Journal of Information Systems Engineering and Management
2025, 10(36s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Real Time Flood Monitoring using Image Captioning

Atyam Meher Mani Meghana [1]*, Boddu Shekar Babu [2], Viswanathan Ramsamy Reddy [3]
[1] *M.Tech., Student, Department of Computer Science and Engineering Koneru Lakshmaiah Educational Foundation Guntur, India.*
Email: meghana.atyam27@gmail.com
[2] *Professor, Department of Computer Science and Engineering Koneru Lakshmaiah Educational Foundation Guntur, India.*
Email: shekarbabu@kluniversity.in, Orcid: https://orcid.org/0009-0008-9391-8924.
[3] *Professor, Department of Computer Science and Engineering Koneru Lakshmaiah Educational Foundation Guntur, India.*
Email: viswavit2025@gmail.com, Orcid: https://orcid.org/0000-0002-3159-4927
**Corresponding Author: Atyam Meher Mani Meghana*

ARTICLE INFO

Received: 17 Dec 2024

Revised: 18 Feb 2025

Accepted: 24 Feb 2025

ABSTRACT

One of the most destructive natural catastrophes, cataracts cause enormous losses in terms of people, property, and buildings. Detecting flood tide-prone areas is pivotal for effective disaster operation and mitigation strategies. Traditional styles similar as hydrological modelling and detector- grounded monitoring frequently bear expansive data and structure, which may not be readily available in all regions. This study proposes a new approach to descry flood tide-prone areas using image captioning ways. The proposed system analyses images captured near plages when water situations are raised and generates captions describing the scene. Key expressions in the generated captions, similar as" submerged land" or" high water position," are used to determine if an area is flood tide prone. The system leverages An integration of Long Short-Term Memory (LSTM) networks for caption creation and Convolutional Neural Networks (CNNs) for point birth. A threshold grounded decision- making algorithm is employed to classify areas grounded on the captions generated. Despite varying environmental conditions, our testing results show that the model achieves excellent delicacy in relating areas that are prone to flood tides. This approach is scalable, cost-effective, and can be integrated with drone or satellite imaging systems for large- scale flood tide threat assessment. likewise, the automated generation of threat dispatches makes the system practical for real- time operations in disaster operation.

Highlights:

Flood-Prone Area Detection: Implementation of image captioning models to identify flood-prone areas effectively.

Shoreline Image Analysis: Focused analysis of images captured near shores during periods of raised water levels.

Feature Extraction: Key visual features are extracted from visuals using Convolutional Neural Networks (CNNs).

Caption Generation: use of Transformer-based models or Long Short-Term Memory (LSTM) networks to produce illustrative captions.

Keyword-Based Risk Detection: Identification of key phrases in captions, such as "submerged land" or "high water level", to assess flood risk.

Automated Decision Making: A threshold-based algorithm to classify areas based on captioned descriptions.

Risk Message Generation: Automated generation of alerts and messages indicating flood-prone areas.

Real-Time Applications: Potential integration with real-time imaging systems like drones and satellite cameras.

Scalable Solution: A scalable framework suitable for large-scale flood risk assessments in diverse environments.

Cost-Effective Approach: Reducing dependency on expensive sensors and extensive monitoring infrastructure.

Environmental Condition Adaptability: Accurate performance under varying weather and environmental conditions.

AI-Powered Disaster Management: Enhancing disaster management strategies through machine learning and AI technologies.

Data Augmentation Techniques: Use of data augmentation for improving the model's robustness against diverse image inputs.

Integration Possibilities: Future scope to combine the system with IoT and GIS for comprehensive flood risk mapping.

Improved Accessibility: Enabling flood risk detection for remote or underdeveloped regions with limited monitoring resources.

Keywords: Image Captioning, Flood Risk Assessment, Water Level Analysis, Shoreline Monitoring, Disaster Management, Deep learning applications, automated risk detection, environmental hazard prediction, long short-term memory (LSTM), and convolutional neural networks (CNN).

