A Survey on Flood and Landslide Prediction using Machine Learning

Navaneeth K Biju, Rohith R, Sabarinath A M,
Varun V, 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 models are usually designed to operate independently, focusing on specific tasks. However, when feature distributions 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

For ease of study the literature are categorized based on the two strategies of transfer learning: 1. Pre-trained Models as Feature Extractors, 2. Fine Tuning Pre-trained Models.

Tehrani et al. [10]Detecting landslide-affected areas quickly and accurately is critical in emergency response. Traditionally, landslide detection relied on manual analysis of aerial images and field surveys, which are labor-intensive, time-consuming, and costly. The integration of ML in landslide detection automates this process by analyzing satellite, UAV, and airborne images, saving time and increasing precision. Two main types of ML approaches used here are pixel-based and object-based methods. Pixel-based methods evaluate each pixel in an image individually, classifying it as landslide or non-landslide based on attributes like brightness or texture. Conventional ML algorithms such as Random Forest (RF) and Support Vector Machines (SVM) are often applied in this approach. object-based methods consider groups of related pixels, segmenting images into distinct regions to detect larger spatial patterns that represent landslides.

Among machine learning techniques, supervised learning methods—such as Recurrent Neural Networks (RNNs), which

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 03 | March - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

excel at handling sequential data—are especially effective for analyzing time-series data related to landslides. For instance, Long Short-Term Memory (LSTM) networks, a type of RNN, can model long-term dependencies, making them useful for predicting landslides triggered by prolonged rainfall or gradual shifts in ground saturation. By integrating real-time sensor data, temporal ML models can dynamically assess landslide risks, supporting immediate decision-making in disaster management.

Meena et al. [6] U-Net Architecture for Landslide Detection The U-Net model, originally designed for biomedical image segmentation, features an encoder-decoder architecture. The encoder path progressively extracts image features through multiple convolutional and max-pooling layers, downsampling the spatial dimensions. Conversely, the decoder path up-samples these features, restoring the spatial resolution through up-convolution layers and concatenating outputs

The U-Net architecture's skip connections and fully convolutional design make it highly suitable for per-pixel classification in landslide detection. In this case, the U-Net model identified landslide areas by assigning each pixel a probability of being landslide-affected, yielding binary classification maps for further evaluation.

Fig. 1 showcases results from a U-Net model applied to landslide detection. Each row represents outputs from different U-Net configurations, labeled with varying depths (16, 32, 64, and 128). The left column displays the original input data, while the right column shows the model's predictions regarding landslides. Black areas represent non-landslide regions, while white areas indicate detected landslides. As the model complexity increases from 16 to 128, one can observe variations in detection capabilities. With deeper models, the predictions may become more refined, potentially enhancing detection accuracy. This comparative approach highlights how different configurations affect the model's performance in identifying landslides

Heo et al. [2] A study titled "Multi-Hazard Assessment for Flood and Landslide Risk in Kalimantan and Sumatra: Implications for Nusantara, Indonesia's New Capital" offers a comprehensive model for predicting flood and landslide risks in Indonesia's Kalimantan and Sumatra regions. By employing advanced machine learning methods to create accurate hazard maps, this study provides crucial insights into disaster risk management for Indonesia's new capital and offers a framework applicable to other regions vulnerable to natural disasters.

Comparative analysis of the machine learning models demonstrated RF's superior ability to delineate risk areas accurately.The hazard maps generated from the RF model classify regions into four risk levels—low, moderate, high, and very high—for floods, landslides, and multi-hazard scenarios. This classification helps prioritize interventions and optimizes resource allocation, allowing urban planners and disaster management officials to focus on the most vulnerable areas. For instance, areas categorized as "very high" risk might necessitate infrastructure reinforcement, enhanced drainage systems, and strict land use regulations.

