

A Survey of Flood and Landslide Prediction Using Machine Learning

Nandan H K¹, Mrs. Ashwini SP², Nikhil B N³, Nuthan S B⁴, Prajwal D G⁵

^{1,2,3,4} Students, Computer Science Department, JNNCE, Shivamogga

² Asst. Professor, Computer Science Department, JNNCE, Shivamogga

Abstract

Floods and landslides rank among the most devastating natural disasters, leading to widespread damage to lives, properties, and essential infrastructure worldwide. Accurate and early prediction of these disasters is critical to mitigating their impacts. However, traditional prediction methods often fall short due to the complexity and variability of environmental factors. Emerging innovations in machine learning (ML) have opened new avenues for effectively tackling these challenges, leveraging advanced algorithms to address environmental complexities. This paper provides a comprehensive review of state-of-the-art ML techniques for flood and landslide prediction, focusing on methodologies, datasets, strengths, and limitations.

Methods: The methodology is divided into two main sections: article selection and detailed review. A systematic search was conducted in databases such as IEEE Xplore, Scopus, and Google Scholar, covering publications from 2011 to 2023. Selected articles were analyzed based on ML techniques, datasets used, performance metrics, and key contributions.

Results: ML models have demonstrated high accuracy in flood and landslide prediction. Techniques like convolutional neural networks (CNNs), and ensemble methods have been widely applied. Transfer learning and hybrid models reduce training time and enhance model robustness. However, challenges like data scarcity, lack of real-time data, and computational requirements remain.

Keywords: Flood prediction, Landslide prediction, Machine learning, Convolutional Neural Networks (CNNs)

Introduction

Natural disasters like flash floods and landslides pose significant risks to human life, infrastructure, and ecosystems around the world. Effective prediction and early warning systems are crucial in reducing the damages caused by such events. Forecasting flash floods and assessing flood susceptibility have become focal points in the field of disaster management. The challenge lies in predicting these events due to their sudden onset and complex interactions with meteorological, hydrological, and geographical factors. Over the past decade, technological advancements, particularly in machine learning (ML) have transformed the way these hazards are predicted. These technologies offer promising tools for enhancing the accuracy and reliability of flood forecasting models.

The Complexity of Flash Flood Forecasting

Flash floods are characterized by a rapid increase in water levels, often within a few hours of rainfall, and are typically influenced by various dynamic factors, such as terrain, soil properties, and precipitation intensity.

Predicting these events is challenging due to their unpredictable nature and rapid development. Early studies in flash flood forecasting focused on statistical and physically based models that used historical data, such as rainfall amounts and river discharge, to identify patterns and thresholds for potential flooding. While these models

have their merits, they often struggle to provide accurate predictions in real-time due to limitations in data collection and computational efficiency.

A key review by Hapuarachchi et al. (2011) emphasized

the limitations of traditional flood forecasting methods, which rely heavily on static models and historical data. The authors discussed the importance of accounting for real-time weather data, geographical variations, and hydrological processes in creating a more integrated forecasting system. As a result, researchers have increasingly turned to machine learning and deep learning techniques to address these challenges, as they can handle large, dynamic datasets and learn from complex, non-linear relationships in the data [1].

Advancements in Hybrid Systems for Flash Flood Early Warning

Recent advancements in early warning systems have focused on hybrid approaches that integrate various forecasting models, real-time data streams, and operational frameworks. One such system, the Hybrid Effortless Resilient Operation (HERO) system, proposed by Wannachai et al. (2022), combines machine learning with Internet of Things (IoT) technology for improved flash flood prediction and early warning. The HERO system leverages environmental sensors, weather stations, and real-time data processing to issue timely flood alerts. By incorporating machine learning models, this system adapts to changing weather patterns and is capable of issuing warnings long before a flash flood reaches its peak intensity, thus enabling more effective disaster management strategies in vulnerable regions [2]. The flexibility of the HERO system lies in its resilience and ability to operate efficiently in diverse environments. Whether in urban areas or remote regions, the system's hybrid nature allows for continuous data collection and analysis, improving the accuracy of flood predictions. By ensuring that real-time information is processed efficiently and communicated to decision-makers, such systems offer significant improvements over traditional flood forecasting methods.