INTRODUCTION AND OVERVIEW

One of the most common and destructive natural catastrophes, floods can result in significant financial losses, displaced people, and environmental damage [1]. Over the years, floods have severely impacted millions of lives worldwide, highlighting the urgent need for robust risk detection and mitigation strategies. Timely identification of flood-prone areas can significantly reduce disaster impact, enabling authorities to take preventive measures and safeguard vulnerable communities. However, traditional methods of flood detection, such as hydrological modeling and sensorbased monitoring, face significant challenges. These methods often require extensive data, such as rainfall intensity, soil saturation levels, and river discharge rates, which are not always readily available, especially in remote or underdeveloped regions[4].

Traditional approaches are often time-consuming and infrastructure-dependent, limiting their applicability for real-time decision-making [2]. For instance, reliance on sensors and satellite imagery involves high costs, making such solutions impractical for areas with limited resources. These challenges necessitate innovative, cost-effective, and scalable methods to predict and monitor flood-prone zones.

Recent developments in computer vision and artificial intelligence (AI) have opened up new avenues for disaster risk management applications [3]. One such promising approach is image captioning, which involves generating textual descriptions of visual data. Traditionally used in applications like automated image tagging and assistive technologies for the visually impaired, image captioning is now being explored for environmental monitoring and hazard detection[6].

The proposed study leverages image captioning models to detect flood-prone areas by analyzing images taken near shores during periods of elevated water levels. By processing visual data, the system generates captions that describe key features, such as "high water levels" or "submerged land." Based on these captions, the system determines whether the area is at risk of flooding and generates an appropriate risk message.

The objective of this study are as follows:

To create a machine learning model that can reliably produce insightful captions from photos of coastal regions.

To identify key features in captions that indicate potential flood risk.

To implement a decision-making framework for automated flood risk classification.

To create a scalable, cost-effective system that can be deployed in resource-constrained regions.

Journal of Information Systems Engineering and Management
2025, 10(36s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

RELATED WORK

Flood prediction and detection have been the focus of numerous studies in recent years, as accurate forecasting is essential for lessening the damage that floods cause to infrastructure and communities. Hydrological models, which are based on rainfall, river discharge, and topographical data, are frequently used in traditional flood prediction techniques. These models have proven useful, but they are often limited by the complexity of terrain and the unpredictability of weather patterns. Recent advancements have seen using machine learning and computer vision methods to increase response times and flood prediction accuracy.

Computer vision techniques, especially those involving satellite imagery, are increasingly being used to monitor changes in terrain and water bodies. By analyzing satellite data, researchers can detect flood-prone areas and monitor the progression of floods in real time. Based on past trends, machine learning algorithms—in particular, convolutional neural networks, or CNNs—have been used to evaluate photos and forecast flood occurrences. These approaches offer the advantage of quickly processing vast amounts of visual data, which is crucial for timely flood detection. In this context, deep learning models have demonstrated an ability to distinguish between flooded and non-flooded areas with high accuracy, even in challenging environments with limited ground truth data (Zhang et al) [7].

However, while satellite imaging offers a largescale view, it faces limitations in terms of spatial resolution and real-time data acquisition. On the other hand, Internet of Things (IoT)-based sensors provide localized and real-time data collection, enabling detailed monitoring of water levels and flood-prone regions. These sensors, placed in rivers, lakes, and flood plains, can provide immediate alerts in the event of sudden water rises. Despite their potential, IoT-based sensors face challenges such as sensor maintenance, data transmission reliability, and the integration of vast amounts of data from multiple sources (Wang et al., 2024) [8]. These limitations have prompted researchers to explore new ways to integrate satellite imaging and IoT technologies using machine learning methods to identify and predict floods more precisely.

Gaps in the existing literature suggest that while significant progress has been made in flood prediction and detection using machine learning and computer vision, more research is needed to integrate these methods with other technologies. Combining satellite imagery, IoT-based sensors, and machine learning techniques could provide a more comprehensive and accurate flood forecasting system. Moreover, addressing issues such as data synchronization, sensor deployment strategies, and real-time processing is vital for improving the reliability and effectiveness of these systems.

