Flood and Landslide Prediction using Machine Learning

Navaneeth KB, Rohith R, Sabarinath AM, Varun V, Mrs. Jisha C Thankappan
Dept. of CSE, College of Engineering Kidangoor, Kottayam, Kerala, India navaneethkb1707@gmail.com, rohith712004@gmail.com,
sabarinathm22@gmail.com, varunv1324@gmail.com, jisha.ct@ce-kgr.org

***Abstract*—Floods and landslides are highly destructive natural disasters, causing severe damage to lives, infrastructure, and economies. Their increasing frequency and intensity, driven by climate change, highlight the urgent need for advanced predictive systems to mitigate their impact. This survey examines the application of machine learning (ML) techniques for flood and landslide prediction, utilizing diverse data sources such as meteorological records, soil conditions, topography, remote sensing imagery, and historical incidents.**

Various ML models, including Random Forest (RF), Convolutional Neural Networks (CNNs), and Attention-UNet, are reviewed for their effectiveness in risk assessment, spatial mapping, and prediction accuracy. Traditional models like RF provide robustness and simplicity, while advanced architectures like Attention-UNet excel in capturing complex spatial dependencies, making them ideal for high-resolution disaster mapping. Hybrid and ensemble models further enhance prediction reliability by overcoming the limitations of individual techniques.

The integration of real-time sensor data and transfer learning improves model adaptability to dynamic and data-scarce environments. These systems offer actionable insights, empowering policymakers and emergency responders to optimize resource allocation, plan mitigation strategies, and enhance disaster preparedness. Moreover, ML applications in disaster management highlight the potential of interdisciplinary approaches, combining geospatial analysis, environmental science, and artificial intelligence. This survey underscores the transformative potential of ML in advancing flood and landslide prediction. By addressing challenges like data scarcity and computational complexity, it aims to support the development of more accurate, scalable, and efficient disaster management solutions essential for building resilient communities in an era of increasing environmental risks

***Index Terms*—Machine Learning, Disaster Prediction, Risk Assessment**

I. INTRODUCTION

Traditional machine learning and deep learning mod- els are usually designed to operate independently, focus- ing on specific tasks. However, when feature distribu- tions change, these models require complete retraining. Transfer learning helps overcome this limitation by allowing knowledge from one task to be applied to another, improving performance.

Machine learning offers powerful tools for analyzing large datasets from diverse sources, including meteorological data, terrain features, soil properties, and historical disaster records. By utilizing techniques such as random forests and neural networks, our project seeks to enhance flood and landslide detection, ultimately strengthening disaster prediction capabilities. The ability of machine learning to uncover complex patterns and correlations within data enables insights that conventional methods may miss.

Beyond immediate disaster response, our project emphasizes long-term resilience planning. By examining historical disaster trends, we can identify recurring patterns and devise mitigation strategies that support urban development and infrastructure enhancements. This holistic approach to disaster management focuses not only on emergency response but also on proactive prevention and preparedness.

Our primary objective is to reduce the impact of natural disasters on vulnerable communities by enhancing prediction accuracy and improving decision-making processes. By integrating machine learning into flood and landslide forecasting, we aim to transform disaster management approaches and bolster community resilience against increasing environmental threats. Through data-driven strategies, we aspire to build a safer future where at-risk communities are better equipped to cope with the challenges of climate change.

II. LITERATURE SURVEY

Machine learning has emerged as a powerful approach to natural disaster prediction, with researchers exploring diverse methodological strategies. Tehrani [1] highlight the transition from labor-intensive manual image analysis to automated landslide detection using machine learning techniques. Meena [2] introduced the U-Net architecture, originally designed for biomedical image segmentation, demonstrating its effectiveness in pixel-level landslide detection. Heo [3] developed a comprehensive model for flood and landslide risk assessment in Indonesia's Kalimantan and Sumatra regions, utilizing Random Forest to classify regions into four risk levels: low, moderate, high, and very high. Mosavi [4] explored various machine learning models including Artificial Neural Networks (ANNs), Support Vector Machines (SVM), Decision Trees (DT), and Adaptive Neuro-Fuzzy Inference Systems (ANFIS), emphasizing the importance of data preprocessing and the challenges of model generalization.