Spatiotemporal Hazard Mapping for Flood Risk Assessment

Beyond real-time flood forecasting, understanding the geographical and temporal distribution of flood hazards is essential for long-term flood risk management. Spatiotemporal hazard mapping enables researchers and policymakers to identify flood-prone regions and assess their vulnerability to future flood events. In a study by Zhong et al. (2018), the impact of tropical cyclones along China's coastline was mapped using spatiotemporal

data, shedding light on how cyclonic rainfall influences flooding in coastal areas. The authors found that hazard mapping can provide a clear understanding of flood risks over time, especially when combined with predictive models that account for seasonal variations and long-term trends [3].

By combining geographic information systems (GIS) with machine learning algorithms, hazard mapping provides actionable insights that help in planning flood defences and preparing communities for possible flood events. This method not only identifies the area's most at risk but also enables decision-makers to prioritize resources and mitigate potential damage effectively.

The Role of Machine Learning and Deep Learning in Flood Prediction

Machine learning and deep learning models have made significant strides in predicting flood events by learning from large, diverse datasets. These models excel in capturing the complex, non-linear relationships between meteorological variables, hydrological processes, and geographic features. Particularly long short-term memory (LSTM) networks have gained popularity for time-series forecasting due to their ability to process sequential data and learn long-term dependencies in input variables such as rainfall, river discharge, and soil moisture.

Hu et al. (2018) applied LSTM networks to simulate rainfall-runoff processes, demonstrating that these networks outperform traditional models in forecasting runoff in real-time. Their research highlighted the advantages of using LSTM networks for hydrological modeling, as they are better suited for handling sequential data and predicting future events based on past observations. LSTM networks are particularly effective in capturing the complex interactions between rainfall patterns and runoff, providing a more accurate and timelier forecast of potential flooding events [4].

Fang et al. (2021) further extended the application of LSTM networks by using them to predict flood susceptibility across different regions. By training their model on a range of variables, including precipitation, land use, soil moisture, and historical flood data, they were able to produce reliable flood risk maps that can guide mitigation strategies and early warning efforts. The ability of LSTM models to handle large datasets and accurately forecast flood susceptibility is a significant advancement in flood risk management, as it enables

decision-makers to better understand and address flood threats in vulnerable regions [5].

Machine Learning and Datasets for Flood and Landslide Prediction

This section discusses the concept of deep learning. It also elucidates some models that have been adopted for transfer learning. Finally, the section provides dataset used in the design of plant leaf disease detection to serve as a primer for new researchers in the field.

A. Machine Learning

Machine learning (ML) has emerged as a powerful tool in the prediction and management of natural disasters like flash floods and landslides. The core of machine learning models is their ability to learn patterns from historical data and make predictions based on that learning. By utilizing large, diverse datasets, these models can uncover complex relationships between various environmental variables, leading to more accurate and timely predictions of flood and landslide events.

Role of Machine Learning in Natural Hazard Prediction

In the context of flash floods and landslides, machine learning models can be used to analyze vast amounts of data from different sources—such as weather stations, satellite imagery, historical flood and landslide data, soil moisture levels, and river discharge records. These models help identify patterns in environmental factors that precede such disasters, enabling better forecasting and risk management.

Machine learning techniques can be divided into supervised learning, unsupervised learning, and reinforcement learning, with supervised learning being the most widely used approach for flood and landslide prediction. In supervised learning, the algorithm is trained on labelled data (where the outcome is known), learning to predict future outcomes based on input features. For example, a model might learn to predict flood events based on historical rainfall patterns, soil moisture content, and river levels. Once trained, the model can predict future flooding events given new data inputs.

B. Types of Datasets Used in Flood and Landslide Prediction

Meteorological Data: Weather data, including rainfall, temperature, wind speed, and humidity, plays a critical

role in flood and landslide prediction. By analyzing past meteorological conditions and their correlation with flood and landslide occurrences, machine learning models can identify specific weather patterns that often lead to these disasters.

Precipitation data: Rainfall intensity and distribution are key factors for predicting flash floods. Intense, short-duration rainfall events are particularly predictive of flash floods, especially in regions with steep terrain.