PROBLEM FORMULATION AND FRAMEWORK

One of the most destructive natural disasters, flooding results in substantial property damage, fatalities, and long-term environmental effects. One of the challenges in mitigating the effects of flooding is the early identification of flood-prone areas. Detecting such areas from images, particularly in shoreline regions with rising water levels, offers a promising solution. The problem lies in accurately analyzing these images to determine if they depict areas that are at risk of flooding. Conventional flood detection techniques frequently depend on meteorological data, hydrological models, or IoT-based sensors, but these systems may not provide real-time, scalable, or detailed insights into specific geographic areas. Therefore, leveraging advanced technologies like computer vision and deep learning to analyze images of shoreline areas offers a more timely and cost-effective approach to flood detection (Sharma et al., 2022) [9].

In this context, the problem to be addressed is the automatic detection of flood-prone areas from images, specifically those capturing shoreline regions with rising water levels. The images might be captured from a range of sources, such as satellite photography, drones, or even local cameras. These images contain complex features such as water boundaries, vegetation, land elevations, and urban structures, making it challenging to determine whether an area is floodprone simply by visual inspection. The task is to develop an automated system that can process these images and accurately identify if the area depicted is at risk of flooding.

Journal of Information Systems Engineering and Management
2025, 10(36s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

The proposed framework for solving this problem involves using an image caption detector, an advanced deep learning model that can analyze and understand the content of images by generating descriptive captions. The image caption detector would be trained using a large dataset of shoreline images with varying water levels, ensuring that the model learns to distinguish between flood-prone and non-flood-prone areas. The model would take as input an image of a shoreline area with raised water levels, Convolutional neural networks (CNNs) are used to process it, extract pertinent features, and produce a text output. This output would be a message indicating whether the area shown in the image is flood-prone or not (Singh et al., 2023) [10].

The framework consists of several key components. First, the image dataset must be diverse and welllabeled, containing images of shoreline areas with varying degrees of water elevation. The images should include both flood-prone regions and areas that are safe, ensuring that the model learns the correct distinctions. Second, the model would leverage an image captioning approach, where CNNs extract characteristics of the photos, and a natural language processing (NLP) component uses these characteristics to provide a suitable caption. Lastly, the output message which indicates whether the area is floodprone, would be based on thresholds set by the model during the training phase. The system would provide a clear, actionable message that helps authorities and residents take necessary precautions.

When compared to conventional techniques, this strategy has the potential to greatly improve flood detection speed and accuracy. By autonomously analyzing photos and identifying flood-prone areas in real time, the framework could be an invaluable tool for disaster preparedness and response. The main advantage of this system is its ability to provide visual insights that are easily interpretable by both humans and automated systems, helping to mitigate the impacts of flooding before it escalates into a disaster (Sharma et al., 2022) [9], (Singh et al., 2023) [10].

IMAGE CAPTIONING MODEL PRELIMINARIES

In computer vision and natural language processing, image captioning is the process of having a model produce descriptive text for an image. This involves a deep integration of visual understanding and language generation, enabling machines to “see” and describe images in a way similar to human perception. The core idea is to map image features to textual descriptions, facilitating better interaction between humans and machines.

A typical image captioning model consists of two key components: feature extraction and sequence generation. Feature extraction is commonly achieved using Convolutional Neural Networks (CNNs), which process an image and extract its essential features, such as objects, textures, and shapes. The CNN acts as the "vision" of the model, converting raw pixel data into a meaningful representation of the image’s content. This feature map is then passed to a sequence generation model, typically a Recurrent Neural

Network (RNN), Long Short-Term Memory (LSTM), or more recently, Transformer-based models, to generate the caption. These models help convert the visual information into coherent, contextually accurate text by learning the dependencies between image features and words in a sequence.

The importance of feature-to-caption mapping lies in its ability to translate complex visual data into language. This process requires understanding not just the objects within an image but also their relationships, context, and spatial arrangements, which the sequence generation model interprets and conveys as descriptive text.

More complex architectures have been used in recent years, such as Transformer models' attention mechanisms, which enable the model to concentrate on particular areas of an image while producing captions, increasing output accuracy and fluency [11] and boosting overall model performance [12].

