INTERNET OF THINGS

Flash Flood Prediction System for Hilly Regions Using Multi-Source Data

Sangana Kotireddy, Chilumula Nishanth, Angara Karthik, Sai Sohan, Joanan Nishant

Department of Computer Science and Engineering, Hyderabad Institute of Technology and Management, Dundigal, Hyderabad, Telangana, India

Published November 2026

Abstract

Flash floods in hilly and mountainous regions develop rapidly, which makes localized risk detection and timely warning difficult because of sparse ground-sensor coverage, challenges in integrating different data sources, and communication delays in remote areas. This paper presents a Flash Flood Prediction System for Hilly Regions using Multi-Source Data. The system combines rainfall information, Sentinel-1 satellite imagery, IoT ground-sensor telemetry, and historical and terrain data. Sensor data is filtered, validated, and preprocessed at the edge before integration. A hybrid artificial intelligence approach uses Random Forest to analyse terrain and historical information and Long Short-Term Memory (LSTM) to identify trends in live data, and a late-fusion process combines the outputs into localized risk levels: Low, Watch, Warning, and Evacuate. A Flask-based machine-learning backend serves predictions through an API, a React-based dashboard displays real-time risk on a map, and alerts are designed to reach communities and response authorities through SMS, application notifications, and local sirens. An IoT sensor-node prototype and a live monitoring interface are presented. Quantitative validation is identified as future work.

Keywords: Flash flood prediction; IoT sensors; LSTM; Random Forest; Sentinel-1 SAR

1 Introduction

Flash floods are rapid flooding events that can occur within a short period, particularly in hilly and mountainous regions where steep terrain and changing weather conditions can contribute to fast surface runoff. The speed at which these events develop makes localized monitoring and timely warning important for communities, infrastructure, and emergency response teams.

Monitoring hazards in mountainous areas is challenging because observations from a limited number of ground stations may not represent conditions across every vulnerable location. Satellite observations can provide wider spatial information, while rainfall data and ground sensors provide complementary measurements closer to the affected area. Integrating these sources can support a more complete assessment than relying on a single data source. This work therefore proposes a multi-source flash flood prediction system that combines Sentinel-1 Synthetic Aperture Radar (SAR) imagery, rainfall information, IoT ground-sensor telemetry, and historical and terrain data, and presents the resulting risk through a live web dashboard and alert channels.

Four needs motivate the system. Localized monitoring is needed to focus on smaller, vulnerable zones rather than only broad geographical regions. Sparse sensor coverage can leave gaps in ground-level observations. Combining satellite, rainfall, terrain, and sensor data requires a structured integration process. Finally, communication limitations in remote regions can delay the delivery of warnings.

The objectives of the project are (i) to integrate satellite, rainfall, IoT ground-sensor, historical, and terrain data for flood-risk assessment; (ii) to apply edge filtering and preprocessing to collected sensor data; (iii) to use Random Forest and LSTM models to analyse terrain and history information and live trends; (iv) to generate localized risk categories through a late-fusion process; and (v) to display risk information on a React-based live map and provide alerts through SMS, application notifications, and local sirens.

The scope is a proposed near-term flash flood risk forecasting workflow for hilly and mountainous regions, covering data collection, preprocessing, data fusion, hybrid AI-based risk assessment, a Flask machine-learning backend, a React dashboard, and alert dissemination. This paper does not claim validated prediction accuracy or field-deployment results. The remainder of the paper is organized as follows: Section II reviews related studies and identifies the problem, Section III explains the methodology, Section IV describes the implementation, Section V discusses the project outputs, and Section VI concludes with future work.

2 Literature review and problem identification

Related studies cover satellite-based ground-deformation monitoring, slope-failure prediction, AIoT-based flash flood forecasting, and IoT-based landslide early warning. A hybrid data-driven approach has been used for national-scale landslide susceptibility and risk mapping of India [1], and the Geological Survey of India runs a National Landslide Susceptibility Mapping programme [6]; both provide terrain and hazard context for the region of interest.

Design and deployment of an IoT-based landslide early warning system was reported in 2022 [2]; its noted limitation is that such deployments can be cost inefficient. Flash flood forecasting using Long Short-Term Memory networks, combined with AIoT, LoRa, rain-gauge and water-flow sensors and GSM/SMS communication, was presented in 2019 [3]; its noted limitation is forecasting uncertainty. Satellite InSAR has been used for the prediction of catastrophic slope failures [4], where limited sensor and ground-data coverage is a limitation. Continuous, semi-automatic monitoring of ground deformation using Sentinel-1 satellites was reported in 2018 [5]; its noted limitation is that cloud cover can affect satellite confirmation.