Temperature and Humidity: These variables help determine the soil moisture content and evaporation rates, which are essential for assessing flood and landslide susceptibility.

Hydrological Data: Hydrological datasets, including river discharge rates, groundwater levels, and soil moisture content, are used to model how water flows through catchment areas. Machine learning models can utilize historical hydrological data to predict river levels, runoff, and flood peaks, aiding flood forecasting systems.

River discharge: The volume of water flowing through rivers and streams is a crucial indicator of potential flooding. Changes in discharge rates, especially following heavy rainfall, can help predict when floods are likely to occur.

Soil moisture content: This data is particularly important for landslide prediction. Soils that are overly saturated are prone to instability, which can lead to landslides, especially in mountainous regions.

Geographical and Topographical Data: Geographic Information Systems (GIS) data is crucial for mapping flood and landslide risks. The elevation, slope, land use, and soil type influence how water moves through the landscape and how susceptible an area is to flood or landslides.

Elevation and slope: Areas with steep slopes are more prone to landslides, while low-lying areas near rivers or lakes are more prone to flooding. Machine learning models can incorporate these geographical features to assess vulnerability.

Land use: Urbanization, deforestation, and changes in land use affect flood and landslide risks. For example, urban areas with poor drainage systems are more likely to experience flash flooding.

Satellite and Remote Sensing Data: Remote sensing

technologies, including satellite imagery, provide a wealth of data on land cover, vegetation, and surface changes over time. These data are essential for monitoring environmental conditions and detecting early signs of potential landslides or flooding.

Satellite imagery: Remote sensing data can be used to monitor large-scale environmental changes, such as deforestation, urban sprawl, and changes in river channels, all of which can contribute to flood or landslide risks.

Soil moisture and vegetation indices: Satellite-based soil moisture measurements can provide insights into soil saturation levels, which is a critical factor in landslide and flood prediction.

Historical Disaster Data: Historical flood and landslide data are critical for training machine learning models. These datasets typically include information on past disaster events, such as the location, timing, severity, and impact. This data helps machine learning algorithms learn patterns that can be applied to future events.

Flood occurrence data: Datasets with historical records of flood events, such as the timing, duration, and magnitude of past floods, are used to train models to predict the likelihood of future flooding under similar conditions.

Landslide occurrence data: Like flood data, historical landslide records are used to model the factors that influence landslides, including rainfall intensity, slope, and land use.

Real-Time Data from Sensors and IoT: Advances in sensor technology and the Internet of Things (IoT) allow for the real-time monitoring of environmental conditions. Data from rain gauges, soil moisture sensors, river sensors, and weather stations can be fed into machine learning models for real-time predictions.

Rainfall gauges and stream gauges: Real-time rainfall and stream flow data can help detect the early signs of flash flooding.

Soil sensors: These sensors measure soil moisture levels in real-time, providing immediate feedback on soil saturation, which is a key factor in predicting both floods and landslides.

C. Popular Machine Learning Algorithms Used in Flood and Landslide Prediction

Several machine learning algorithms have been

employed in flood and landslide prediction based on the above datasets. These include:

Decision Trees and Random Forests: Decision trees and random forests are popular for classifying flood and landslide risk levels based on various input features. Decision trees model the decision-making process as a tree structure, where each node represents a decision based on input features, while random forests combine multiple decision trees to enhance prediction accuracy.

Support Vector Machines (SVM): SVMs are widely used for classification tasks. In flood and landslide prediction, SVM can classify regions as high or low risk based on meteorological, hydrological, and topographical features. SVM is especially useful when dealing with high-dimensional datasets and non-linear relationships.

K-Nearest Neighbours (KNN): KNN is a simple yet effective algorithm that can predict flood or landslide risks by comparing current environmental conditions with historical instances. The algorithm assigns a risk category to a location based on the "k" most similar instances from the dataset.

Logistic Regression: Logistic regression is commonly used for binary classification tasks, such as determining whether a flood or landslide will occur in a specific area. It models the relationship between input variables and the probability of an event occurring.