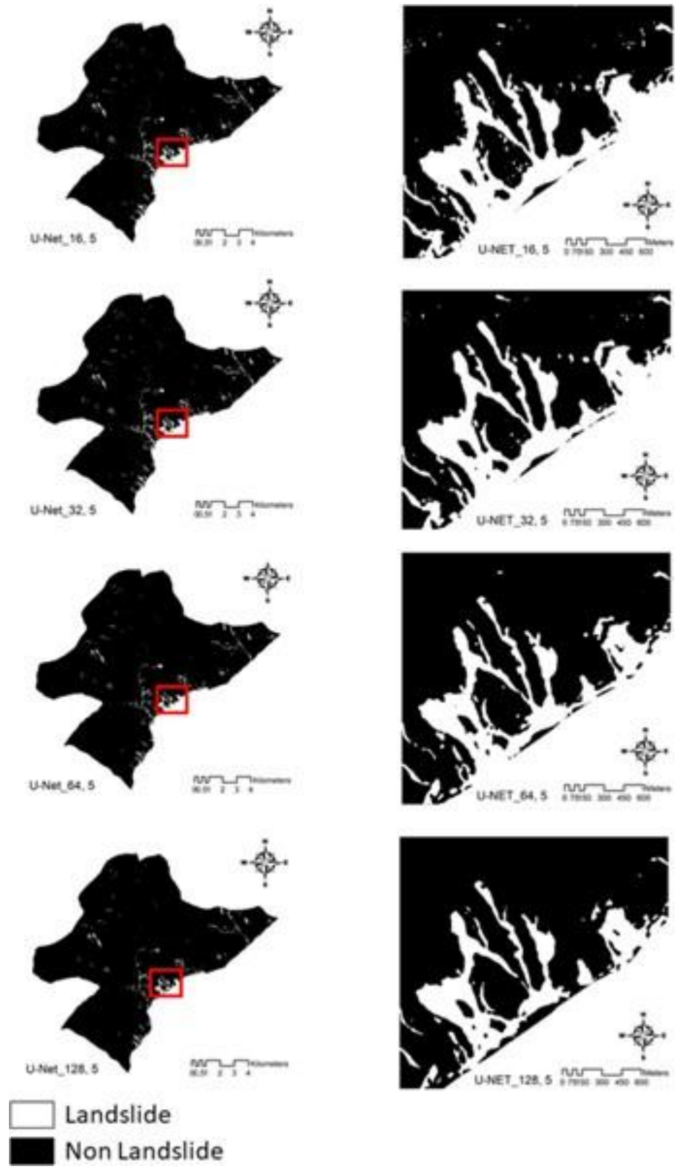

Fig. 1. Landslide detection results using U-Net model in sampled area in the test zone using dataset. [6]

Fig. 2 outlines a structured approach to hazard risk assessment and management using machine learning. It begins with data collection and preprocessing, focusing on environmental factors and inventory maps related to hazards and non-hazards. A variety of machine learning algorithms are employed, such as k-Nearest Neighbors, Naive Bayes, and Random Forest, with the data split into 70% for training and 30% for testing.Following model evaluation, the validation results highlight the optimal model, emphasizing the importance of various environmental factors and providing descriptive statistics. The analysis also includes generating a hazard risk map, which can represent both single and multi-hazard scenarios for different regions. Finally, the outcomes have risk management implications, assisting in urban planning and sustainable development efforts while providing foundational data for policymakers regarding the role of machine learning in hazard management.

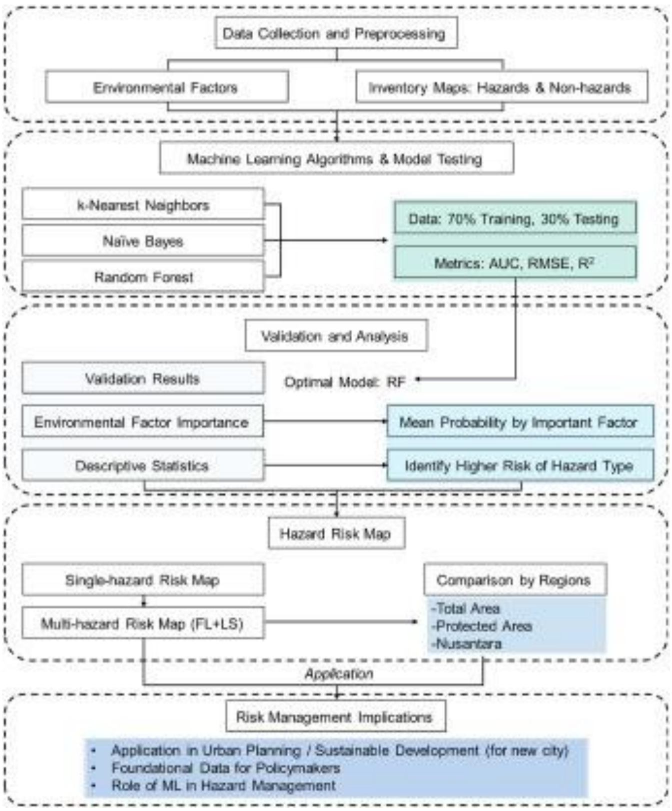

Fig. 2. Methodology framework. [2]

The hazard maps generated from the RF model classify regions into four risk levels—low, moderate, high, and very high—for floods, landslides, and multi-hazard scenarios. This classification helps prioritize interventions and optimizes resource allocation, allowing urban planners and disaster management officials to focus on the most vulnerable areas. For instance, areas categorized as "very high" risk might necessitate infrastructure reinforcement, enhanced drainage systems, and strict land use regulations.