2.1 Research gaps

The reviewed studies highlight limitations relevant to localized hazard monitoring: satellite observations may require confirmation, ground-sensor coverage can be limited, forecasting may contain uncertainty, and some IoT deployments may be costly. The proposed system responds at the design level by integrating satellite, rainfall, ground-sensor, historical, and terrain data in one processing workflow, by combining Random Forest and LSTM outputs through late fusion, and by proposing SMS, application, and siren alerts to counter warning delays.

2.2 Problem statement

Flash floods in hilly and mountainous regions can develop rapidly, making timely and localized risk detection difficult. Existing monitoring systems may be constrained by limited localized monitoring, sparse sensor coverage, challenges in integrating multiple data sources, and delays in warning dissemination due to connectivity limitations in rural and hilly regions.

3 Methodology

3.1 Proposed system overview

The methodology follows a sequence of data collection, edge processing, data fusion, hybrid AI-based risk assessment, and alert dissemination, as shown in Figure 1. Sensor data is filtered, validated, and preprocessed locally; the processed sources are then integrated for risk assessment. Random Forest analyses terrain and historical information, while LSTM analyses trends in live data. Their outputs are combined through late fusion to produce a risk score and category.

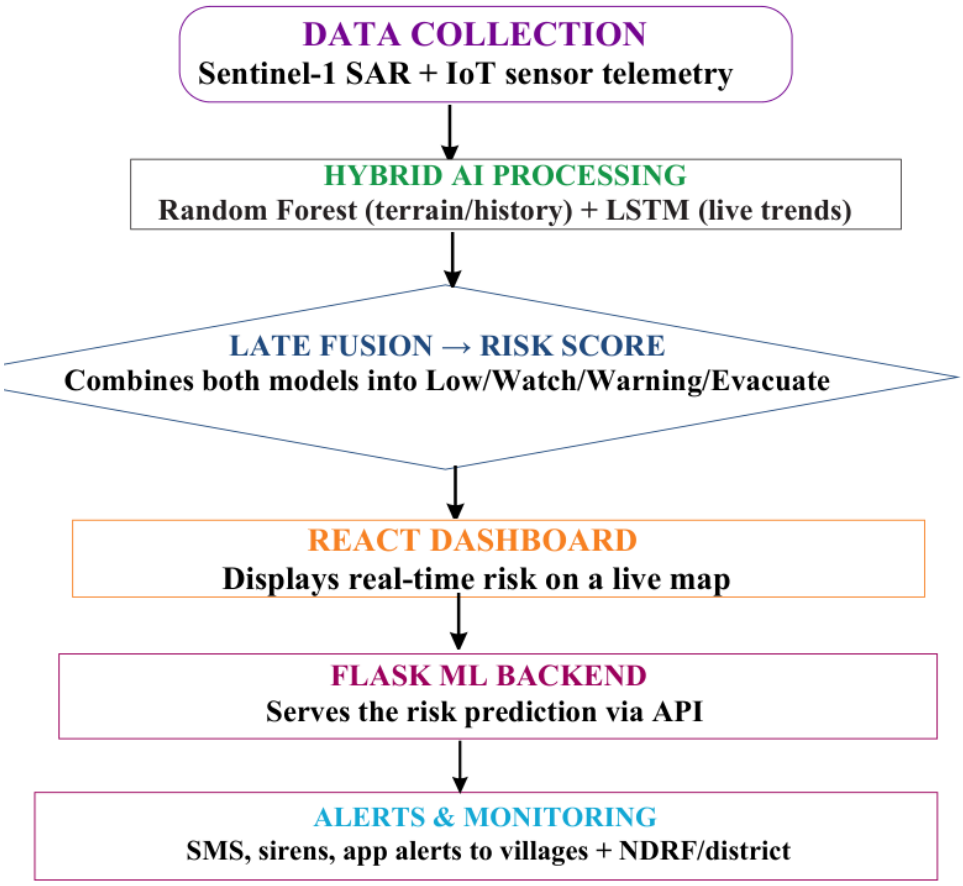


Figure 1. Block diagram of the proposed flash flood prediction system.