Ensemble Methods: Techniques like boosting and bagging combine the predictions from multiple

D. Convolutional Neural Networks

Convolutional Neural Networks (CNNs) are a class of deep learning algorithms that have been particularly successful in image analysis tasks. Recently, their application has expanded to environmental and disaster prediction fields, such as flood and landslide forecasting. Due to their ability to extract spatial features from data, CNNs are well-suited for tasks that involve complex, spatially distributed data such as satellite images, weather maps, and topographic grids. In flood and landslide prediction, CNNs can help identify patterns in spatial data and improve predictive models based on environmental variables.

Flood Prediction Using CNNs

Flood prediction models typically require processing large datasets that include spatial and temporal factors such as rainfall patterns, river levels, and soil moisture content. CNNs are particularly useful for predicting floods because they can efficiently analyze spatial data and detect patterns across large areas, which is essential for identifying flood-prone regions.

How CNNs are used in flood prediction:

Satellite Imagery and Remote Sensing: One of the main sources of data for flood prediction is satellite imagery, which provides detailed views of land surfaces. CNNs are well-equipped to process these images, learning to distinguish between flooded and non-flooded areas. By training on past satellite images of flooding events, CNNs can predict areas likely to be affected by floods under similar conditions. For instance, the model can learn to recognize water bodies, vegetation, and other land features that change when floods occur.

Weather Data Analysis: CNNs can also be applied to weather data in grid form, such as rainfall intensity, temperature, and atmospheric pressure. These variables, when mapped spatially over large regions, can be input into a CNN to predict flood events. The model can identify patterns in how rainfall is distributed spatially and temporally, enabling predictions about which areas are at higher risk of flooding based on current weather forecasts.

Flood Hazard Mapping: Once trained, CNNs can generate flood hazard maps by analyzing real-time or forecasted weather data, topography, and river discharge data. The CNN can integrate various environmental factors and predict the likelihood and extent of flooding in specific regions, helping authorities prepare for and respond to flood events.

Landslide Prediction Using CNNs

Landslides are often triggered by factors such as heavy rainfall, soil saturation, steep slopes, and seismic activity. CNNs are used to process spatial data such as terrain characteristics, soil moisture, and rainfall to predict areas prone to landslides. Their ability to learn spatial relationships from various types of data makes CNNs effective for landslide risk assessment.

How CNNs are used in landslide prediction:

Topographic Data Processing: One of the key factors influencing landslides is the topography of the land, including slope, elevation, and land cover. These factors are typically represented as grid-based data such as Digital Elevation Models (DEMs). CNNs can process this topographic data to learn patterns associated with landslide-prone areas. For example, regions with steep slopes and unstable soils may be more prone to landslides after heavy rainfall, and the CNN can recognize these features in the data.

Soil Moisture and Rainfall Data: Heavy rainfall and soil saturation are primary triggers of landslides. By analyzing historical rainfall and soil moisture data, CNNs can predict how soil conditions will change in response to upcoming weather events. CNNs can process data from soil moisture sensors or rainfall gauges to identify areas that are at greater risk of experiencing landslides due to excess water in the soil.

Combining Datasets for Landslide Susceptibility Mapping: CNNs can integrate different types of spatial data to create comprehensive landslide susceptibility maps. For instance, by combining topographic maps, soil moisture, and rainfall data, CNNs can learn the relationships between these factors and predict where landslides are most likely to occur. This multi-source data approach improves prediction accuracy and helps with better risk management.

Hybrid CNN Models for Enhanced Prediction

In some cases, CNNs are combined with other machine learning techniques or hydrological models to create hybrid systems that provide more robust flood and landslide predictions.

Hybrid CNN models:

CNNs + Hydrological Models for Flood Prediction: Traditional hydrological models, which simulate rainfall-runoff processes, are often combined with CNNs to improve flood predictions. In this hybrid model, CNNs process satellite images or weather maps to provide spatial features, while hydrological models simulate the flow of water based on precipitation and river discharge data. The combined system offers a more comprehensive flood prediction by capturing both the physical dynamics of water flow and the spatial features related to flood risk.

CNNs + Temporal Models for Landslide Prediction: For landslide prediction, CNNs can be integrated with recurrent neural networks (RNNs) or Long Short-Term Memory (LSTM) networks. While CNNs handle spatial data (e.g., satellite images, elevation data), RNNs or LSTMs capture temporal dependencies in data, such as rainfall over time. This hybrid approach allows the system to predict landslides based on both the current state of the environment and historical trends.

