Hybrid Flood Forecasting System Integrating Machine Learning and Geospatial Analysis: A Study of Vijayawada, India

Bunga Varshith[1], Dr. G. Narasimha Rao[2]

[1]MTech Student, Andhra University College of Engineering(A), Visakhapatnam, Andhra Pradesh

[2]Ad-HocProfessor, Andhra University College of Engineering(A), Visakhapatnam, Andrha Pradesh

Abstract - **Floods are the natures destructive force causing major human, economic and agricultural loss. A flood forecasting system can minimize such losses; a Decision Support System (DSS) for Flood Forecasting in Vijayawada (India) is presented in this paper. The proposed system can predict the flood by using Sentinel-1 SAR-GRD for flood extent mapping, SRTM-DEM for flood depth estimation, and a 5-year-size historical dataset of hydrological (India-WRIS) and meteorological (Open-Meteo) variables. A geospatial analysis was performed of the severe September 2024 flood event, and results including an inundation area of 65.35 km2 and corresponding depth of inundation were derived. This ground truth generated by the satellite was used to improve the historical dataset which was used as an input to train a two-stage machine learning architecture. This architecture first applies the random forest classifier to estimate the probability of occurrence of flood, and if the risk is high, the second stage applies a random forest regressor to estimate the potential inundation area. The final DSS is an interactive dashboard that integrates live 7-day weather forecasts from the Open-Meteo API to provide a dynamic flood risk assessment, thereby demonstrating a reliable and scalable framework for data-driven disaster management.**

Keywords- Flood Forecasting, Decision Support System (DSS), Machine Learning, Sentinel-1 SAR, Geospatial Analysis, Random Forest Classifier, Random Forest Regressor.

I. INTORDUCTION

Floods have been the most frequently occurring and damaging natural disaster events with significant economic, environmental and human impacts, particularly damaging property, infrastructures and agricultural lands worldwide. Flood forecasting is an important part of disaster preparedness and risk reduction. Floods normally happen during heavy rains which is caused by a large amount of precipitation and an increase in river runoff. Vijayawada is a major economic centre of the state of Andhra Pradesh situated on the banks of Krishna River. The topographical location of the city makes it highly susceptible to floods, recent September 2024 flood which caused widespread damage is the evidence for it. Traditional flood warning systems are helpful but typically give limited information for disaster management authorities to evaluate the potential scale of an event and allocate resources effectively. Traditional hydrological model based on statistical methods, whereas Machine Learning techniques have shown promising results in overcoming the limitations of statistical methods. By using data from meteorological, hydrological, and geospatial sources, the machine learning models would help to develop accurate, reliable and timely flood forecasting models. This research was conducted to develop a Decision Support System (DSS) through the integration of geospatial analysis and prediction modeling. The aim was to create a system that takes flood warnings a stage further and gives a precise forecast of the effects of the flooding.

II. PREVIOUS WORK

Significant research has been conducted on flood prediction, with a clear move towards using data-driven and machine learning (ML) methods to boost accuracy and provide early warnings. Many studies have examined the performances of different supervised machine learning algorithms. Lawal et al. [2] established that Logistic Regression was considered the best model to predict floods in Kebbi

State, Nigeria, which was better than Decision Tree and Support Vector Classifier when using 33 years of rainfalls as input data. Likewise, Syeed et al. [7] examined four models in Bangladesh and observed that models trained on a recent 10-year dataset performed better, with Binary Logistic Regression being the top performer. When it comes to commonly used ensemble techniques, Reddy et al. [3] found that Random Forest (RF) was more effective for reducing flood impact, achieving better accuracy than the Decision Tree model. A common theme in these studies is the importance of good data preparation and feature selection for achieving high accuracy. Gauhar et al. [1] showed this by using Pearson and Spearman correlation analysis to find rainfall and cloud coverage as the most important predictors before applying a k-Nearest Neighbors (k-NN) model in Bangladesh, which had a high accuracy of 94.91%. Going beyond weather data, Myrchiang et al. [5] showed the benefit of using geospatial information, such as Digital Elevation Models (DEM), slope, and curvature, for flood prediction in Assam, India. Their work showed that a Random Forest model using this enriched dataset had a high accuracy of 96.26%. Similarly, Vimala et al. [8] used a broad dataset with more than 20 features, including dam condition and urban growth, and found that a Logistic Regression model had the best accuracy of 99.87% among six tested methods, including advanced boosting models such as XGBoost and LGBM. New hybrid approaches have been introduced to make predictions more accurate. Sakthi et al. [4] created a hybrid HMM-ML model that combines a Hidden Markov Model (HMM) with standard ML classifiers. This framework acted as a fine-tuner, greatly improving the accuracy of all tested models and achieving a top accuracy of 97.33% when used with an HMM-Random Forest setup. These studies show a growing trend towards using more detailed datasets and hybrid models to better understand the complex nature of flood events, while also highlighting the key role of rainfall and river data as the main predictors [6]. Our research is based on these findings by combining a wide range of hydrological and meteorological features with satellite-based ground-truth data to measure flood severity, addressing a major challenge in quantitative flood prediction.

