, 30 m for regional studies).
- Resampling methods: bilinear for continuous rasters, nearest neighbor for categorical rasters (LULC, soil type).

4. Noise Reduction and SAR Preprocessing

- For SAR: apply speckle filtering (Lee, Frost) and radiometric calibration, then convert to backscatter sigma naught or VV/VH ratio as needed.
- For optical: atmospheric correction if raw Level-1 data used (e.g., Sen2Cor for Sentinel-2).

5. Temporal Aggregation & Windowing (for time-series models)

- Create lagged rainfall features (e.g., cumulative rainfall 24h, 72h, 7d) and moving averages.
- Align temporal indices for datasets with different timestamps (e.g., hourly vs daily).

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, November 2025
Impact Factor: 7.67

6. Missing Data Handling

- Interpolate missing gauge readings using IDW or Kriging for spatially correlated variables.
- For time-series gaps, use linear interpolation or model-based imputation if gaps are short; discard or flag records if gaps are large.

7. Normalization & Scaling

- Apply min–max scaling (0–1) or z-score standardization depending on model requirements (neural nets typically benefit from standardization).
- Keep transformations invertible (store mean/std) for test-phase and operational deployment.

8. Data Augmentation (for scarce labeled datasets)

- For image-based training: patch extraction, rotation, flipping, random cropping, multi-temporal augmentation.
- For class imbalance (few hazard events vs many non-events): oversampling (SMOTE), undersampling, or class-weighted loss functions.

9. Label Preparation

- For classification: create binary or multi-class labels (e.g., low/medium/high hazard) using thresholds from inventories or expert rules.
- For segmentation: prepare pixel-wise masks from event polygons.

C. Feature Engineering & Extraction (purpose → concrete features and computation)

Why feature engineering matters.

- Good features capture the physical drivers of hazards and often make simpler models perform well. In many geohazard studies, derived indices outperform raw variables.

Common features extracted and how they are computed

1. Terrain-Based Features

- **Slope**: computed from DEM using a 3×3 kernel (degrees or percentage).
- **Aspect**: orientation of slope; encoded as sin/cos components for ML.
- **Curvature**: plan and profile curvature to detect concave/convex surfaces.
- **TWI (Topographic Wetness Index)**: ln(As / tan(b)) where As is upslope contributing area and b is local slope.
- **SPI, TRI**: computed using standard GIS hydrological toolboxes (ArcGIS, SAGA, GRASS).

2. Hydrological Time-Series Features

- **Antecedent Precipitation Index (API)**: aggregated rainfall over past n days (commonly n = 1,3,7,14).
- **Rainfall Intensity Metrics**: peak hourly rainfall, 24-h max, number of rainy hours > threshold.
- **Flow Accumulation & Proximity to Streams**: derived from DEM/hydrology.

3. Spectral & Vegetation Indices

- **NDVI = (NIR - RED) / (NIR + RED)** — vegetation health indicator.
- **NDWI = (GREEN - NIR) / (GREEN + NIR)** — moisture/water presence indicator.
- **Change indices**: differencing NDVI/NDWI across dates to detect land-cover change or post-event scars.

4. Soil & Geology-derived Features

- Categorical or ordinal variables: soil texture class, lithology code, rock strength indicators.

IJARSCT

International Journal of Advanced Research in Science, Communication and Technology

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

ISSN: 2581-9429

Volume 5, Issue 2, November 2025

Impact Factor: 7.67

- If numeric geotechnical parameters (cohesion, internal friction) are available, include directly.

5. Anthropogenic & Land-Use Features

- LULC class (one-hot encoded), impervious surface fraction, distance to roads/settlements.

6. Contextual/Neighborhood Features

- Local statistics computed in moving windows (e.g., mean slope in 3×3 or 7×7 neighborhood), to capture spatial context.

D. Feature Selection and Dimensionality Reduction

Because environmental datasets frequently contain highly linked, duplicated, or noisy variables, feature selection is an essential step in machine learning-based hazard assessment. Model interpretability is improved, overfitting is decreased, computational efficiency is increased, and generalization performance across various geographic locations is improved by keeping only the most pertinent predictors.

