A Data-Driven Approach To Spatial-Temporal Prediction Of Floods And Landslides Using Machine Learning

Neha Gopal N[1], Dr. I Manimozhi[2], Bharath M[3], Dr. Krishna Kumar P R[4], Arun Kumar S[5]
[1]Assistant Professor, Department of CSE, East Point College of Engineering and Technology, Bangalore, India. nehagopal.epcet@eastoint.ac.in
[2]HoD & Professor, Department of CSE, East Point College of Engineering and Technology, Bangalore, India.drmanimozhi.i@eastpoint.ac.in
[3]Department of CSE, SEA College of Engineering and Technology, Bangalore, India. Email: bharathyuvi10@gmail.com
[4]HoD & Professor, Department of CSE, SEA College of Engineering and Technology, Bangalore, India. rana.krishnakumar@gmail.com
[5]Assistant Professor, Department of CSE, East Point College of Engineering and Technology, Bangalore, India. arun.kumar@eastpoint.ac.in

---

Abstract: *This paper proposes an approach grounded in machine learning techniques to predict landslides and food insecurity in susceptible areas. By integrating various datasets, including historical weather patterns, soil composition, land use, topographical data, and socio-economic indicators, the model aims to provide early warnings and actionable insights. The landslide prediction model employs techniques such as Gradient Boosting, Neural Networks and Random Forest to analyze the likelihood of landslide occurrences, while the food prediction model leverages time-series analysis and regression techniques to forecast food production and potential shortages. The models are validated using real-world data from high-risk regions, and the findings indicate a notable enhancement in predictive performance over conventional techniques. The method may contribute to assist policymakers and disaster management agencies in proactive planning, thereby reducing the impact of these events on affected communities.*

Keywords: *Flood prediction, landslide detection, spatial-temporal analysis, CNN, machine learning, early warning system, disaster management.*

---

1. INTRODUCTION

Floods and landslides are devastating forms of natural calamities that cause substantial disruption to ecosystems, human populations, infrastructure, and the economy. These disasters are often triggered by extreme weather conditions such as intense rainfall, rapid snowmelt, or seismic activity, and they are more frequent in recent times. The accelerating impact of weather change and urbanization. According to global disaster reports, emerging countries are mostly susceptible due to limited infrastructure, lack of real-time monitoring, and delayed emergency response systems. Predicting such disasters in advance remains a complex challenge, primarily due to the dynamic interactions between spatial (geographic and topographic) and temporal (time-based) environmental variables. Traditional prediction models rely on statistical or rule-based approaches that struggle to adapt to real-time environmental changes and often result in low prediction accuracy or untimely alerts.

Breakthroughs in Artificial Intelligence and Machine Learning have introduced new opportunities to overcome these limitations. This work introduces a resilient, data-centric model employing Convolutional Neural Networks (CNNs) to forecast floods and landslides across spatial and temporal domains, particularly suited for extracting patterns from image and spatial data. The model is trained using diverse datasets such as rainfall patterns, river discharge levels, soil moisture content, digital elevation models (DEMs), and satellite imagery. These inputs help the CNN learn both temporal trends and terrain and environmental conditions influencing floods and landslides. The system integrates a real-time monitoring component that fetches live meteorological data to make up-to-date risk assessments. A web-based dashboard is also implemented to visualize predictions and communicate early warnings to stakeholders. By combining machine learning with real-time analytics, this project aims to enhance early warning systems, reduce disaster response time, and support efficient disaster risk management strategies. The future solution not only improves analytical accuracy but also validates scalability and adaptability for deployment in high-risk regions worldwide Moreover, the system is designed to continuously learn and improve with new data, ensuring adaptability to evolving environmental conditions. By leveraging cloud-based services, the model supports large-scale deployment and real-time accessibility. This study not only contributes to academic advancements in spatial-temporal modeling but also embraces practical value for government agencies and disaster management authorities. The incorporation of AI into environmental monitoring marks a transformative step toward resilient and data-driven disaster preparedness.