Mosavi et al. [7]ML models into single and hybrid methods, detailing their application across different flood scenarios. Key ML models discussed include Artificial Neural Networks (ANNs), Support Vector Machines (SVM), Decision Trees (DT), and Adaptive Neuro-Fuzzy Inference Systems (ANFIS). Each model has distinct features that make it more or less suitable for specific tasks, depending on factors such as lead time, dataset size, and computational efficiency.

Flood prediction models typically rely on historical flood records and real-time data from rain gauges, radar systems, or satellites. Key predictive variables include rainfall intensity, river inflow, soil moisture, and water level. In high-resolution settings, remote sensing data is often used to improve the spatial accuracy of flood prediction. Data preprocessing steps, such as normalization and decomposition, are essential to enhance model training and reduce noise.

Although ML models offer considerable advantages, several challenges remain. The generalization of ML models is often limited by the availability of high-quality, diverse datasets. For regions with sparse data, model performance may be inconsistent. Hybrid models, while accurate, can be computationally intensive and require expert knowledge for optimal implementation. Future research directions include the development of transfer learning techniques to adapt models to new locations with limited data, improved model interpretability to facilitate policy decisions, and integration with physical models to enhance prediction accuracy in complex environments.

Fig. 3 outlines the essential steps in a machine learning project workflow. The process starts with data acquisition, involving the collection of pertinent information from multiple sources. This data is then subjected to data preprocessing, which involves cleaning and transforming it to ensure quality and consistency. After preprocessing, the building model phase commences, defining the algorithms and structures to be used. Next, the model undergoes training, during which it learns patterns and adjusts its parameters based on the training dataset. Once training is complete, the model is evaluated in the testing phase with a separate dataset to assess its accuracy and effectiveness. This structured process may require iterative refinements, ensuring a robust and reliable machine learning model.

Harsh et al. [5]Landslide and Flood Prediction Using Machine Learning Spark Framework by Mihir Sandeep Kungulwar, Harsh Sanjeev Mishra, Shahzer Ahmed Khattak, and Uzair Nazir Bhat, published in The International Journal of Innovative Research in Science, Engineering, and Technology (May 2019), explores the application of machine learning for predicting floods and landslides. Using 30 years of historical rainfall data, the authors aimed to classify rainfall patterns and assess return periods through statistical analysis. Big data analytics played a crucial role, leveraging Apache Spark to handle large datasets and extract actionable insights.

The study focused on integrating meteorological data, including rainfall, temperature, and humidity, which was collected and processed for machine learning. They applied the random forest algorithm, chosen for its high accuracy and quick classification capabilities, making it especially suitable for environmental forecasting tasks. The research reviewed previous studies on big data and machine learning for predictive modeling in disaster-related fields and identified gaps in systematic approaches for optimizing big data analytics, which this work seeks to address. Results showed the model's robustness, achieving an accuracy of 93.4%.

Khan et al. [4] The research centers on the Indus River Basin, recognized as one of the world's largest transboundary river basins. Covering Pakistan, India, China, and Afghanistan, this basin is crucial for regional water resources. To predict floods accurately, the study collects data on key hydroclimatic variables including,Digital Elevation Model (DEM)-Used for