The studies reveal a range of machine learning approaches across different contexts. Harsh [5] leveraged Apache Spark to analyze 30 years of historical rainfall data, achieving an impressive 93.4% accuracy using the random forest algorithm. Khan [6] conducted a comprehensive study on the Indus River Basin, comparing multiple machine learning models including ANN, KNN, SVM, Naive Bayes, and Random Forest. Nguyen [7] provided a systematic review of flood prediction methods, classifying approaches into hydrologic models and machine learning models, and advocating for hybrid techniques that combine both approaches. Huu [7] evaluated four machine learning algorithms for landslide susceptibility prediction, with Random Forest demonstrating the highest accuracy at 79.19% and an Area Under the Curve (AUC) of 0.76. Nguyen et al. [8] focused on landslide susceptibility mapping in Pithoragarh District, India, finding that Linear Discriminant Analysis (LDA), Logistic Regression (LR), and Radial Basis Function Network (RBFN) could effectively predict landslide-prone areas.

The literature review also highlighted critical insights and future research directions. Ahmed [8] conducted a groundbreaking study on landslide risks in refugee camps in Bangladesh's Cox's Bazar district, utilizing geospatial data and machine learning models to assess infrastructure vulnerability. Kappi [9] conducted a bibliometric analysis revealing the significant impact of machine learning methodologies, particularly convolutional neural networks (CNNs), decision trees, and ensemble methods. The research consistently emphasizes the potential of interdisciplinary approaches, integrating computer science with environmental sciences, geology, and telecommunications. Challenges remain, including limited high-quality datasets, computational intensity of complex models, and the need for location-specific adaptability. Future research directions include developing transfer learning techniques, improving model interpretability, and creating more precise, real-time predictive systems. Remote sensing, geographic information systems (GIS), and Internet of Things (IoT) devices are emerging as crucial technologies for enhancing disaster prediction capabilities, with the ultimate goal of supporting sustainable urban development and improving climate resilience.

The depth of machine learning applications in dis- aster prediction continues to expand, with researchers exploring increasingly sophisticated methodologies. Huu [7] developed a comprehensive approach to landslide susceptibility modeling, comparing four distinct machine learning algorithms: Logistic Regression (LR), Multi-Layer Perceptron (MLP), Gradient Boosted Trees (GBT), and Random Forest (RF). Their study revealed nuanced insights into model performance, with Random Forest emerging as the most effective method, achieving an impressive accuracy of 79.19% and an Area Under the Curve (AUC) of 0.76. Nguyen [7] further demonstrated the potential of single machine learning models, focusing on Logistic Regression, Linear Discriminant Analysis, and Radial Basis Function Networks for landslide susceptibility mapping in Pithoragarh District, India. Their research highlighted the importance of considering local conditions and specific geographical factors in developing predictive models.

The future of machine learning in disaster prediction lies in addressing current limitations and pushing technological boundaries. Mosavi [4] identified key challenges, including the generalization of ML models and the need for high-quality, diverse datasets. Researchers are increasingly focusing on hybrid modeling approaches that combine machine learning techniques with traditional hydrologic models to improve predictive accuracy. Nguyen [10] advocated for ensemble machine learning approaches to reduce prediction uncertainty and improve generalizability. The studies consistently point to several critical research directions: developing transfer learning techniques to adapt models to new locations with limited data, improving model interpretability for policy decision-making, and integrating machine learning with physical models to enhance prediction accuracy in complex environmental scenarios. The ultimate goal remains creating more precise, real-time predictive systems that can provide early warnings, support sustainable urban development, and enhance climate resilience across diverse geographical contexts..

III. METHODOLOGY

Our methodology follows a systematic approach for flood and landslide prediction using machine learning. We collect diverse datasets, including meteorological records, soil data, topography, remote sensing imagery, and historical incidents. Data preprocessing involves cleaning, normalization, feature selection, and spatial mapping. Various ML models, such as Random Forest, CNNs, and Attention-UNet, are employed for risk assessment and prediction, with hybrid models enhancing accuracy.