The examined studies used a variety of feature selection techniques to find significant predictors. When DEM-derived variables (slope, curvature, aspect) shown strong multicollinearity, filter-based techniques like Pearson/Spearman correlation and Mutual Information were frequently employed to remove redundant features. The impact of each feature to lowering uncertainty in class labels for landslide and flood susceptibility mapping was measured using Information Gain and Gain Ratio.

Studies that needed improved feature subsets for predictive performance employed wrapper-based techniques, most notably Recursive Feature Elimination (RFE) using SVM or Random Forest. Simultaneous model training and feature ranking were made possible by embedded methods like L1-regularized logistic regression (Lasso) and feature significance scores from tree-based models (Random Forest, XGBoost).

Principal Component Analysis (PCA) and other dimensionality reduction techniques were widely used in studies that used high-dimensional satellite images or produced dozens of hydrological indices derived from DEM. Correlated variables could be transformed into orthogonal components with maximum variance retained thanks to PCA. Because convolutional layers automatically train hierarchical features from raw pixel data, feature selection was frequently implicit in deep learning-based image segmentation.

E. Model Selection and Training Strategies

The effectiveness of flood and landslide hazard prediction is largely dependent on the model chosen. Specialized model families are needed for various datasets and types of hazards, and a variety of machine learning and deep learning architectures were employed in the examined studies.

Traditional Models of Machine Learning

Because of its resilience to noise, capacity to manage nonlinear interactions, and interpretability through feature importance.

- **Random Forest (RF)** is the most used model. Grid search was usually used to adjust RF settings like maximum depth and the number of trees (n_estimators).
- **Support Vector Machine (SVM):** Particularly useful for smaller datasets; cross-validation was used to optimize hyperparameters (C, γ) and kernel selection (RBF, polynomial).
- **Gradient Boosting Models (XGBoost,LightGBM):** These ensemble methods showed superior performancefor susceptibility mapping because they capture complex feature interactions. They were often regularized using subsampling and tree-pruning.

Deep Learning Models

- **Convolutional Neural Networks (CNN):** These networks are primarily used to segment floods and detect landslides using satellite data. CNNs eliminate the need for human feature engineering by automatically learning spatial features.
- **UNet and Encoder-Decoder Architectures**: Frequently used for pixel-by-pixel segmentation of landslide or flood damage. Due to class imbalance, studies employed IoU-based loss functions or Dice Loss for training.
- **LSTM Networks:** Used for flood forecasting and rainfall-runoff modeling. By analyzing successive rainfall time series, LSTMs were able to extract long-term temporal dependencies.
- **Hybrid CNN-RF or XGboost**: CNN-extracted spatial features are combined with conventional ML classifiers in hybrid CNN–RF or CNN–XGBoost frameworks to increase interpretability and accuracy.

Protocols For Training:

The following training procedures are involved:

Systematic hyperparameter tuning with Bayesian optimization, random search, or grid search.

Regularization strategies to prevent overfitting include early halting, dropout (for neural networks), and penalty terms.

Class-weighted losses, SMOTE resampling, or probabilistic thresholds can be used to address class imbalance.

The spatial or temporal nature of the hazardbeing modeled, computational resources, input feature types, and data availability all had an impact on the model selection.

F. Model Evaluation and Validation Procedures :

Reliable assessment methods are crucial for determining the actual predictive power of machine learning models, particularly when it comes to hazard prediction, where choices have an impact on infrastructure safety and human life. A range of evaluation techniques, each according to the type of hazard and machine learning model, were employed in the examined studies.

Evaluation Metrics :

Metrics that were often used included:

- AUC-ROC: Measures classification performance across thresholds; most commonly used for susceptibility modeling.
- Accuracy, Precision, Recall, and F1-score are used to classify landslides and floods.
- Kappa Coefficient :By contrasting expected and predicted accuracy, the Kappa Coefficient aids in the assessment of unbalanced datasets.

For segmentation tasks like flood extent mapping, IoU (Intersection-over-Union) and Dice Score are crucial.

RMSE, MAE, and NSE: Used in rainfall-runoff predictions and hydrological modeling.