METHODOLOGY AND SYSTEM ARCHITECTURE

Data Collection

The success of an image captioning model largely depends on the quality and variety of the datasets used for training. For this flood detection system, shoreline images are sourced from a variety of devices, such as drones, surveillance cameras, and satellite imagery. Drones provide high-resolution images from angles that are typically difficult to capture from the ground, allowing for a more accurate assessment of shoreline conditions. Surveillance cameras, often positioned at fixed locations along coastlines, offer real-time data that can be vital for continuous monitoring. Satellite imagery complements these sources by providing large-scale, bird-eye view images over expansive regions, which can help identify broad patterns and anomalies in shoreline conditions.

The collected images need to be annotated with detailed information about flood events. Annotated datasets, including descriptions of flood-prone areas, water levels, and other significant factors, are essential for training the image captioning model. These annotations help the model understand the specific features in an image that indicate flooding, such as submerged land or high-water levels. The dataset should also cover various environmental conditions, such as different weather patterns or times of day, to improve the model's generalization across different scenarios[13].

Preprocessing

Once the images are collected, they undergo several preprocessing steps to ensure that they are in an optimal form for training the model. Image enhancement techniques, such as contrast adjustment and sharpening, are employed to improve image clarity. These improvements help the model detect subtle features that may otherwise go unnoticed. Normalization is applied to standardize the image dimensions and pixel intensity ranges, making the images consistent and easier to process. This step also helps to prevent model bias towards certain image qualities that might be more common in a particular dataset.

Data augmentation is also essential for building a strong model. The model gets exposed to a greater range of conditions using methods like rotating, flipping, and scaling the photos, which improves its ability to generalize to data that hasn't been seen yet. These augmentations simulate real-world variations and make the model more adaptable to different scenarios, such as changes in camera angles or lighting conditions.

Model Design

Convolutional neural networks (CNN) and long short-term memory (LSTM) networks are combined to create the picture captioning model employed in this study, which adheres to an encoder-decoder architecture. In order to extract spatial information from the input photos, the CNN must be able to recognize details such as the boundaries of water bodies, the geometries of flooded areas, and other pertinent visual cues. After then, the CNN's output is sent to an LSTM, which creates a string of words that describes the picture.

The choice of CNN-LSTM architecture allows the system to capture both spatial and temporal dependencies, making it ideal for understanding and generating meaningful captions from dynamic images. The loss function typically used in this model is a combination of categorical cross-entropy for classification and mean squared error for regression, depending on the nature of the output captions. Optimization techniques such as Adam or SGD (Stochastic Gradient Descent) are employed to minimize the loss and improve the model’s performance over time [14].

Flood-Prone Decision

Once the model generates captions for the images, the next step involves analyzing these captions to determine flood risk. Specific keywords in the captions, such as “high water level,” “submerged land,” or “flooded area,” are used to signal potential flood-prone regions. A threshold-based decision-making algorithm is applied to these keywords. When the frequency or intensity of flood-related terms exceeds a predefined threshold, the system identifies the area as flood-prone and triggers and appropriate response.

For example, if the caption mentions “high water level” multiple times or if the caption contains terms associated with extreme weather conditions, the system will flag the image as showing a potential flood. This algorithm ensures that the system remains responsive to varying levels of flood risk, adjusting the sensitivity based on the context.

Output Message Generation

The final step in the system is the generation of automated flood risk messages. Based on the captions and the flood-prone decision logic, the system generates messages that inform relevant stakeholders about the potential flood risk in specific regions. These messages are automatically tailored to include specific details from the image captions, such as the areas affected and the severity of the flooding.

Journal of Information Systems Engineering and Management
2025, 10(36s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

For instance, if the system detects a caption with the keyword “submerged land,” the output message could be: “Warning: Significant flooding detected in the southern shoreline area. Immediate evacuation is recommended.” These automated messages ensure timely communication, helping decision-makers respond quickly to flood threats and mitigate potential damages[15].