SYSTEM ARCHITECTURE

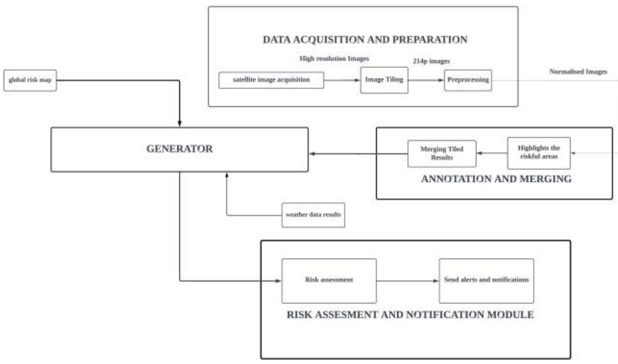

Fig. 1. System Architecture

The Project has been mainly divided into four mod- ules.

1) *Data Acquisition and Preparation:*

- Functionality : Fetches images from Plaftorms like LandSat,GoogleEarth.
- Features : This module contains submodules which include satellite image acquisition ,image tiling and preprocessing.
- Image tiling: The images from the landsat or google earth engine are in high resolution format which are converted into 214p images.
- Preprocessing: Here the images are made more clear and more accuarate,and it outputs normalised images.

2) *Annotation and Merging:*

- Merging Tiled Results: Processed image tiles are combined to reconstruct large-scale maps.
- Risk Identification: The system highlights areas prone to floods or landslides based on ML predic- tions.
- Spatial Alignment: Ensures different datasets (im- agery, historical records) align correctly.
- TensorFlow Integration: Deep learning models pro- cess the data to detect potential disaster-prone zones.

3) *Generator:*

- Machine Learning Processing: ML models analyze data to predict flood and landslide risks.
- Risk Map Creation: A flood/landslide risk map is generated to visualize high-risk zones.
- Integration with Global Risk Map: The local risk map is aligned with global disaster databases.
- Weather Data Incorporation: Real-time weather data enhances prediction accuracy.
- Dynamic Updates: The system updates risk maps based on incoming new data.

4) *Risk Assessment and Notification Module:*

- Risk Evaluation: The system determines the severity of potential disasters based on ML predictions.
- Threshold Analysis: Compares predictions against predefined risk levels to trigger alerts.
- Automated Notifications: Sends warnings to rele- vant authorities and local populations.

IV. IMPLEMENTATION

1) *Hardware Implementation:*

- Our project requires specific hardware resources for efficient training, deployment, and storage. For training our model, we utilize Google Colab's free T4 GPU, which significantly accelerates computa- tion and reduces processing time
- During deployment, a minimum of an Intel i5 or AMD Ryzen 5 processor is required to ensure smooth execution, with 8GB of RAM as the base- line and 16GB recommended for optimal perfor- mance—Colab provides up to 15GB of RAM for training
- For storage, we rely on Google Drive (500MB free space) to store datasets and model checkpoints, while at least 10GB of free space on a local machine is necessary for deployment-related files and processing
- These hardware resources collectively enable effi- cient model training, seamless real-time analysis, and reliable data management in our project.

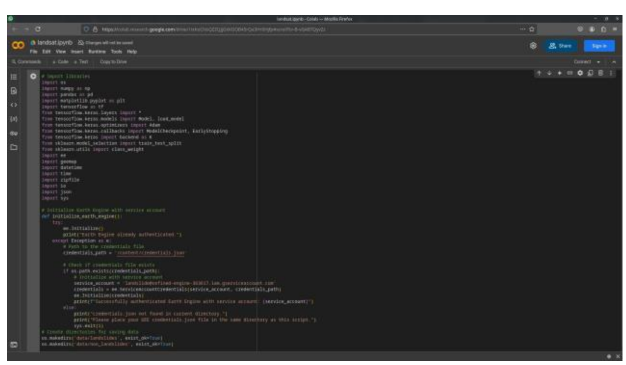

Fig. 2. Google Colab

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 06 | June - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

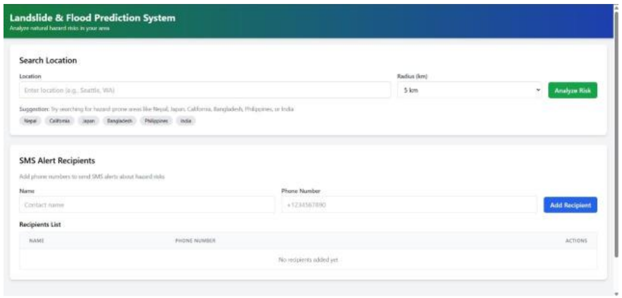

Fig. 3. Opening nterafce

2) Software Implementation:

- We use Python 3.7 and above, which comes pre-installed in Google Colab, ensuring compatibility with all required libraries
- TensorFlow 2.x is essential for training and deploying the Attention U-Net model, while Scikit-learn aids in preprocessing, evaluation metrics, and other machine learning utilities
- Geemap is integrated for geospatial data visualization and processing, and Scipy provides scientific computing functions and optimization techniques
- Additionally, we require API keys and credentials: a Google Earth Engine service account (credentials.json) for accessing satellite data and a Google Maps API Key for geocoding functionalities in the Flask application
- These software components collectively support our project's data processing, model training, and deployment workflows.

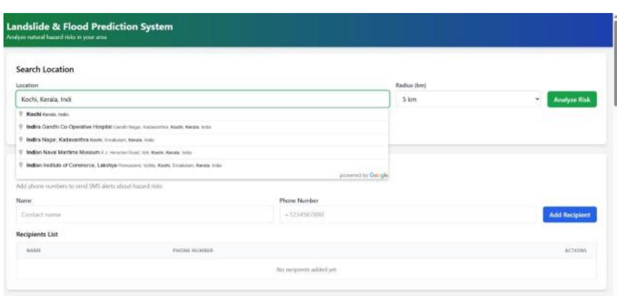

Fig. 4. Location Search

3) System Workflow:

- Satellite Image Collection:Gathering high-resolution satellite images from sources like Google Earth .
- Image get normalised by removing noise, and generating labeled masks for training.
- Attention U-Net marks high-risk regions prone to floods and landslides.
- Merging weather, topographical, and satellite data to enhance prediction accuracy.
- Risk maps are generated by using data from the satellite images and weather data.
- Based on the results ,we assign risk levels as low,medium and high .
- After final assessment real-time notifications are sent for users and authorities.

V. RESULTS

- The risk assessment for Kochi, Kerala, indicates a high flood risk with a 0.67/1.00 risk score and 75% confidence, primarily due to water proximity, soil permeability, and rainfall.
- Residents should monitor weather updates, have an evacuation plan, and consider flood-proofing measures.
- In contrast, the landslide risk is low with a 0.21/1.00 risk score and 85% confidence, as factors like slope angle and vegetation have minimal impact.
- While landslides are unlikely, staying aware of warning signs and past events is recommended. Overall, flood preparedness is crucial, while land- slide concerns are minimal.

The results are shown in Figure 5,6 and 7 .