Validation Techniques

Train-Test Split: A simple yet popular method.

- K-Fold Cross-Validation: Reduces variance and enhances generalization; spatial k-fold was employed in several studies to prevent bias resulting from geographic autocorrelation.
- Temporal Validation: To simulate real-world prediction, forecasting models (like LSTM) were trained using historical data and tested using more recent occurrences.
- Spatial Transfer Testing: To assess a model's capacity for generalization, it was trained in one area and tested in another.

The need of uniform validation techniques in future work is highlighted by the large variation in evaluation quality between studies.

G. Comparative Synthesis and Integration

Synthesizing information from the examined research to create broad insights was the methodology's last phase.

1. Comparison of Studies

Research was contrasted using:

- Metrics for model performance
- Types and resolutions of data
- Workflows for feature engineering
- Variability in geography and climate

2. Thematic Grouping

All of the literature was divided into the following categories:

- Models for forecasting floods
- Models of flood susceptibility
- Models of landslide susceptibility
- Segmentation and detection of landslides
- Multi-hazard and hybrid strategies

3. Finding Recurring Patterns

The synthesis showed:

- For tabular data, RF and XGBoost are consistently powerful.
- CNN is the greatest for visuals.
- For temporal forecasts, LSTM works best.
- Single algorithms are outperformed by ensemble models.
- There aren't many multi-hazard integrated models.

This served as the basis for determining research needs and directing integrated machine learning frameworks' future course.

IV. RESEARCH GAPS

The examined literature identifies a number of enduring gaps, restrictions, and unsolved issues that limit the operational reliability, scalability, and generalizability of current models, despite significant advances in machine learning-based flood and landslide hazard assessment. The need for more reliable, comprehensible, and comprehensive hazard assessment frameworks is highlighted by these research gaps. The main gaps found are described in more detail below.

Lack Of Integrated Multi Hazard Frameworks :

The lack of unified machine learning frameworks that can concurrently model landslide and flood hazards is one of the biggest research gaps. Despite the known causal association between strong rainfall, soil saturation, flash floods, and slope failure, almost all studies concentrate solely on either landslide susceptibility mapping or flood prediction. Current models duplicate effort and have lower predictive ability because they do not take use of shared environmental drivers (such as soil moisture, rainfall intensity, and DEM-derived characteristics). The necessity for multi-hazard, multi-output models that can discover interdependencies among hazards is highlighted by this gap.

Limited Multi-Source and Multi-Modal Data Fusion:

Despite the widespread use of remote sensing and environmental datasets, the majority of research only use a small subset of inputs, such as DEM, rainfall, and LULC. Few incorporate multi-sensor fusion, like merging:

SAR combined with optical imaging

IoT stream gauges and rainfall sensors

Radar rainfall combined with soil moisture probes

DOI: 10.48175/IJARSCT-29780

Satellite series with many temporal dimensions
The models' capacity to represent intricate physical processes is limited by the absence of multi-modal fusion, which also weakens their resilience in severe weather. Advanced fusion techniques including graph neural networks (GNNs), hybrid DL-ML pipelines, and attention mechanisms are needed for future models.

Poor Spatial Transferability of Models :

The spatial dependence of ML models has been noted as a significant difficulty. Even with similar danger features, models trained in one region (such as the Western Ghats) frequently perform poorly when applied to another (such as the Himalayas).
The following causes this lack of transferability:
Variations in terrain structure, rainfall patterns, and geology
Variability in the quality of the data
Overfitting to the environment in the area
This field is severely neglected because there are very few research that examine transfer learning, domain adaptability, or model generalization strategies.

Insufficient Ground-Truth Inventories and Standardization :

Incomplete flood extent maps, unreliable landslide inventories, and non-standard data gathering procedures plague many research. This results in:
Unbalanced classes
Labels that make noise
Inadequate model calibration
Difficulties in comparing findings from different research
To enable repeatable machine learning research, standardized, high-quality, open-source danger inventories with consistent temporal and spatial resolution are obviously needed.