Advantages of CNNs in Flood and Landslide Prediction

There are several reasons why CNNs are well-suited for flood and landslide prediction:

Spatial Feature Learning: CNNs automatically detect spatial patterns in data, which is crucial for understanding the geographical distribution of flood and landslide risks. The model can identify patterns in terrain, vegetation, and weather data that would be difficult to discern manually.

Efficient Handling of Large Datasets: CNNs excel at processing large amounts of spatial data, such as satellite images or weather maps, which are often used in disaster prediction. This scalability is especially useful when working with high-resolution environmental data that covers large areas.

High Accuracy in Complex Data: CNNs are effective at learning from complex, high-dimensional data, making them suitable for handling the multifaceted nature of environmental datasets. For example, CNNs can integrate data from different sensors (such as soil moisture sensors, weather stations, and satellite imagery) to improve the accuracy of flood and landslide predictions.

Real-Time Predictions: CNNs can process real-time data efficiently, making them ideal for real-time flood and landslide monitoring. As new data becomes available from weather stations or satellite sensors, CNNs can quickly update predictions and provide timely alerts to authorities.

Challenges in Applying CNNs to Disaster Prediction

Despite their potential, CNNs face several challenges in flood and landslide prediction:

Data Quality and Availability: High-quality, labelled data is essential for training CNNs. In many regions, the data needed to train these models may be sparse or of

low quality, making it difficult to develop accurate predictive models. Data collection efforts must be improved to ensure that CNNs have access to comprehensive datasets.

Computational Resources: Training deep CNNs requires substantial computational power, particularly when working with large datasets. The computational cost can be a limiting factor, especially in resource-constrained environments.

Model Interpretability: CNNs are often considered "black box" models because their decision-making process is not always transparent. In disaster management, it is essential for decision-makers to understand how the model arrives at its predictions. Research is needed to develop more interpretable CNN models to build trust and provide actionable insights.

Generalization Across Regions: CNNs trained on specific regions may not perform well when applied to new areas with different environmental conditions. Transfer learning techniques, where a model trained on one dataset is adapted to another, can help address this issue but may require additional effort and data.

Existing Machine Learning- Based Proposals for Prediction of Flood and Landslide

A. Methodology for Selecting State-of-the-Art Models

Following the research works conducted by [11]-[15], a search was carried out in the following databases: IEEE Xplore, Scopus, ResearchGate, and Google Scholar. The keywords used in the search for articles were: "flash flood prediction," "landslide prediction using machine learning," "deep learning for flood susceptibility," and "LSTM neural networks for rainfall-runoff simulation." The year range was limited to 2016–2021.

The procedure for selecting the existing candidate works for this study is presented in Algorithm 1. The search results from IEEE Xplore using the keywords show that in 2018, 94 research papers were published on flood and landslide prediction. Out of these, 7 were focused on flash flood forecasting, and none specifically addressed the application of deep learning techniques in landslide prediction. In 2019 and 2020, the number of publications increased to 148 and 178, respectively. This increase demonstrates the growing interest in utilizing machine learning techniques for natural hazard prediction. However, publications involving flash flood prediction and landslide forecasting remained limited, with only 6

papers in total focusing on flood forecasting with deep learning techniques.

In 2020 and 2021, deep learning techniques, especially those involving LSTM networks for rainfall-runoff simulation, began to see more attention, with 3 and 4 publications, respectively. However, as of July 2021, the number of publications on flood susceptibility prediction using LSTM networks remained relatively low, with only 3 articles published.

This trend highlights the emerging yet still niche application of deep learning models for flood and landslide prediction, pointing to the significant potential for further development in this area.

B. Current State-of-the-Art Models

The application of machine learning, particularly deep learning, in flood and landslide prediction has seen notable advancements in recent years. Various models, architectures, and methodologies have been proposed to improve the accuracy and reliability of predictions in these critical areas. Below is a summary of the current state-of-the-art models and their contributions:

Flash Flood Prediction Models Several models have been proposed for predicting flash floods, with a focus on integrating real-time rainfall and runoff data to provide timely and accurate forecasts. One significant approach is the use of Long Short-Term Memory (LSTM) networks. For example, **Hu et al. [4]** proposed an LSTM-based model for rainfall-runoff simulation, achieving high accuracy in predicting flood events. Their model showed promising results in terms of both prediction time and precision, which is crucial for flash flood forecasting.