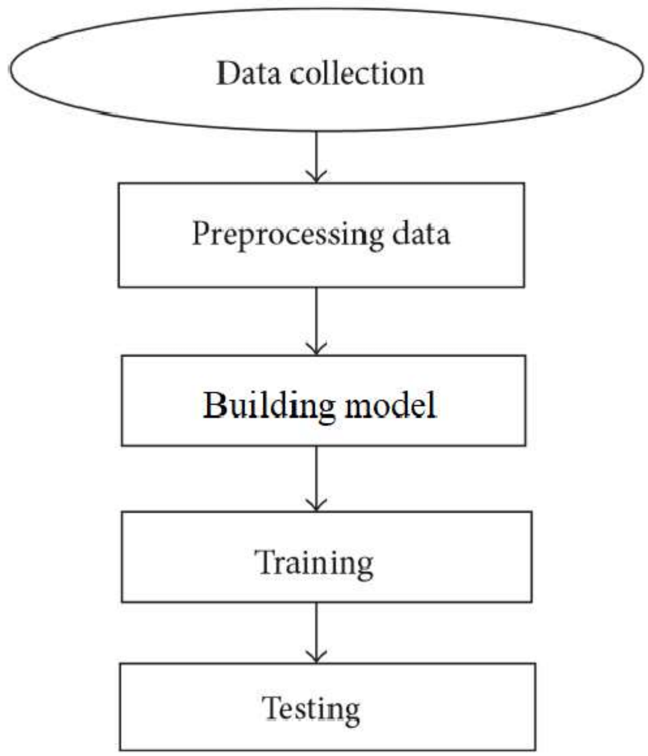

Fig. 3. Basic flow for building an ML model. [7]

watershed delineation,Meteorological Data-Temperature and precipitation records,Hydrological Data-Monthly streamflow measurements. These datasets, spanning from 1985 to 2013, provide a robust historical record for model training and validation.

The study utilizes a variety of machine learning (ML) models—artificial neural networks (ANN), k-nearest neighbors (KNN), support vector machines (SVM), Naive Bayes (NB) and random forest (RF)—each selected for its distinct advantages in capturing flood risk patterns. To enhance model performance, data preprocessing included class balancing, which addresses the unequal distribution of flood (yes/no) events, reducing model bias and improving predictive accuracy. Each ML method was chosen for specific strengths: KNN, which classifies data points based on their nearest neighbors, offers simplicity and effectiveness in handling small datasets. SVM is recognized for its robust accuracy and generalization, making it particularly well-suited for complex flood prediction tasks. The Naive Bayes model, based on Bayes' theorem, assumes predictor independence, allowing for efficient and straightforward flood risk classification. ANN, known for modeling complex flood dynamics, leverages neural networks' capacity to handle non-linear relationships within data. Finally, Random Forest (RF), a common tool in hydrology, utilizes an ensemble of decision trees to boost prediction accuracy and minimize overfitting. The models' effectiveness was evaluated using metrics like accuracy, precision, recall, F1 score, and the Matthews Correlation Coefficient (MCC), ensuring a thorough assessment of their predictive performance.

Fig. 4 shows a bar chart presents the distribution of predictions made by different machine learning models: KNN, SVM, NB, ANN, and RF. Each model is compared across four categories: predicted "YES," predicted "NO," actual "YES," and actual "NO." The pink section represents the percentage of predicted "YES" outcomes, while the light blue indicates the predicted "NO" results. The blue bars represent the proportion of true "YES" cases, while the dark green bars indicate the actual "NO" instances. An analysis of the distributions across models reveals variability in prediction accuracy and the models' tendencies to classify outcomes. For example, one might notice trends where certain models consistently predict more "YES" outcomes compared to others. This visual representation aids in assessing how well each model aligns with actual outcomes, providing valuable insights for improving model selection and accuracy.

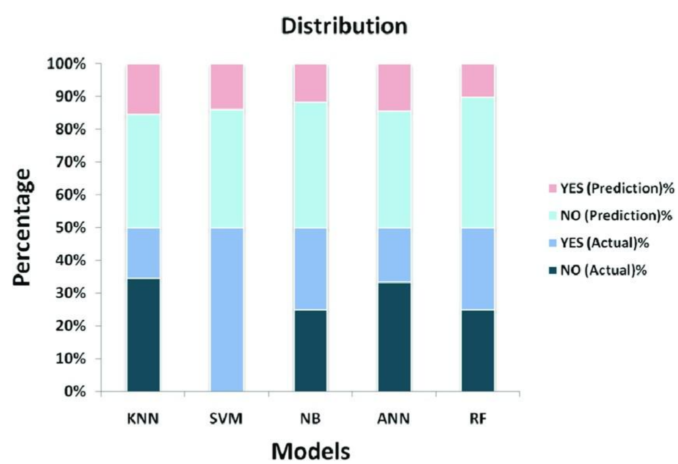

Fig. 4. Distribution diagrams for ML models. [2]