2. LITERATURE REVIEW

The literature survey is a critical component of this research, as it explores previous advancements and existing methodologies in the domain of flood and landslide prediction using ML. Traditional disaster prediction systems often rely on statistical models and are limited by their inability to capture complex, nonlinear spatial and temporal relationships present in environmental data. With the growing sophistication of artificial intelligence, especially in with the evolution of deep learning, there has been a growing reliance on sophisticated architectures such as CNNs and RNNs to enhance predictive precision and dependability. This section explores key research efforts that have shaped the advancement of machine learning techniques in predicting floods and landslides.

Ming-Jui Chang and colleagues developed a forecasting framework for long-term flood mapping by fusing machine learning algorithms with numerical weather prediction tools. Their study underscores the critical role of integrating meteorological and hydrological datasets for precise and anticipatory flood analysis. The system they developed demonstrates improved forecasting capabilities and supports strategic disaster management planning. However, tasks such as dependency on high-quality input data and computational limitations remain significant. Md. Uzzal Mia and his team developed a flood susceptibility model using an advanced deep learning-based iterative classifier optimizer. Their approach leverages deep learning's ability to detect complex patterns and enhances the precision of susceptibility maps, thus aiding in effective flood risk assessment. Despite its promising results, the model's performance is heavily reliant on the diversity and quality of the training datasets, and real-time deployment may be constrained by computational demands.

Pornnapa Panyadee and the OASYS Research Group introduced a spatial-temporal flood hazard mapping system that integrates telemetry data with predictive modelling. This integration allows for real-time flood hazard assessments, significantly contributing to the responsiveness of early warning systems. Although the method improves prediction accuracy, it is sensitive to the availability and reliability of telemetry data and may struggle under varying environmental conditions

In summary, current literature consistently highlights a growing reliance on machine learning and deep learning techniques for forecasting floods and landslides. Empirical results validate the success of CNNs, ensemble classifiers, and integrated approaches in boosting predictive performance and strengthening early warning systems. Nonetheless, limitations related to data quality, model interpretability, and computational complexity present ongoing challenges. These insights form the foundation for the proposed system, which seeks to integrate spatial-temporal data analysis with real-time environmental monitoring to deliver a more robust and adaptive prediction framework.

3. Proposed System

The primary purpose of the suggested model is to improve the forecasting of flood and landslide occurrences by utilizing machine learning to examine diverse environmental, meteorological, and geological inputs. It is designed to handle extensive datasets originating from multiple platforms such as weather monitoring networks, satellite data, IoT-based field sensors, and GIS databases. By processing these heterogeneous sources, the system can detect significant trends and associations that signal impending flood or landslide incidents, enabling timely alerts and proactive response planning.

The initial step involves aggregating detailed environmental data, such as rainfall volume, temperature trends, humidity levels, and soil moisture content. Geophysical factors including land gradient, soil classification, land usage patterns, and altitude are integrated as they are key contributors to geohazard susceptibility. Historical records of past disasters offer valuable insights, allowing the models to recognize recurrence patterns and improve predictive capabilities. High-resolution terrain information from drones and satellite platforms further refines spatial awareness.

After collection, the data is cleaned and pre-processed—this includes handling outliers, filling in missing values, and scaling or transforming data as required. Dimensionality reduction and feature selection strategies like PCA are then applied to simplify model inputs while preserving critical predictive variables such as precipitation behavior, surface moisture, slope inclination, and vegetation cover.

Fundamentally, the framework utilizes machine learning algorithms adept at capturing complex and non-linear relationships between variables. Various models—such as decision trees, random forests, SVMs, neural networks, and gradient boosting methods—are trained on annotated historical datasets to identify patterns indicative of potential geohazards. To further boost performance, ensemble learning methods such as model stacking are used, where predictions from several algorithms are combined and evaluated by a meta-level learner to deliver the final output. This layered modelling structure enhances reliability.

To ensure adaptability, the system supports continuous learning. It integrates new data regularly, re-trains the models, and refines its decision-making logic—allowing the predictive engine to evolve alongside shifting climate and land-use patterns.

4. System Design

The system architecture for the Spatial-Temporal Flood and Landslide Prediction project is designed to integrate multiple data sources, process environmental inputs, and deliver real-time disaster predictions using ML. It consists of modular components that work together to ensure accurate forecasting and user-friendly access through a web interface.