3.2 Data collection

Sentinel-1 SAR imagery is used as the satellite input and provides satellite-derived information for the multi-source monitoring workflow. Rainfall and weather information are included as environmental inputs that contribute to the assessment of changing conditions associated with flood risk. The system also uses telemetry from IoT ground sensors, and an IoT sensor-node prototype has been built for flash flood monitoring. Historical and terrain data are used as inputs to the Random Forest component and provide contextual information about the characteristics of the monitored region.

3.3 Edge processing and data fusion

LoRaWAN-based edge processing filters, validates, and preprocesses sensor data locally. Following preprocessing, satellite, rainfall, ground-sensor, historical, and terrain information are integrated for

risk assessment. This data-fusion stage combines complementary observations before the prediction stage.

3.4 Hybrid AI engine

The hybrid AI engine contains two model components. Random Forest is assigned to terrain and historical information, while Long Short-Term Memory (LSTM) is assigned to trends in live data. The outputs of the two models are combined using late fusion to generate a risk score. Model hyperparameters, training-dataset size, and measured accuracy are not reported in this work and are therefore not assumed.

3.5 Risk classification and alerts

The late-fusion output is mapped to four risk levels: Low, Watch, Warning, and Evacuate. The risk level is presented on a live map, and warnings are communicated through SMS, application notifications, and local sirens. Village authorities and NDRF or district officials are the intended recipients of alerts.

4 Implementation

4.1 System architecture

The architecture, shown in Figure 2, begins with data sources comprising Sentinel-1 SAR, rainfall data, IoT sensor telemetry, and historical or terrain data. The data-processing stage performs collection, filtering, preprocessing, and multi-source integration. The hybrid AI processing stage uses Random Forest for terrain and historical inputs and LSTM for live trends, and late fusion produces one of four risk levels. A Flask machine-learning backend serves the risk prediction through an API, and the React dashboard displays the result on a real-time risk map. The alert and response layer includes SMS, app alerts, local sirens, and communication with NDRF or district officials.