Nguyen et al. [9] Flooding, with its profound environmental and socioeconomic impacts, necessitates accurate prediction methods. The journal article "Flood Prediction using Hydrologic and ML-based Modeling: A Systematic Review" classifies these methods into hydrologic models, which simulate physical water-flow processes, and machine learning (ML) models, which are data-driven. Hydrologic models, including 1D, 2D, and 3D simulations, vary in complexity: 1D models efficiently simulate linear water channels but struggle in complex scenarios; 2D models, with better spatial resolution, capture flow dynamics over grid layouts but are more computationally demanding; and 3D models, which offer the most detailed simulations, are optimal for assessing urban flooding though computationally intensive and suitable for localized issues. ML approaches, like Artificial Neural Networks (ANNs) and Support Vector Machines (SVMs), identify flood patterns from extensive datasets without relying on physical assumptions, making them adaptable across various contexts.

Hybrid models, which combine hydrologic and ML techniques, show strong potential for enhancing real-time forecasting and addressing complex urban flood scenarios. Ensemble ML approaches further improve generalizability and reduce prediction uncertainty. A bibliometric analysis revealed a significant research emphasis on ML-based flood modeling compared to hydrologic models, highlighting ML's role in advancing flood prediction capabilities. The study advocates for integrating these techniques, especially in hybrid models, to leverage ML's data-driven accuracy and hydrologic models' physical realism. Future studies should prioritize enhancing data resources to achieve more precise, localized flood forecasts, thereby highlighting the transformative impact of machine learning in real-time flood prediction and climate resilience.

Huu et al. [9]In a study on predicting landslide susceptibility, four machine learning algorithms were implemented and evaluated: Logistic Regression (LR), Multi-Layer Perceptron (MLP), Gradient Boosted Trees (GBT), and Random Forest (RF). Logistic Regression, a foundational statistical model, predicts the likelihood of landslides by applying a logistic function to weighted inputs. The MLP, a neural network model with a single hidden layer, captures complex patterns and feature interactions. Gradient Boosted Trees, an ensemble model, sequentially trains trees to minimize prior errors, which is effective for data with nonlinear relationships. Lastly, Random Forest, another ensemble model, uses multiple decision trees in parallel, aggregating their outputs for final classification; in the study, it demonstrated the highest accuracy and Area Under the Curve (AUC) scores.

For model evaluation, accuracy and the AUC of the ROC curve were used to measure each model's capability to differentiate between at-risk and safe locations. Random Forest achieved the highest accuracy percentage of 79.19 and an AUC of 0.76, showing reliable performance and suitability for binary classifications like landslide risk prediction. Gradient Boosted Trees performed well, slightly below RF, confirming its effectiveness in processing complex patterns. The MLP outperformed LR but underperformed compared to ensemble models like RF and GBT. Furthermore, data preprocessing techniques like scaling and outlier filtering were found to have minimal impact on model performance, suggesting that the models were adaptable to input variations. introduced a convolutional neural network-based technique for systematically classifying plant disease symptoms. MobileNet_V3_Large was chosen for more transfer learning experiments. The following three stages of the stepwise TL method are used: 1. Weights from the source domain are transferred and frozen, with the exception of the classifier, which is replaced. 2. After each training iteration, the loss is calculated. Another layer block is unfrozen if the loss computed has not decreased in the last 10 epochs. 3. If the layers are remained frozen owing to a constant loss, they will all unfreeze. On the Pepper and PlantVillage datasets, the suggested approach achieved 99.69% and 99.69% accuracy, respectively.

Nguyen et al. [8] The study evaluated three single machine learning models—Linear Discriminant Analysis (LDA), Logistic Regression (LR), and Radial Basis Function Network (RBFN)—for landslide susceptibility mapping in Pithoragarh District, India, aiming to test their simplicity and adaptability in predicting landslide-prone areas. Using a landslide inventory of 398 historical events, the data was divided into a percentage of 70 for training and a percentage of 30 for validation. The study considered ten key landslide-influencing factors, including slope, aspect, curvature, elevation, land cover, lithology, geomorphology, proximity to rivers and roads, and overburden depth, while accounting for local conditions, such as fault-controlled rivers. Each model was trained and validated using Weka and GIS software, with statistical metrics like accuracy, specificity, sensitivity, Positive Predictive Value (PPV), Negative Predictive Value (NPV), Kappa, RMSE, and AUC assessing their performance. Results indicated that all models performed well, achieving AUC values over 0.90, with LR standing out as the best predictor (AUC = 0.926) for landslide susceptibility in the region. While RBFN excelled in PPV and sensitivity, LR achieved the highest validation results overall, producing landslide susceptibility maps categorized into five risk levels. Areas within 200 meters of roads or rivers were particularly vulnerable to landslides, consistent with other studies linking infrastructure to increased landslide risk. The findings suggest that, despite the availability of complex ensemble and hybrid models, single models like LR can offer reliable, cost-effective predictions and are useful for planning and early warning in landslide-prone areas. Future studies are encouraged to reevaluate single ML models for different regions, as landslide-influencing factors may vary with local conditions.