Limited Use of Explainable Artificial Intelligence (XAI) :

The majority of cutting-edge models, particularly deep learning models, are opaque and difficult to understand. Interpretability frameworks like SHAP, LIME, Grad-CAM, or permutation significance were only used in a small number of studies.
Inability to be interpreted:
decreases confidence between disaster management authorities and policymakers
makes identifying model faults challenging.
restricts scientific knowledge of the causes of hazards
To improve decision-making transparency, XAI techniques must be included into hazard prediction systems.

Neglect of Uncertainty Quantification :

Although hazards are inherently linked to uncertainty, the studies reviewed seldom measure prediction uncertainty. Many models generate deterministic results without providing:
Confidence levels
Prediction intervals
Probabilistic hazard maps
Without quantifying uncertainty, the application of ML models in disaster management becomes precarious. Techniques like Bayesian neural networks, Monte Carlo dropout, ensemble variance analysis, and probabilistic modeling are not fully utilized.

Limited Real-time and Operational Deployment

Many research studies focus on developing and validating models but often overlook the challenges of practical implementation. Effective hazard management in real-world scenarios necessitates:

real-time data acquisition

automated preprocessing systems

rapid inference

scalable cloud-based infrastructures

Only a limited number of studies have established near-real-time frameworks utilizing IoT sensors or cloud services like Google Earth Engine. Consequently, there remains a significant gap between academic research and its operational application.

Underrepresentation of High-Resolution, Local-Scale Studies :

While regional studies are prevalent in the literature, there is a scarcity of high-resolution local studies focusing on urban areas, steep landscapes, or isolated regions. The challenges faced include:

a lack of detailed local DEMs

absence of municipal drainage data

limited access to field-verified inventories.

Expanding machine learning research to local scales could greatly enhance micro-level hazard planning and infrastructure design.

Insufficient Evaluation Using Robust Validation Techniques:

A worrying number of studies only use random train-test splits, which don't accurately represent spatial autocorrelation. Model accuracy is frequently exaggerated in the absence of appropriate temporal holdout testing or spatial cross-validation.

Although they are infrequently employed, robust validation techniques like leave-one-region-out CV, spatial block CV, and time-split validation are crucial for reliable hazard modeling.

Lack of Standard Benchmarks and Benchmark Datasets :

Flood and landslide hazard assessment lacks defined benchmark datasets and leaderboards, in contrast to other machine learning domains like computer vision and natural language processing. This makes it impossible to compare meaningfully across:

Algorithms

Geographical areas

Resolutions of data

Methods of preprocessing

In this discipline, developing open benchmark datasets and consistent evaluation procedures is still crucial.

V. RESULTS AND DISCUSSIONS

The combined results of the literature review on machine learning-based flood and landslide hazard assessment are presented in this part. Results in a review paper reflect collective patterns, comparative observations, and performance trends recorded across several studies, in contrast to experimental research where "results" represent outcomes generated by the author's own model. Interpreting these results, highlighting technological advancements, and identifying recurring advantages and disadvantages of current methods are the objectives of the debate.

A. Performance Trends Across Machine Learning Models

In both flood and landslide prediction, machine learning models clearly outperform conventional statistical and heuristic methods, according to the evaluated studies. Random Forest (RF) frequently emerged as a top-performing algorithm among traditional machine learning techniques because of its resistance to noise and capacity to capture

ISSN: 2581-9429
Volume 5, Issue 2, November 2025
Impact Factor: 7.67

nonlinear relationships. In most landslide susceptibility investigations, RF obtained AUC values between 0.85 and 0.95, and it performed similarly well in flood susceptibility mapping.

Support Vector Machine (SVM), on the other hand, performed well, especially in research with smaller, more organized datasets. When dealing with high-dimensional or noisy environmental layers, SVM performance decreased unless considerable feature preprocessing was used.

Due to their capacity to represent intricate feature interactions, Gradient Boosting models (XGBoost, LightGBM) frequently outperformed RF in flood mapping studies, demonstrating greater accuracy. These models showed consistent performance in a variety of geographic locations and had accuracy levels exceeding 90%.