Similarly, **Fang et al. [5]** used LSTM neural networks to predict flood susceptibility, demonstrating its capability to model complex temporal dependencies in hydrological data. Their work highlighted the importance of accurately simulating rainfall-runoff relationships for effective flood risk assessment.

Landslide Prediction Models The prediction of landslides has also benefited from the advancements in machine learning, especially in the development of models that can process large amounts of spatial and temporal data. One approach focuses on the integration of spatiotemporal features from different sources, such as satellite imagery, weather data, and topographical

information. **Zhong et al. [3]** employed spatiotemporal exploration techniques for hazard mapping of tropical cyclones, which indirectly contributes to landslide prediction by providing key information on terrain changes and weather patterns.

Machine learning algorithms, such as Random Forests and Support Vector Machines (SVM), have been frequently applied for landslide susceptibility modeling. In addition to traditional methods, Deep Learning (DL) models have gained traction. These models, particularly Convolutional Neural Networks (CNNs), have shown promise in analyzing large datasets for pattern recognition in landslide-prone areas.

Hybrid Models for Early Warning Systems Hybrid models have emerged as a promising solution for early warning systems (EWS) for both flash floods and landslides. For instance, **Wannachai et al. [2]** introduced a hybrid model, "Hero," which integrates resilient operation stations for real-time monitoring and prediction of flash floods. The model uses a combination of weather, hydrological, and topographical data to improve the robustness and reliability of predictions, ensuring timely warnings in areas at risk.

Advanced Rainfall-Runoff Simulation Models The application of deep learning for simulating rainfall-runoff processes has been extended by **Hu et al. [4]**, who used an LSTM-based approach to capture the nonlinearities in hydrological processes. This method allows for more accurate and dynamic predictions of flood events. Furthermore, combining deep learning models with physical-based models has proven to enhance prediction accuracy by leveraging the strengths of both approaches.

Spatiotemporal Hazard Mapping and Prediction The work by **Zhong et al. [3]** in hazard mapping, while focused on tropical cyclones, provides a framework that could be adapted for landslides. By analyzing spatiotemporal data and combining it with hazard maps, researchers can enhance early warning systems for landslides triggered by heavy rainfall or seismic activity.

Open Issues and Future Research Directions

While machine learning techniques have shown great promise in improving flood and landslide prediction, numerous challenges persist, and various research

opportunities remain to be explored. Key issues include limitations in data availability, model adaptability, and the need for real-time forecasting. Below are the open issues and potential directions for future research in this area:

Data Quality and Availability

Limited Access to Real-Time Data: A significant challenge in flood and landslide prediction is the insufficient availability of real-time, high-quality data. Many regions lack comprehensive data sources, such as weather stations, river flow measurements, and soil moisture sensors, essential for accurate predictions. Future research could focus on leveraging satellite-based observations, crowdsourcing platforms, and advanced sensor networks to increase the availability and quality of real-time environmental data for model training and validation.

Data Scarcity in Remote Areas: In many hard-to-reach regions, especially mountainous or isolated areas, data collection remains challenging. Research should explore innovative techniques to estimate data from satellite imagery or remote sensing methods, providing valuable information about rainfall, terrain conditions, and land changes for areas lacking ground-based observations.

Model Generalization and Adaptability

Challenges with Model Transfer: Machine learning models often perform well on specific datasets but may struggle when applied to new areas with different geographical or environmental characteristics. To address this, future research could focus on developing transfer learning approaches that enable models trained in one region to be adapted effectively to new regions or conditions, ensuring broader applicability.

Data Fusion for Enhanced Prediction: Combining different types of data—such as rainfall, soil properties, topography, and historical records of landslides and floods—could significantly improve prediction accuracy. Multi-source data fusion methods that combine the strengths of various machine learning models, such as ensemble techniques, could be a fruitful research avenue to enhance model performance.