Kappi et al. [3]One of the key findings of this study is the high citation impact of certain research contributions, indicating influential work that has significantly shaped the field. Papers focused on machine learning methodologies, such as convolutional neural networks (CNNs), decision trees, and ensemble methods, were highly cited, underscoring the effectiveness of these methods in predictive tasks. In particular, CNNs have been instrumental in analyzing remote sensing images for disaster detection, while decision trees and ensemble techniques have been used to map landslide susceptibility in various geographic contexts. The study also found that the average citation growth rate over the analyzed period was slightly negative, suggesting potential shifts in citation patterns, possibly due to the expanding availability of AI tools and new developments in the field. While computer science and engineering dominate the research areas within AI-driven disaster prediction, interdisciplinary approaches involving environmental sciences, geology, and telecommunications have also emerged, further broadening the scope of this field.

remote sensing and geographic information systems (GIS) are instrumental in mapping flood and landslide susceptibility, while IoT devices provide real-time environmental data essential for early warning systems. These technological advancements are pivotal in improving the precision and scope of disaster prediction models, especially for geographically diverse or remote areas prone to natural hazards.

Despite significant progress the study acknowledges certain limitations in existing research while analyzing highly cited papers offers valuable insights it may overlook emerging studies or innovative methodologies published in lesser-known journals to address this the authors recommend broadening future research by incorporating a wider range of databases and alternative analytical techniques such as thematic analysis this approach could help uncover underexplored areas and recent trends potentially leading to novel interdisciplinary strategies for disaster prediction additionally the study highlights the need for ai models designed for specific disaster types and calls for real-world assessments to ensure these models effectively contribute to disaster management efforts.

Ahmed et al. [1]This study examines refugee camps in bangladeshs coxs bazar district an area highly vulnerable to landslides to build a predictive model for landslide risks geospatial data such as elevation lithology the normalized difference vegetation index ndvi and the topographic wetness index twi were utilized elevation data came from a global 30-meter digital elevation model while ndvi was obtained through satellite imagery twi calculated from flow accumulation data helps assess water movement across the landscape a critical factor in landslide susceptibility additionally the study incorporated infrastructure data from over 17000 camp facilities including healthcare centers schools tube wells latrines and general infrastructure each facility was analyzed alongside landslide risk data to determine its vulnerability status

The study applied four machine learning modelslogistic regression lr multi-layer perceptron mlp gradient boosted trees gbt and random forest rffor data processing and training various preprocessing methods including min-max scaling standardization and normalization were used to improve predictive performance among these models random forest proved to be the most effective achieving an accuracy of 7619 and an area under the curve auc of 076 its high reliability makes it particularly useful for identifying landslide-prone zones in camps allowing for timely interventions and reinforcements to reduce potential risks

Fig. 5 displays a ROC curve compares the performance of four machine learning models: Random Forest, Gradient Boosted Trees, Multi-Layer Perceptron, and Logistic Regression. The x-axis shows the False Positive Rate (FPR), while the y-axis shows the True Positive Rate (TPR). Each line represents how well a model differentiates between positive and negative classes. The Random Forest model (black curve) achieves the highest AUC (0.76), indicating the best performance among the models. The Gradient Boosted Trees (red curve) follow with an AUC of 0.71, showing moderate performance. The Multi-Layer Perceptron (orange curve) performs slightly worse with an AUC of 0.69. Finally, the Logistic Regression (gray curve) achieves the lowest AUC at 0.67, demonstrating the weakest performance. The dashed diagonal line represents random guessing, where the true positive rate equals the false positive rate. Models closer to the top-left corner of the graph perform better. Random Forest's curve being farthest from the diagonal highlights its strong predictive capability.