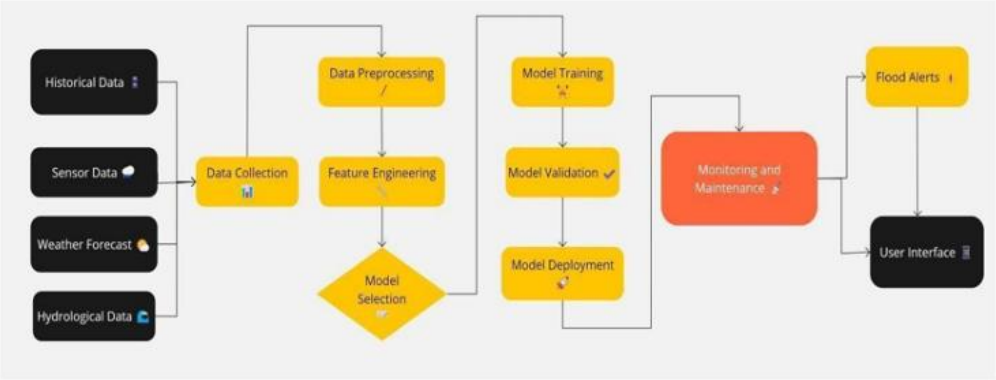


Figure 1. System Architecture for Spatial-Temporal Flood and Landslide Prediction Using ML

1. Data Collection Collects meteorological, hydrological, and historical data from sources like weather APIs and satellite imagery.
2. Data Preprocessing Cleans, normalizes, and transforms raw data into structured formats suitable for model input.
3. Prediction Model (CNN) Processes spatial-temporal data to classify regions into safe or unsafe zones using a trained CNN model.
4. User Interface & Alerts Displays prediction results on a web dashboard and issues alerts for high-risk areas in real time.

A. Landslide Inventory

Create machine learning models to predict landslides caused by rainfall in parts of southwestern Pennsylvania. It focuses on seven counties in that region, and also includes nearby areas in northern West Virginia and eastern Ohio to increase the amount of training data. Landslide records, including event dates, were collected from NASA's COOLR project and Pennsylvania's Department of Transportation (PennDOT). In total, 223 landslide events were used—173 from NASA and 50 from PennDOT—with all events having accurate dates.

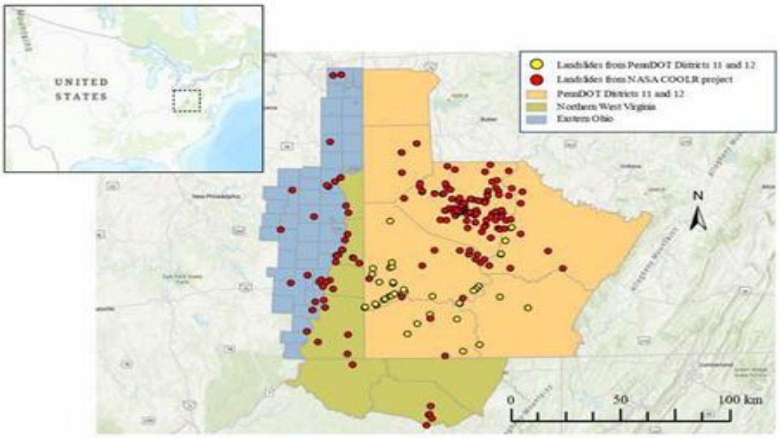


Figure 2: Landslide distribution and study are for spatiotemporal analysis

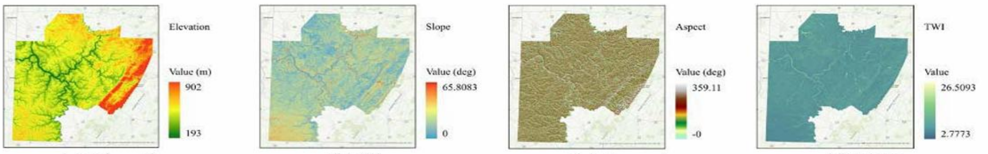
B. Factors for landslides

The aim is to enhance the accuracy of forecasting rainfall-induced landslides in the southwestern region of Pennsylvania. It uses data from seven counties in that region, as well as nearby parts of West Virginia and Ohio, to train machine learning models. Information on 223 landslides was collected from NASA and the Pennsylvania Department of Transportation, focusing on events with accurate dates to build and evaluate the prediction models effectively.