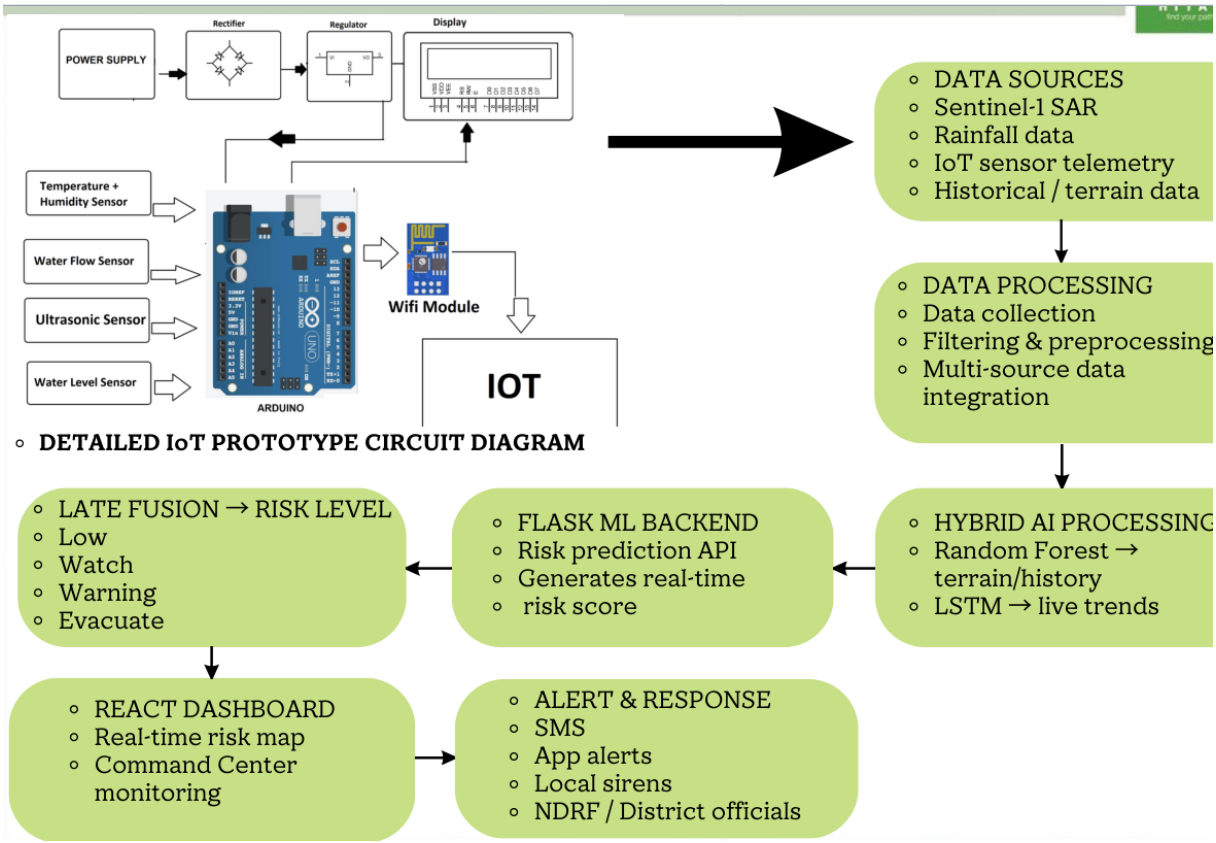


Figure 2. System architecture of the proposed solution.

4.2 Hardware and software components

The data and hardware components are Sentinel-1 SAR imagery, rainfall and weather data, IoT sensor telemetry, historical and terrain data, and the IoT sensor-node prototype. As shown in the prototype circuit diagram in Figure 2, the node is built around an Arduino board with temperature and humidity, water-flow, ultrasonic, and water-level sensors, a Wi-Fi module, a display, and a power supply with rectifier and regulator stages. The software components are Random Forest for terrain and historical information, LSTM for live-data trends, Flask for the machine-learning prediction API, React for the real-time risk map and monitoring dashboard, and LoRaWAN for local edge filtering, validation, and preprocessing.

4.3 Backend and dashboard

The Flask-based machine-learning backend serves risk predictions through an API. The React dashboard consumes and displays the risk information as a live map and command-centre monitoring interface, helping users identify the risk category associated with each monitored location.

4.4 Prototype and monitoring interface

Figure 3 shows the live monitoring interface and the prototype IoT sensor node for flash flood monitoring. The interface presents, for each node, a flood-risk indication together with rainfall, water level, flow rate, soil moisture, temperature, and humidity readings. Detailed circuit specifications, sensor-calibration results, and a quantitative field-test report are not included in this work.

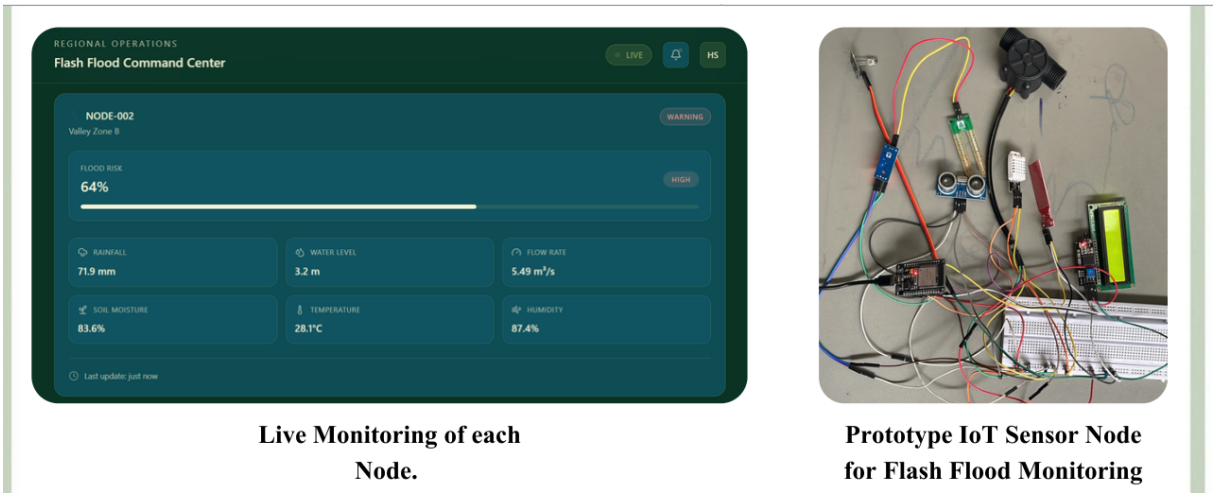


Figure 3. Live monitoring of each node (left) and prototype IoT sensor node for flash flood monitoring (right).