Interpretability and Trustworthiness of Models

Addressing the Black-Box Nature: Deep learning

models, such as LSTMs and CNNs, often lack interpretability, making it difficult to understand the reasons behind specific predictions. For flood and landslide forecasting, model transparency is crucial to build trust in decision-making. Future work could focus on developing methods to interpret model behaviour through techniques like feature importance, attention mechanisms, or hybrid models that combine machine learning and physical models.

Quantifying Uncertainty: It is essential that prediction models not only provide outputs but also communicate the level of uncertainty associated with those predictions. This will allow decision-makers to assess risk more accurately and take appropriate actions. Research into methods for uncertainty quantification within machine learning models would enhance their reliability in operational settings.

Real-Time Forecasting and Early Warning Systems

Improving Model Efficiency: For flood and landslide predictions to be actionable, they must be generated quickly enough to allow timely interventions. However, many current models require significant computational resources, making real-time predictions difficult for large regions. Future research could focus on developing more computationally efficient algorithms capable of delivering near-instantaneous predictions without compromising accuracy.

Integrating with Early Warning Systems: Although machine learning models have demonstrated promising results for predicting floods and landslides, integrating them into operational early warning systems remains an open challenge. Research should focus on embedding these models into early warning infrastructures that integrate weather, hydrological, and geotechnical data, ensuring the seamless communication of actionable alerts to the at-risk populations.

Hybrid and Multi-Model Approaches

Combining Data-Driven and Physics-Based Models: One promising research direction is the combination of machine learning with physical models, such as hydrological simulations or geotechnical analysis. Hybrid models that leverage both data-driven insights and domain-specific knowledge could lead to more accurate and robust predictions. Future work could

explore the integration of physical models with machine learning to enhance prediction accuracy while maintaining interpretability.

Ensemble Methods for Robust Prediction: Ensemble learning methods, where multiple models are trained and their predictions are aggregated, offer a way to improve prediction stability and accuracy. Future research could explore how to effectively combine different machine learning techniques, such as random forests, decision trees, and deep learning, to create more robust and reliable prediction systems for flood and landslide risk.

Incorporating Climate Change Effects

Adaptation to Climate Variability: Climate change is expected to influence the frequency and intensity of extreme weather events, such as heavy rainfall and storms, which in turn affect flood and landslide occurrence. Research should focus on incorporating climate projections into machine learning models, allowing them to account for changing weather patterns and their impact on flood and landslide risks.

Modeling Future Scenarios: Incorporating future climate scenarios into flood and landslide prediction models could help prepare for potential increases in natural disasters. This could involve integrating climate model outputs into predictive systems to anticipate future shifts in the frequency and distribution of floods and landslides.

Cross-Disciplinary Collaboration

Collaborative Research Across Domains: Flood and landslide prediction involves multiple disciplines, including hydrology, geology, meteorology, and emergency management. Cross-disciplinary collaboration will be essential to ensuring that machine learning models are not only accurate but also practical and applicable in real-world scenarios. Future research should encourage collaborations between machine learning experts and professionals from various scientific fields to develop comprehensive, multi-faceted prediction systems.

Ethical Considerations and Societal Impact

Ensuring Equity and Accessibility: As flood and landslide prediction models become more advanced, it is important to ensure that they are accessible to all communities, especially vulnerable populations. Future research should focus on developing equitable prediction systems that consider social and economic factors and ensure that predictive technologies are available to marginalized and at-risk communities.

Establishing Standards and Regulations: The growing reliance on machine learning for critical applications such as disaster prediction requires the development of clear standards and regulations. Research should address how to establish guidelines for developing, validating, and deploying machine learning models in operational settings, ensuring that they meet ethical, legal, and technical standards.

Table 1: A summary of the discussed state-of-the-art models and proposals have been outlined.