B. Effectiveness of Deep Learning Models :

In research using satellite imagery and temporal rainfall–runoff sequences, deep learning demonstrated significant benefits. When compared to traditional machine learning techniques, CNN-based picture classification and segmentation models consistently generated maps of landslide and flood extent that were more accurate. UNet architectures had high Intersection-over-Union (IoU) scores and accurately captured flood borders even in complex terrain, making them especially useful for mapping flood inundation.

By identifying long-term dependencies in rainfall and streamflow data, LSTM networks fared better for flood forecasting than statistical and regression-based hydrological models. LSTMs showed increased accuracy in forecasting short-term flood development and peak flow episodes.

Nevertheless, extensive training data, GPU resources, and meticulous hyperparameter adjustment were needed for deep learning models. Deep learning notably benefits from a large amount of high-quality labeled data, since studies with restricted training datasets found lower performance or overfitting.

C. Strengths and Limitations of Existing Approaches:

1. Advantages:

Classical hazard mapping techniques are consistently outperformed by ML models.
Even with diverse data, ensemble models (XGBoost, RF) offer consistent performance.
Pixel-level mapping and automated feature extraction are areas where deep learning (CNN, UNet) shines.
Strong flood predicting skills are demonstrated by LSTM networks.
Accuracy is greatly increased by combining different data sources (DEM, rainfall, and SAR).

Restrictions

When applied to new regions, the accuracy of almost all models drastically decreases due to poor transferability.
Despite the fact that floods and landslides are interrelated, most research only model one of these hazards.
The accuracy of many studies is limited by their reliance on unreliable or incomplete inventories.
Large labeled datasets are necessary for deep learning, yet they are frequently unavailable in developing nations.
Because few models incorporate real-time data, their operational viability is diminished.
Reliability in real-world decision-making is limited by the infrequent quantification of model uncertainty.

VI. FUTURE SCOPE

Although there is a lot of room for innovation in the development of machine learning for flood and landslide hazard assessment, there are still a number of undiscovered technological, methodological, and practical directions. More integrated, scalable, interpretable, and operational frameworks that enable practical catastrophe management in addition to advancing scientific understanding must be the focus of future research. The following important next directions are suggested in light of the deficiencies found in this review:

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, November 2025
Impact Factor: 7.67

Integrated Multi-Hazard Frameworks :
The majority of research models landslides and floods independently. Future studies should create integrated, multi-hazard machine learning frameworks that capture common triggers like soil moisture, rainfall, and topographical features. In areas where risks coexist, unified models can increase prediction accuracy.

Multi-Source Data Fusion :
Future research should concentrate on sophisticated data fusion, combining data from UAVs, IoT sensors, radar rains, satellite imaging, and soil moisture probes. Attention networks and multimodal fusion models are examples of deep learning architectures that can better utilize a variety of datasets.

Improved Model Transferability :
The inadequate regional generalization of models needs to be addressed in future research. Methods like domain adaptation, transfer learning, and meta-learning can assist in creating models that function consistently even in regions with less data.

Explainable and Transparent AI :
Models that are both accurate and comprehensible are required. Future systems should use XAI technologies (SHAP, LIME, Grad-CAM) to boost operational adoption, give clear explanations, and increase decision-makers' trust.

Uncertainty-Aware Predictions :
Probabilistic results are necessary for hazard evaluations, not merely deterministic forecasts. To enable risk-informed decision-making, future research should use ensemble variance, Monte Carlo, and Bayesian deep learning to quantify uncertainty.

Real-Time and Scalable Deployment :
Real-time early warning systems require scalable designs that make use of cloud computing, edge inference, and automated pipelines. Future research should focus on deployment-ready frameworks that can be updated continuously in the event of extreme weather.

Enhanced Hazard Inventories :
The availability of standardized, high-quality danger inventories will determine future developments. Open-source datasets, community reporting, and better field data gathering will significantly increase model reliability and lower bias.

VII. CONCLUSION

In order to identify key trends, developments, and constraints, this review study looked at 20 years of research on machine learning-based flood and landslide hazard assessment. In hazard modeling, machine learning has become a revolutionary technique that offers notable advancements over conventional statistical and heuristic methods. While deep learning models, especially CNN, UNet, LSTM, and hybrid frameworks, have enabled high-resolution flood and landslide detection as well as sophisticated rainfall–runoff forecasting, classical techniques like Random Forest, SVM, and Gradient Boosting have shown strong predictive capabilities for susceptibility mapping.