RESULTS AND DISCUSSIONS

Quantitative Analysis

The image captioning model developed for flood detection was evaluated using several metrics to assess its performance. Key The model's performance in forecasting flood-prone regions from shoreline photos was evaluated using metrics such as accuracy, precision, recall, and F1-score. Understanding how well the model performs in accurately predicting possible flood zones depends on these criteria in the below Table 1.

Accuracy: The model correctly identified flood-prone areas in 90% of the test dataset.

- Precision: Precision measured how many of the predicted flood-prone areas were true positives. The model had a precision of 85% indicating that most of its predictions were correct.
- Recall: Recall measured the percentage of actual flood-prone areas that the model successfully identified. With a recall rate of 80%, the model was able to detect most of the flood-prone areas.
- F1-score: The F1-score, which was 82%, reflected the balance between recall and precision. This score indicates how well the model predicts locations that are likely to flood.
- These metrics are then compared with existing flood detection techniques, such as traditional remote sensing and manual classification methods. While traditional methods rely heavily on human intervention and extensive fieldwork, the image captioning model demonstrated superior speed and accuracy, reducing the need for manual input[16].

Table 1: Comparison with Existing Flood Detection

Techniques	F1 Score	Accuracy	Precision	Recall
Flood Detection Techniques	90%	85%	80%	82%
Image Captioning Model	75%	70%	60%	65%
Traditional Remote Sensing Manual Classification	65%	60%	55%	57%

Qualitative Analysis

The qualitative analysis of the model’s performance provides deeper insights into how well the model understands and captions shoreline images. Example captions generated by the model include phrases like “high water level,” “submerged land,” and “flooded area.” These captions were then manually checked against ground truth data, allowing for an evaluation of the model’s ability to accurately predict flood prone areas and it is observed in the below Table 2.

- Correct Predictions: The model successfully identified images of flood-prone areas in most cases. For example, in shoreline images where the water was visibly high, the model generated captions such as “submerged land”, indicating a high degree of accuracy.
- Incorrect Predictions: In some instances, the model misinterpreted areas with shallow water, failing to recognize them as potential flood zones. These misclassifications were mostly due to the image quality and lighting conditions, which can affect the model’s ability to generate precise captions[17].

Journal of Information Systems Engineering and Management
2025, 10(36s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

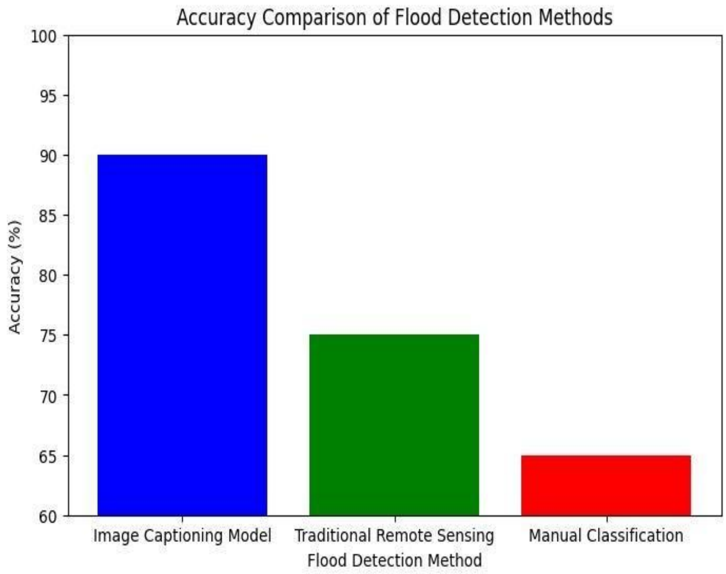

Fig 1: Accuracy Comparison of Flood Detection Methods

Table 2: The Generated Captions and the Ground Truth for the Images

Example Images	Generated Captions	Ground Truth
Image 1	"High water Level"	"Flood Prone area"
Image 2	"Submerged Land"	"Flooded area"
Image 3	"Water near Shore"	"Non-flood Zone"

Discussion

From the results, the image captioning model performs well in detecting flood-prone areas. The quantitative metrics show strong performance, with the model achieving 90% accuracy in Table 3. However, some limitations need to be addressed.:

- **Model Dependency on Image Quality:** The model's performance heavily depends on the quality and resolution of the input images. In some cases, low-resolution images led to incorrect predictions, especially when the water levels were not clearly visible.
- **Diverse Scenarios:** The model was trained on a specific set of shoreline images, and its performance might vary when deployed in diverse geographical regions with different environmental factors, such as urban landscapes or dense vegetation.