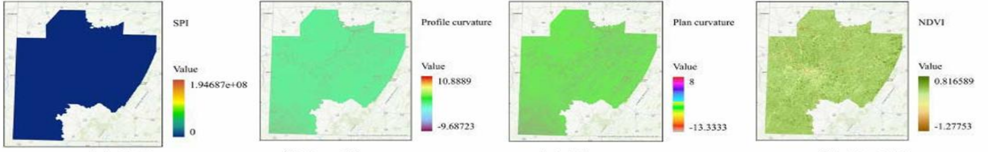
Table 1

No.	Causative Factor	No.	Causative Factor
1	Elevation	8	NDVI (Normalized Difference Vegetation Index)
2	Slope	9	Clay Content
3	Aspect	10	Sand Content
4	Topographic Wetness Index (TWI)	11	Bulk Density
5	Stream Power Index (SPI)	12	Field Capacity
6	Profile Curvature	13	Multi-scale Topographic Position Index (mTPI)
7	Plan Curvature	14	Texture Classification

The study utilized 14 different causative factors to model landslide susceptibility in the target region. These included terrain-related variables such as elevation, slope, aspect, curvature (both profile and plan), and topographic indices like TWI (Topographic Wetness Index), SPI (Stream Power Index), and mTPI (Multi-scale Topographic Position Index). Vegetation influence was assessed using NDVI (Normalized Difference Vegetation Index). Soil characteristics such as clay content, sand content, bulk density, field capacity, and texture classification were also considered. Together, these variables provide a comprehensive dataset for building robust predictive models aimed at rainfall-triggered landslide events.



(a) Elevation (b) Slope (c) Aspect (d) TWI

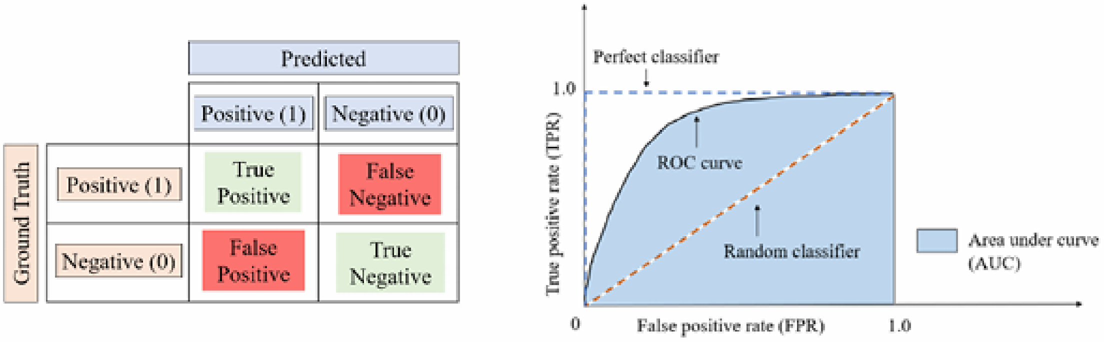


(e) SPI (f) Profile curvature (g) Plan curvature (h) NDVI

Figure 3

C. Evaluation Metrics

To assess landslide susceptibility, four classification models—Logistic Regression (LR), Support Vector Machine (SVM), Random Forest (RF), and Gradient Boosting—were implemented. Model performance was evaluated using five-fold cross-validation, with data partitioned into 80% for training and 20% for validation. The evaluation relied on metrics such as Accuracy and Precision, which are computed from the confusion matrix by comparing predicted outcomes with actual labels. Additionally, the Area Under the ROC Curve (AUC) was used to capture model performance across varying classification thresholds, reflecting the trade-off between true and false positive rates. All indicators are normalized between 0 and 1, where higher values reflect greater predictive capability..



(a) (b)

Figure 4. Illustration of Matrix and ROC Curve

D. Spatiotemporal Sampling Method

This study refines landslide prediction by improving how samples are selected for training machine learning models. Rather than just choosing landslide (positive) and non-landslide (negative) points based on spatial features, it introduces a spatiotemporal approach. Rainfall data—being time-sensitive—is added to better capture the causes of landslides. Non-landslide samples are taken not just from different places, but also from different time periods before a landslide occurs, offering richer patterns. To balance the dataset, the number of negative samples is reduced to match the positive ones, ensuring more accurate and fair model training.