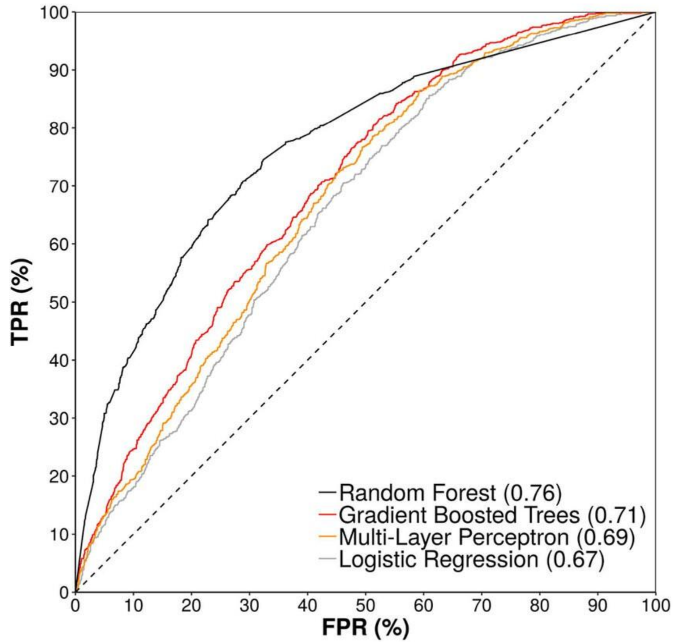

Fig. 5. ROC curve of machine learning algorithms. [1]

Fig. 6 displays two ROC curve plots, labeled (a) and (b), representing the performance of four models—MLP, AB-RBFN, MB-RBFN, and DG-RBFN—during the training and validation phases, respectively. The x-axis shows ”1-Specificity” (false positive rate), while the y-axis shows ”Sensitivity” (true positive rate), which are standard measures for evaluating classification models. In subplot (a), corresponding to the training phase, the AUC (Area Under the Curve) values indicate high performance: DG-RBFN achieves 0.969, MLP achieves 0.963, MB-RBFN achieves 0.953, and AB-RBFN achieves 0.936. Subplot (b), corresponding to the validation phase, shows slightly lower AUC values for all models: DG-RBFN (0.931), MB-RBFN (0.929), AB-RBFN (0.926), and MLP (0.913). These values are presented at a 95% confidence interval, ensuring statistical reliability of the classification results. The DG-RBFN model consistently demonstrates superior performance in both phases, as its curves are closest to the top-left corner, reflecting better sensitivity and specificity. The training phase (a) shows tighter and more optimal curves compared to the validation phase (b), where performance slightly declines. The differences between the training and validation AUC values highlight a potential performance drop when models are generalized to unseen data. Overall, the ROC curves and AUC values effectively illustrate the classification capabilities and robustness of the models, with DG-RBFN emerging as the best performer across both phases.

III. COMPARATIVE STUDY

Multiple machine learning methods have shown great potential in forecasting floods and landslides each bringing its

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 03 | March - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

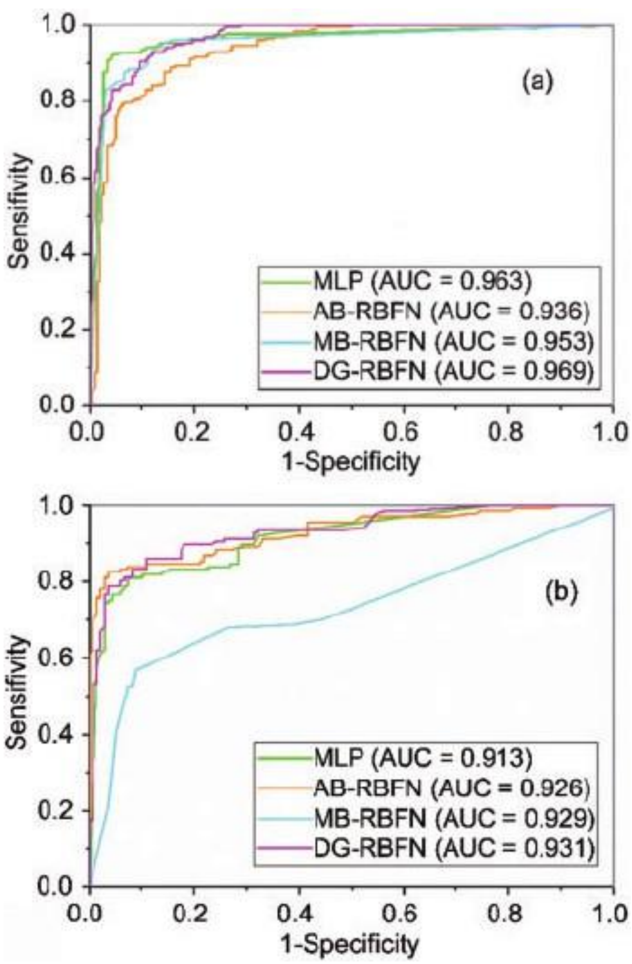

Fig. 6. AUC values of the models during the (a) training phase and (b) validation phase. [8]