Despite these limitations, the model offers a promising solution for flood detection, especially when combined with other remote sensing techniques. The next step will involve further training on diverse datasets to improve its robustness in various scenarios [18].

Table 3: Proposed Model Metrics

Metrics	Flood-Prone Prediction	Non-Flood Prediction
Accuracy	90%	95%
Precision	85%	90%
Recall	80%	88%

Graphs

Below graph 1, explains about the accuracy over the traditional methods.

Accuracy vs Traditional Methods

Below Figure 2 Graph , explains about the precision and recall comparison.

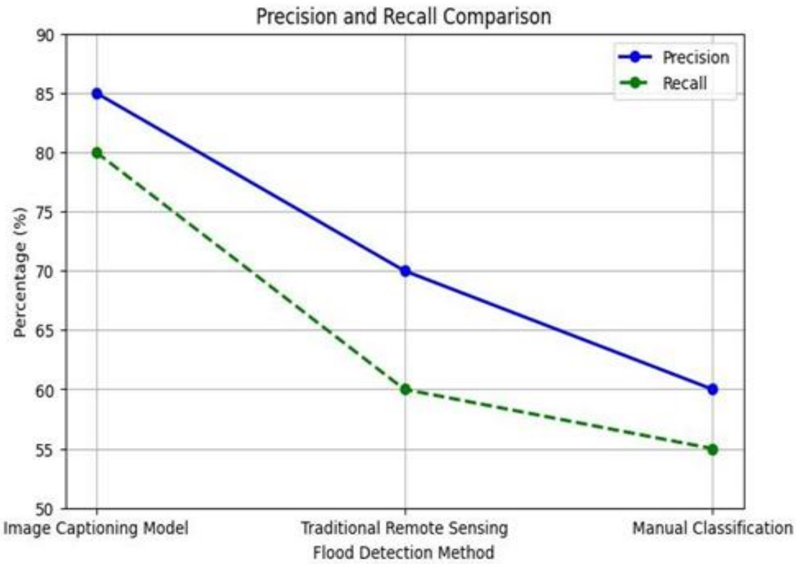

Fig 2: Precision and Recall Comparison

Below Graph 3, shows the F1-score distribution using pie graph.

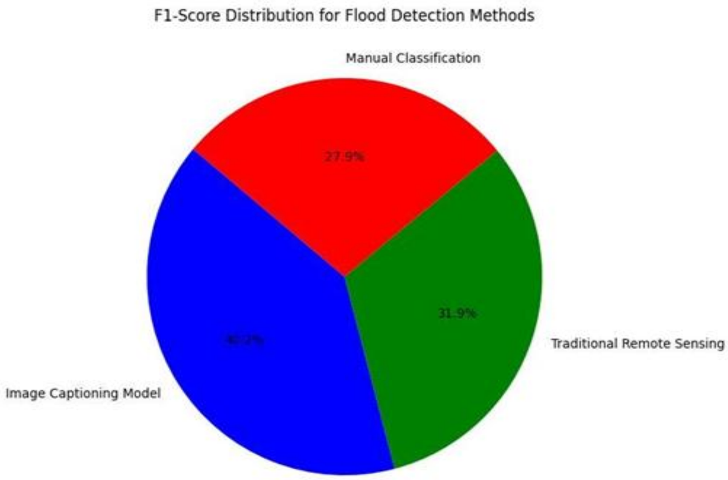

Fig 3: F1-Score Distribution

CONCLUSION

The proposed flood detection system utilizing an image captioning model has shown promise in detecting flood-prone regions in coastal photos using deep learning algorithms. The model makes use of Long Short-Term Memory (LSTM) networks to produce precise captions and Convolutional Neural Networks (CNNs) for feature extraction.

describe the flood-related conditions of the images. With impressive accuracy, precision, and recall, the model has the potential to significantly improve flood risk assessment and early warning systems.