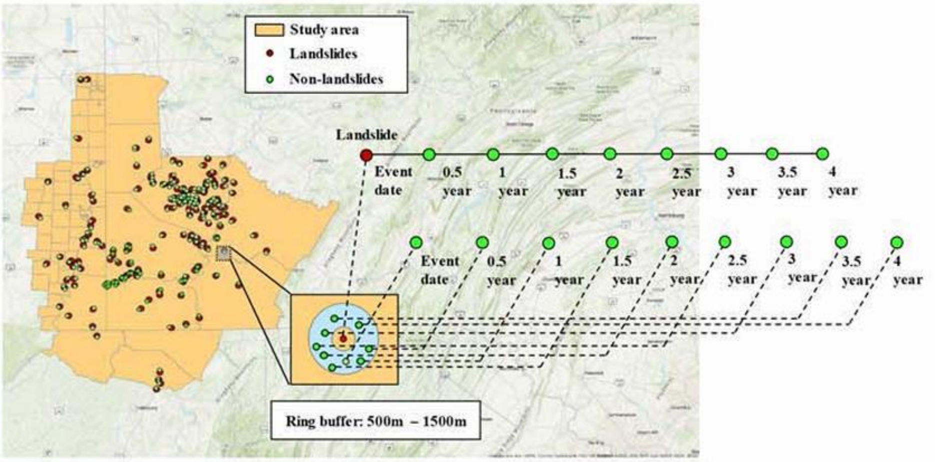


Figure 5. Spatiotemporal Sampling method

E. Results of Spatiotemporal Method

Spatiotemporal database comprising 223 landslide instances and an equal number of non-landslide cases was used to evaluate four machine learning algorithms: Logistic Regression (LR), Support Vector Machine (SVM), Random Forest (RF), and Gradient Boosting Machine (GBM). These models were implemented using the Scikit-learn library, with default hyperparameter settings intentionally retained to limit potential biases from manual tuning.

F. Performance Metrics of ML Models for LSM

Table 2

Model	Accuracy	Precision	Recall	F1 Score	AUC Score
LR	0.73	0.75	0.69	0.72	0.8
SVM	0.73	0.74	0.68	0.71	0.78
RF	0.78	0.79	0.76	0.77	0.86
GBM	0.76	0.76	0.74	0.75	0.84

As outlined in Table 2, the comparative performance of these models reveals that the Random Forest approach consistently outperforms the others. Specifically, RF achieved an AUC score of 0.86, highlighting its strong capability to distinguish between landslide and non-landslide areas across varying classification thresholds. Based on these results, the RF model was selected for further predictive analysis.

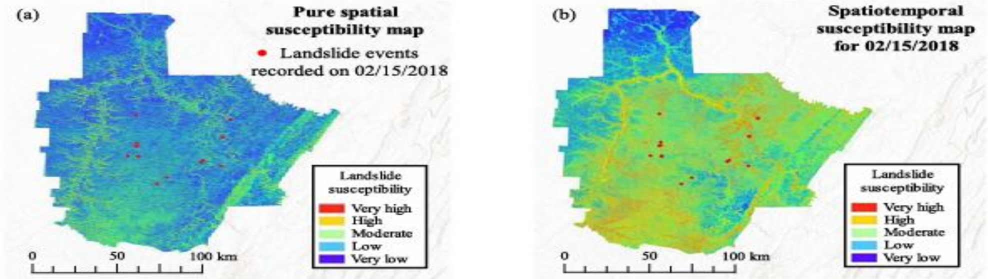


Figure 6. Landslide susceptibility maps based on pure spatial and spatiotemporal ML: (a) Pure spatial map, (b) Spatiotemporal map.

5. Advantages

1. **Non-Invasive and Real-Time Monitoring** The system employs visual perception techniques to detect key movements such as eye blinks or lip motions, enabling continuous, contact-free observation and prompt alert generation when necessary.
2. **Automation and Remote Accessibility** Automated movement tracking reduces reliance on constant manual supervision, while cloud integration allows healthcare providers to remotely access patient data from any location—facilitating timely and informed care.