The review identifies a number of enduring issues in spite of this advancement. The majority of current research relies on small datasets, is still exclusively focused on a specific danger, or lacks reliable validation techniques. Major obstacles to real-world deployment include the lack of integrated multi-hazard models, the dearth of standardized hazard inventories, the restricted application of explainable AI, and the underdevelopment of uncertainty quantification. Additionally, the scalability and operational reliability of models are hampered by their low transferability across geographical areas and climate conditions.

The review's conclusions highlight the need for ML frameworks that are integrated, data-rich, explainable, and uncertain. Multi-source data fusion, real-time sensor networks, sophisticated deep learning architectures, scalable cloud-based computing, and transparent AI techniques should all be incorporated into future hazard assessment systems. These developments have the potential to enable precise early warning systems, improve preparedness for disasters, safeguard vulnerable people, and aid in the planning of climate-resilient infrastructure.
In conclusion, there are still significant chances to create cohesive, reliable, and functional multi-hazard frameworks, even though machine learning has significantly improved flood and landslide hazard assessment. To turn ML-based

hazard prediction into trustworthy, useful intelligence for catastrophe risk reduction, it will be crucial to fill in the highlighted research gaps.

REFERENCES

[1] T. T. Pham, D. T. Bui, H. Pradhan, and A. Revhaug, “Machine learning models for flood susceptibility mapping: A review,”*Environmental Modelling & Software, vol. 103, pp. 1–14, 2018.*

[2] J. Yaseen, B. Deo, and M. Kisi, “Hybrid machine learning for flood prediction using environmental and hydrological data,” *Journal of Hydrology, vol. 596, pp. 125–148, 2021.*

[3] S. Kratzert, G. Klotz, K. Brenner, Z. Prettenthaler, and M. Hochreiter, “Rainfall–runoff modelling using LSTM networks,” *Hydrology and Earth System Sciences, vol. 23, pp. 5089–5112, 2019.*

[4] B. T. Bui, A. Hoang, and P. Prakash, “Random forest and SVM-based landslide susceptibility modelling with GIS,” *Geomorphology, vol. 303, pp. 256–270, 2018*.

[5] Y. Chen, W. Li, and Y. Hong, “Deep learning for landslide detection using high-resolution satelliteimagery,” *Remote Sensing of Environment, vol. 247, pp. 112–323, 2020.*

[6] S. Hong, S. Kim, and J. Park, “A convolutional neural network-based flood extent mapping using Sentinel-1 SAR imagery,” *IEEE Transactions on Geoscience and Remote Sensing, vol. 58, no. 8, pp. 5379–5390, 2020.*

[7] P. Sahana, M. Mohammed, and S. Patel, “Ensemble learning techniques for landslide susceptibility mapping,” *Natural Hazards, vol. 110, pp. 1857–1881, 2022.*

[8] K. Chapi, M. Singh, and A. Mergili,“GIS-based landslide susceptibility assessment using multi-source environmental data,” *Geoscience Frontiers, vol. 12, no. 4, pp. 1012–1030, 2021.*

[9] E. B. Abdi and A. Haghizadeh, “XGBoost and LightGBM-based flood hazard modelling for monsoon-dominated regions,” *Water Resources Research, vol. 57, no. 9, pp. 1–18, 2021.*

[10] M. Dou, M. E. Krampe, and A. B. Smith, “UNet-based flood segmentation: A review and benchmark comparison,” *ISPRS Journal of Photogrammetry and Remote Sensing, vol. 182, pp. 216–230, 2021.*

[11] N. Patel and R. Kumar, “Monitoring rainfall-induced landslides using remote sensing and machine learning,” *Landslides, vol. 18, pp. 329–348, 2021.*

[12] A. Bui, N. Khosravi, and H. Shahabi, “Comparative evaluation of SVM, RF, and logistic regression for flood susceptibility mapping,” *Journal of Environmental Management, vol. 247, pp. 104–116, 2019.*