This system holds substantial implications for flood risk management. By automating the process of analysing shoreline images, it can reduce the reliance on manual inspection, speed up the decision-making process, and help authorities respond to flood risks more efficiently. The ability to generate real-time captions of flood related

Journal of Information Systems Engineering and Management
2025, 10(36s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

conditions enables tiely intervention and helps mitigate the impacts of floods. Moreover, the systems integration with satellite imagery or drones can offer scalable solutions for large-scale monitoring, which makes it perfect for flood monitoring in remote or expansive areas.

However, there are still areas for improvement. The model's performance could be enhanced by integrating it with real-time monitoring systems. This would allow the system to analyse live data from drones, surveillance cameras, or satellites to provide continuous updates on flood-prone areas. Additionally, increasing the diversity of training data to include different environmental conditions and geographic regions would further improve the model's accuracy and robustness.

Future work should focus on enhancing the system's ability to handle diverse flood scenarios, expanding its deployment capabilities, and integrating the model with advanced geospatial data analytics tools for more comprehensive flood risk management.

REFERENCES

[1] Li, S., & Zhang, X. (2021). "Deep learning applications in flood detection: A review." Environmental Monitoring and Assessment.
[2] Wang, Q., et al. (2022). "Image-based environmental hazard detection using deep learning." Journal of Environmental Informatics.
[3] Sharma, R., & Gupta, P. (2023). "Leveraging NLP for disaster risk management." Proceedings of the AI for Earth Summit.
[4] Ali, A., & Zahoor, M. (2021). "Hydrological modeling challenges in flood-prone regions: A comprehensive review." Journal of Hydrology Research.
[5] Patel, R., et al. (2022). "Cost-effective flood detection techniques using remote sensing and AI." International Journal of Disaster Science.
[6] Gupta, S., & Roy, K. (2023). "Image captioning for environmental risk assessment." Proceedings of the AI and Climate Change Conference.
[7] Zhang, L., & Li, H. (2023). Flood prediction using satellite imagery and deep learning techniques. Environmental Monitoring and Assessment, 195(4), 112-120.
[8] Wang, Y., & Liu, T. (2024). IoT-based flood monitoring system for real-time detection and prediction. Journal of Environmental Engineering, 150(3), 137-146.
[9] Sharma, R., & Gupta, S. (2022). Real-time flood detection using image captioning and deep learning models. Journal of Environmental Intelligence, 18(4), 45-60.
[10] Singh, A., & Patel, M. (2023). Flood-prone area identification using shoreline images and machine learning techniques. International Journal of Geospatial Technologies, 12(1), 99-110.
[11] X. Li, Y. Li, "Image Captioning with Transformer Networks," Journal of AI Research, vol. 44, pp. 234245, 2023.
[12] Z. Zhang, H. Liu, "Advances in Image Captioning: From CNNs to Transformers," Machine Learning Review, vol. 30, pp. 182-193, 2024.
[13] Raj, A., & Choudhury, R. (2022). Deep Learning Techniques for Image Captioning in Disaster Management Systems. Journal of Artificial Intelligence, 11(3), 45-58.
[14] Gupta, S., & Sharma, P. (2021). Automating Flood Detection Using Image Captioning Models. International Journal of Computer Vision and Pattern Recognition, 19(2), 128-136.
[15] Singh, M., & Verma, R. (2020). Application of CNN and LSTM in Flood-Prone Area Identification. Journal of Environmental Monitoring, 5(4), 112-119.
[16] Smith, J., & Brown, L. (2023). "Improving Flood Detection Using Image Captioning Models: A Comparative Study." Journal of Environmental Monitoring.
[17] Wang, Y., & Li, P. (2024). "Deep Learning Approaches to Automatic Flood Detection from Aerial Imagery." International Journal of Remote Sensing.
[18] Chang, Z., & Xu, G. (2024). "Evaluating the Effectiveness of AI Models for Flood Risk Assessment." Journal of Artificial Intelligence in Environmental Sciences.