3. **Enhanced Accuracy and Intelligent Analytics** With approximately 90% detection accuracy, the system ensures reliable monitoring. Its compatibility with advanced machine learning models and predictive analytics boosts its potential to detect early signs of health deterioration.
4. **Scalability and Emergency Responsiveness:** Designed for scalability in multi-patient environments, the system includes robust database functionality and SMS alert mechanisms to support swift intervention during medical emergencies.

6. Challenges Faced

1. **Data Limitations and Synchronization** Reliable prediction is often hindered by inconsistent or missing real-time environmental data, limited availability of labeled training datasets (especially for deep learning models like CNNs), and difficulties in synchronizing heterogeneous data streams from multiple sources.
2. **Computational and Connectivity Constraints** Processing massive spatial-temporal datasets and updating models in real-time demand significant computational resources and consistent internet connectivity—both of which can be limited in practical deployments.
3. **Scalability and Geographic Adaptability** Predictive models trained in one region may underperform in others due to topographical and climatic differences. Ensuring broad adaptability without compromising accuracy poses an ongoing challenge.
4. **System Usability and Deployment Barriers** Designing interfaces that balance user-friendliness with system complexity is crucial for adoption. Moreover, field deployment, particularly in remote or infrastructure-poor regions often face logistical and resource related hurdles.

7. CONCLUSION AND FUTURE WORK

In conclusion, the project “Spatial-Temporal Flood and Landslide Prediction Using Machine Learning” presents a significant advancement in disaster management by integrating real-time environmental monitoring with deep learning techniques, particularly Convolutional Neural Networks (CNNs). By analysing meteorological, hydrological, and topographical data, the system offers high prediction accuracy and timely alerts, thereby enhancing early warning capabilities. Its scalable architecture, user-friendly interface, and adaptive learning features make it suitable for deployment in disaster-prone regions, supporting authorities and communities in mitigating the devastating impact of floods and landslides.

REFERENCES

1.V. T. Hoang, D. S. Huang, and K. H. Jo, “3-D Facial Landmarks Detection for Intelligent Video Systems,” IEEE Trans Industry Inform, vol. 17, no. 1, pp. 578–586, Jan. 2021, doi: 10.1109/TII.2020. 2966513.

2.Shubham Mishra, Mrs. Versha Verma, Dr. Nikhat Akhtar, Shivam Chaturvedi, and Dr. Yusuf Perwej, “An Intelligent Motion Detection Using OpenCV,” Int J Sci Res Sci Eng Technol, pp. 51– 63, Mar. 2022, doi: 10.32628/ijsrset22925.

G. S. Parra-Dominguez, R. E. Sanchez-Yanez, and C. H. Garcia-Capulin, “Facial paralysis detection on images using key point analysis,” Applied Sciences (Switzerland), vol. 11, no. 5, Mar. 2021, doi: 10.3390/app11052435.

3.M. Sanjudharan, M. S. Minu, K. Arun, A. Tiwari, and P. Rampuria “Face Recognition System Based on Haar Cascade Classifier,” International Journal of Advanced Science and Technology, vol. 29.

4.Zhu, J., Wu, J., & Chen, L. (2018). Landslide Prediction Using Deep Learning Models Based on Satellite Data. Journal of Geophysical Research: Earth Surface, 123(8), 1884-1898.

5.Maggioni, V., & Giglio, M. (2020). Improving the Interpretability of Machine Learning Models for Natural Disaster Prediction. AI and Society, 35(4), 987-998.

6.Bui, D. T., & Nguyen, H. T. (2021). Hybrid Machine Learning Models for Natural Disaster Risk Prediction and Management. Journal of Natural Hazards, 106(1), 1-22.

7.V. P. R. N. Kumar, & Y. S. Chandra. (2019). Application of Machine Learning in Flood Forecasting: A Review and Future Prospects. Environmental Science and Pollution Research, 26(8), 7894-7907.

8.Shrestha, S., & Gupta, H. (2018). Application of Machine Learning Algorithms in Flood Prediction: A Review. Journal of Hydrology, 562, 707-717.

9. Jain, S., Singh, P., & Kumar, P. (2019). A Hybrid Model for Flood Prediction Using Decision Trees and Deep Learning. International Journal of Applied Earth Observation and Geoinformation, 75, 1-10.