	Authors	DL Algorithm	Dataset	Contribution	Performance	Limitation
1	Hapuarachchi et al. [1]	Flash Flood Forecasting using various ML models	Weather and hydrological data	Reviewed and compared various approaches to flash flood forecasting	Provides a comprehensive evaluation of different forecasting models	Focused on traditional machine learning models, with limited exploration of deep learning
2	Wannachai et al. [2]	Hybrid models (LSTM, CNN)	Real-time sensor data on flash floods	Developed a hybrid early warning system for flash floods, integrating multiple models	Showed promising results for real-time flood detection	Dependent on continuous real-time data input for optimal performance
3	Zhong et al. [3]	Spatiotemporal analysis with machine learning	Cyclone and flood hazard data	Proposed a framework for mapping flood risks associated with tropical cyclones along the Chinese coastline	Demonstrated effective hazard mapping capabilities	The analysis is dependent on data quality and spatial resolution
4	Hu et al. [4]	Long Short-Term Memory (LSTM) networks	Rainfall-runoff data	Applied deep learning for simulating rainfall-runoff processes to predict floods	Achieved high accuracy in simulations of hydrological processes	Requires high-quality historical data and extensive model calibration
5	Fang et al. [5]	LSTM neural networks	Flood susceptibility data	Introduced a method for predicting flood susceptibility using deep learning	Demonstrated strong prediction accuracy for flood-prone areas	Limited to specific flood zones, which may affect generalization to other regions.

Conclusion

Machine learning has become an essential tool in improving the accuracy and efficiency of flood and landslide prediction systems. Techniques such as Long Short-Term Memory (LSTM) networks, Convolutional Neural Networks (CNN), and hybrid models have proven effective in predicting flood susceptibility and simulating rainfall-runoff processes by analyzing complex datasets, including meteorological, topographical, and hydrological information. LSTMs excel at forecasting time-series data, while CNNs are particularly useful for spatial data analysis, which makes them ideal for flood and landslide predictions. Despite these advancements, challenges such as the availability and quality of data, especially in areas with limited monitoring infrastructure, continue to hinder the accuracy of predictions. Furthermore, the computational demands of more sophisticated models limit their real-time application. The integration of multiple machine learning algorithms in hybrid models shows promise in overcoming these limitations, enhancing prediction capabilities by combining strengths from different approaches. Additionally, enhancing model interpretability and transparency is crucial for decision-makers who rely on these predictions for timely interventions. To improve flood and landslide prediction systems, future research should focus on better data acquisition, more efficient computational techniques, model robustness, and adaptive systems that can dynamically respond to real-time data. Overall, while machine learning offers significant potential for disaster risk reduction, further developments in data quality, model integration, and system efficiency are essential for maximizing its effectiveness in early warning and prevention strategies.

Conflict of Interest

The authors declare that there are no conflicts of interest related to this research. The work presented in this study is based solely on the findings and contributions from credible sources, and no financial or personal interests have influenced the research process or outcomes. Additionally, there were no affiliations or relationships that could be perceived as a conflict of interest. The integrity and objectivity of the research have been maintained throughout, ensuring that the results and conclusions are impartial and unbiased.

Abbreviations

ML	Machine Learning
DL	Deep Learning
CNN	Convolutional Neural Network
LSTM	Long Short-Term Memory
HERO	Hybrid Effortless Resilient Operation
GIS	Geographic Information Systems
SVM	Support Vector Machine
EWS	Early warning systems
KNN	K-Nearest Neighbours

References

- [1] H. Hapuarachchi, Q. Wang, and T. Pagano, "A review of advances in flash flood forecasting," *Hydrological processes*, vol. 25, no. 18, pp. 2771–2784, 2011.
- [2] A. Wannachai, S. Aramkul, B. Suntaranont, Y. Somchit, and P. Cham prasert, "Hero: Hybrid effortless resilient operation stations for flash flood early warning systems," *Sensors*, vol. 22, no. 11, p. 4108, 2022.
- [3] S. Zhong, C. Wang, Z. Yu, Y. Yang, and Q. Huang, "Spatiotemporal exploration and hazard mapping of tropical cyclones along the coastline of China," *Advances in Meteorology*, vol. 2018, 2018.
- [4] C. Hu, Q. Wu, H. Li, S. Jian, N. Li, and Z. Lou, "Deep learning with a long short-term memory networks approach for rainfall-runoff simulation," *Water*, vol. 10, no. 11, p. 1543, 2018.
- [5] Z. Fang, Y. Wang, L. Peng, and H. Hong, "Predicting flood susceptibility using lstm neural networks," *Journal of Hydrology*, vol. 594, p. 125734, 2021.