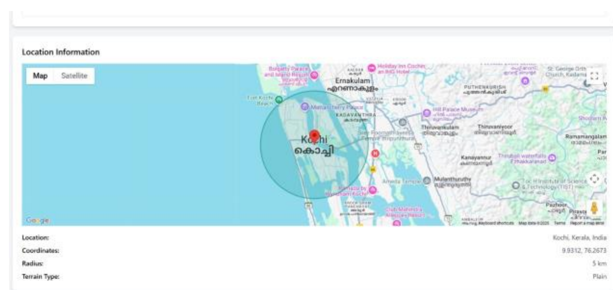

Fig. 5. Map

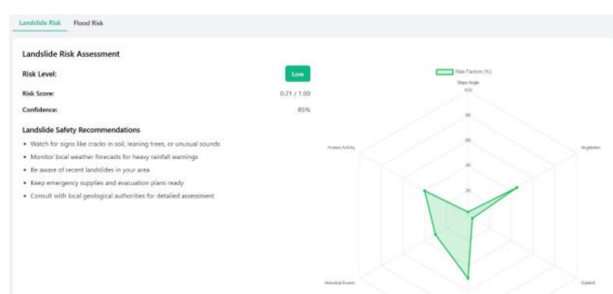

Fig. 6. Lanslide Chart

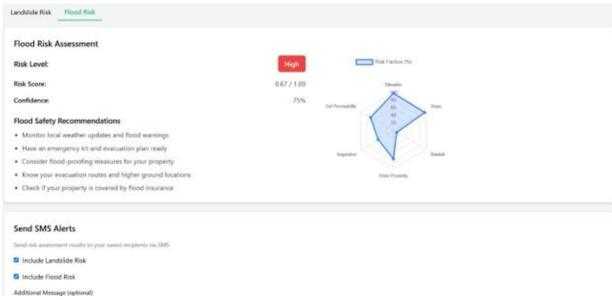

Fig. 7. Flood Chart with Alert System

VI. Conclusion

TThis survey has examined a range of machine learning and deep learning techniques for flood and landslide prediction, highlighting their applications, strengths, and limitations in analyzing environmental data and image tiles. Techniques such as Convolutional Neural Networks (CNNs) and Random Forests (RF) have been foundational in disaster risk prediction, with CNNs excelling in localized feature extraction and RF offering robustness for structured data. However, their lim- itations in capturing global spatial dependencies and high- dimensional image patterns underscore the need for more specialized approaches like Uet, particularly when enhanced with an Attention mechanism, emerges as a powerful model for flood and landslide prediction. Its encoder-decoder archi- tecture and skip connections allow it to retain critical spatial information, while the attention mechanism directs focus to high-risk regions, enhancing both accuracy and efficiency. This 7 makes Attention-UNet exceptionally suitable for applications requiring detailed spatial segmentation and high-resolution risk mapping, as it can prioritize and analyze specific areas prone to floods or landslides with greater precision than traditional models.Overall, this survey indicates that while simpler models offer interpretability and lower computational demands, attention-enhanced UNet holds significant compar- ative advantages for flood and landslide prediction. Its ability to process large-scale geospatial data with spatial prioritization makes it a promising choice for real-time risk assessment and disaster management, providing a more accurate, scalable, and actionable solution for predicting and mitigating the impact of natural disasters

References

[1] F. S. Tehrani, M. Calvello, Z. Liu, L. Zhang, and S. Lacasse, "Machine learning and landslide studies: recent advances and applications," *Natural Hazards*, vol. 114, no. 2, pp. 1197–1245, 2022.

[2] S. R. Meena, L. P. Soares, C. H. Grohmann, C. Van Westen, K. Bhuyan, R. P. Singh, M. Floris, and F. Catani, "Landslide detection in the himalayas using machine learning algorithms and u-net," *Landslides*, vol. 19, no. 5, pp. 1209–1229, 2022.

[3] S. Heo, W. Sohn, S.-J. Park, and D. K. Lee, "Multi-hazard assessment for flood and landslide risk in kalimantan and sumatra: Implications for nusantara, indonesia's new capital," *Heliyon*, vol. 10, p. e37789, 09 2024.

[4] A. Mosavi, P. Ozturk, and K.-w. Chau, "Flood prediction using machine learning models: Literature review," *Water*, vol. 10, no. 11, p. 1536, 2018.

[5] H. M. S. Kungulwar, S. Mishra, U. S. A. Khattak, and N. Bhat, "Landslide and flood prediction using machine learning spark framework," 2019.

[6] M. Khan, A. U. Khan, B. Ullah, and S. Khan, "Developing a machine learning-based flood risk prediction model for the indus basin in pakistan," *Water Practice & Technology*, vol. 19, no. 6, pp. 2213–2225, 2024.

[7] H. D. Nguyen, D.-K. Dang, Q.-T. Bui, and A.-I. Petrisor, "Multi-hazard assessment using machine learning and remote sensing in the north central region of vietnam," *Transactions in GIS*, vol. 27, no. 5, pp. 1614–1640, 2023.

[8] N. Ahmed, A. Firoze, and R. M. Rahman, "Machine learning for predicting landslide risk of rohingya refugee camp infrastructure," *Journal of Information and Telecommunication*, vol. 4, no. 2, pp. 175–198, 2020.

[9] M. Kappi and B. Mallikarjuna, "Artificial intelligence and machine learning for disaster prediction: a scientometric analysis of highly cited papers," *Natural Hazards*, pp. 1–21, 2024.

[10] T. Q. Ngo, N. D. Dam, N. Al-Ansari, M. Amiri, T. Van Phong, I. Prakash, H. Van Le, H. B. T. Nguyen, and B. T. Pham, "Research article landslide susceptibility mapping using single machine learning models: A case study from pithoragarh district, india," 2021.