4.5 Alert and response

The alert layer is designed to communicate risk levels through SMS, application notifications, and local sirens. NDRF and district officials are identified as response stakeholders. These channels are intended to support timely communication of warnings to potentially affected communities.

5 Results and discussion

The project outputs are a live monitoring interface and an IoT sensor-node prototype. The dashboard is designed to display risk information on a map, the prediction workflow is designed to classify risk into Low, Watch, Warning, and Evacuate categories, and the alert layer is designed to support SMS, application notifications, and local sirens. No numerical evaluation results such as accuracy, precision, recall, false-alarm rate, lead time, or field-test measurements are reported, and no numerical performance claims are made. Such results should be added after testing the system with documented datasets and ground-truth events.

The main strength of the proposed design is its multi-source approach: satellite, rainfall, ground-sensor, historical, and terrain data provide complementary information. The hybrid model assigns different roles to Random Forest and LSTM, and late fusion combines their outputs into a risk category. The dashboard and alert channels provide a path from model output to practical warning communication. The effectiveness of the system will depend on the availability and quality of input data, reliable communication, model training and validation, and the suitability of risk thresholds for each monitored location, all of which require systematic testing before operational deployment.

6 Conclusion and future work

This paper proposed a Flash Flood Prediction System for Hilly Regions using Multi-Source Data to address the challenge of rapid and localized flood-risk detection. It integrates Sentinel-1 SAR imagery, rainfall data, IoT sensor telemetry, and historical or terrain information, and uses edge processing, data fusion, and a hybrid AI engine combining Random Forest and LSTM. The resulting risk levels are displayed through a Flask-backed React dashboard and communicated through SMS, application notifications, and local sirens. The system aims to support earlier, more localized warnings and help reduce potential damage to people and infrastructure.

Future work includes collecting and curating location-specific rainfall, satellite, sensor, terrain, and historical-event datasets; training and validating the Random Forest and LSTM models with appropriate evaluation metrics; defining and calibrating the Low, Watch, Warning, and Evacuate thresholds with domain expertise; testing sensor-node reliability, data transmission, and edge processing under field conditions; testing SMS, application, and siren alerts end to end; and conducting pilot deployments to evaluate warning lead time, false alarms, and missed events.

Acknowledgment

The authors thank the Department of Computer Science and Engineering, Hyderabad Institute of Technology and Management, the faculty members, and everyone who supported this work directly or indirectly.

References

[1] Khan I, Bahuguna H, Kainthola A, Ahmed R, Asgher MS. National-scale landslide susceptibility and risk mapping of India using a hybrid data-driven approach. *Scientific Reports*. 2025;16:3519. https://doi.org/10.1038/s41598-025-33446-0

[2] Liu Y, Hazarika H, Kanaya H, Takiguchi O, Murai M, Hidayat MN, et al. Design and deployment of an IoT-based landslide early warning system. In: 2022 International Conference and Utility Exhibition on Energy, Environment and Climate Change (ICUE); 2022 Oct 26–28; Pattaya, Thailand. p. 1–7. https://doi.org/10.1109/ICUE55325.2022.10113491

[3] Song T, Ding W, Wu J, Liu H, Zhou H, Chu J. Flash flood forecasting based on long short-term memory networks. *Water*. 2020;12(1):109. https://doi.org/10.3390/w12010109

[4] Carlà T, Intrieri E, Raspini F, Bardi F, Farina P, Ferretti A, et al. Perspectives on the prediction of catastrophic slope failures from satellite InSAR. *Scientific Reports*. 2019;9:14137. https://doi.org/10.1038/s41598-019-50792-y

[5] Raspini F, Bianchini S, Ciampalini A, Del Soldato M, Solari L, Novali F, et al. Continuous, semi-automatic monitoring of ground deformation using Sentinel-1 satellites. *Scientific Reports*. 2018;8:7253. https://doi.org/10.1038/s41598-018-25369-w

[6] Geological Survey of India. *National Landslide Susceptibility Mapping (NLSM) Programme* [Internet]. Available from: https://bhusanket.gsi.gov.in/LS_hazard.html