own set of strengths and weaknesses random forest rf is especially favored for its reliability and clarity in interpretation it is highly effective in identifying risk zones and creating detailed hazard maps by efficiently processing structured data and categorizing areas into various risk levelsfrom low to very high however rf can encounter difficulties with large-scale spatial datasets since it may not fully capture widespread spatial relationships

Convolutional neural networks cnns are highly effective at processing satellite imagery pinpointing fine details that are crucial for disaster mapping however their focus on localized pixel-level features can mean they miss broader spatial trends important for a full understanding of disaster risks to address this innovative architectures like attention-unet have been developed to enhance the models ability to recognize wider spatial relationships and achieve more accurate segmentation

Attention-UNet, as highlighted in several studies, offers a significant advantage in disaster prediction by incorporating attention mechanisms into the encoder-decoder architecture. This enables the model to focus on high-risk areas, such as unstable slopes or flood-prone regions, while maintaining a global perspective. After thoroughly evaluating, analyzing, and comparing various models discussed in the literature, we determined that the Attention-UNet mechanism is the most suitable choice for our project. Its ability to capture both fine-grained details and broader spatial contexts ensures precise high-resolution risk mapping. Moreover, Attention-UNet's focus on prioritizing critical regions during training leads to more actionable and reliable predictions, making it the ideal candidate for addressing the complex challenges of flood and landslide prediction. Furthermore, its scalability enables efficient application across diverse geographic regions, ensuring broader usability in disaster management efforts. The integration data with advanced machine learning models like Attention-UNet ensures dynamic, up-to-date risk assessments, enhancing disaster response efforts and minimizing damage.

Hybrid and ensemble techniques improve disaster forecasting by merging multiple algorithms to generate more dependable predictions. For instance, combining Radial Basis Function Network (RBFN)-based methods with hydrological models can better tackle intricate challenges such as urban flooding. While these methods demand greater computational resources, they effectively address the shortcomings of standalone models, enhancing scalability and strengthening the resilience of the overall system.

Each machine learning model analyzed provides unique advantages for predicting floods and landslides. Basic models, such as Random Forest (RF), are easy to interpret and computationally efficient, whereas advanced frameworks like Attention-UNet excel in detailed mapping and spatial analysis. Combined methods further boost accuracy, making them crucial for effective disaster management. After thorough evaluation, Attention-UNet stands out as the top choice due to its attention mechanisms and ability to merge local and global data, ensuring precise predictions. The selection of a model ultimately depends on factors like data availability, computational capacity, and the scale of the area at risk.

IV. CONCLUSION

This 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 limitations 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 architecture 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 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 comparative 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] N. Ahmed, A. Firoze, and R. M. Rahman. Machine learning for predicting landslide risk of rohingya refugee camp infrastructure. *Journal of Information and Telecommunication*, 4(2):175–198, 2020.

[2] 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*, 10:e37789, 09 2024.

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

[4] 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*, 19(6):2213–2225, 2024.

[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] 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*, 19(5):1209–1229, 2022.

[7] A. Mosavi, P. Ozturk, and K.-w. Chau. Flood prediction using machine learning models: Literature review. *Water*, 10(11):1536, 2018.

[8] D.-D. Nguyen, N. Tiep, Q.-A. Bui, H. Leˆ, I. Prakash, R.-D. Costache, M. Pandey, and B. Pham. Landslide susceptibility mapping using rbfn-based ensemble machine learning models. *Computer Modeling in Engineering Sciences*, pages 1–10, 11 2024.

[9] 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*, 27(5):1614–1640, 2023.

[10] F. S. Tehrani, M. Calvello, Z. Liu, L. Zhang, and S. Lacasse. Machine learning and landslide studies: recent advances and applications. *Natural Hazards*, 114(2):1197–1245, 2022.