III. METHODOLOGY

The proposed system is a feature-based hybrid method for predicting floods that combines geospatial analysis and time-series aware machine learning models. It is intended to be easy to understand and reliable, using data from multiple sources available through open-source platforms. Unlike other models, which are built on a single type of data, this method uses a mix of satellite-based information, historical records of water flow, water discharge, and weather-related data to better capture the complicated and nonlinear nature of flood events. We introduce a new set of input features for the model, engineered temporal features. These features help the model understand the conditions that lead to a specific day. The main source of these features was the daily historical data. To track changes over time, we calculated lag features, such as Rainfall_lag_1, which shows the conditions from the previous day, and rolling window features, such as Rainfall_3day_sum, which shows the total effect of variables over a recent period. We also used satellite data of the actual flooded area in square meters as input, not as a feature in our regression model, but as the value we are trying to predict, which is a very important improvement on just using yes-no flood labels.

IV. MODEL ARCHITECUTRE

We developed a flood forecasting system as a modular pipeline consisting of three main stages: geospatial analysis, training of predictive models, and the DSS dashboard. Each module is designed to ensure reproducibility and utilizes geospatial software such as ESA SNAP and ArcGIS Pro, along with Python libraries like Pandas, Scikit-learn, and Streamlit, to build a complete and end-to-end early warning system.

STUDY AREA: The study focuses on the city of Vijayawada and its surrounding areas in the state of Andhra Pradesh, India. Vijayawada, located at approximately 16.51° N latitude and 80.65° E longitude, is a significant urban, political, and commercial center, making it essential for disaster management and planning.

DATASET: The system is built on a comprehensive dataset sourced from various open-source providers.

The main data sources include Sentinel-1 SAR imagery from the Copernicus Hub, SRTM DEM from USGS, historical hydrological data for the Krishna River from India-WRIS, and historical meteorological data for Vijayawada from the Open-Meteo API. All geospatial data has been standardized to the WGS 1984 UTM Zone 44N coordinate reference system to ensure accurate spatial alignment.

GEOSPATIAL ANALYSIS: The September 2024 flood event was analyzed using a standard remote sensing workflow within the ESA SNAP and ArcGIS Pro software environments

Sentinel-1 Pre-processing: The raw Sentinel-1 data were processed using the following sequence: Apply Orbit File > Radiometric Calibration (to Sigma0) > Speckle Filtering (Lee Sigma) > Range-Doppler Terrain Correction > Linear to dB Conversion.

The preprocessed Sentinel-1 images were transformed into quantitative measurements of flood extent and depth. A water threshold of -14 dB was determined using a histogram analysis of the pre-flood image. This threshold was applied to the during-flood images using the Raster Calculator in ArcGIS Pro to create a binary water map. The Flood Inundation Map was generated by subtracting the permanent water mask, thus isolating only the newly inundated land. The total pixel count of this final map was used to calculate the actual flooded area (sqm). To estimate the flood depth, a vector boundary of the inundated area was created, and elevation values at the edge of the flood were extracted from the DEM. After filtering out "NoData" values, the mean of these elevations was calculated to establish the Water Surface Elevation (20.04 m). The map for flood depth was then generated by subtracting the ground elevation (from the DEM) from the water surface elevation.

FORECASTING MODEL: The model architecture is a two-stage pipeline designed to address the severe class imbalance in the historical data (many non-flood days vs. very few flood days).

Data Aggregation and Feature Engineering: The Raw historical data (2019-2024) were aggregated into a single daily summary dataset, capturing the maximum or minimum values for each day. This dataset was then enhanced with engineered temporal features (1-day lags and 3-day rolling sums for key variables such as rainfall and water level).

Model Architecture: A two-stage prediction pipeline was developed:

Classifier: This will be the random forest classifier which will be trained using historical data. It was meant to predict the binary result of the occurrence of a flood (Flood_Occurred = 1 or 0). The parameter class weight=balanced was applied so that the model would attach the right weight to the rare flood events.

Regressor: A random forest regressor, trained exclusively on historical data from actual flood occurrences, is used to estimate the actual flooded area, if a flood event is first predicted

A two-step architecture is utilized in the system to improve the accuracy. The first one is where the data is first classified and then a second specialized model is used to derive an exact number. This approach is superior to employing a single model since the data is extremely unbalanced

DSS Dashboard: The final DSS was built using Streamlit. It loads the two trained models, fetches a live 7-day forecast from the Open-Meteo API, runs the two-stage prediction, and displays the results in an interactive dashboard.

V. RESULTS

This section provides the results of the spatial analysis of Sentinel-1 and the DSS dashboard, which showcases flood forecasting. The geospatial analysis of the September 2024 flood event yielded a total inundated area of 65.35 km² within the defined study area. The flood inundation map revels the extent of area that is flooded (Figure 1). The flood depth map (Figure 2) shows different levels of depth ranging from 0 to 10 m, and shows that low-lying areas have flood inundation of over 10 m.
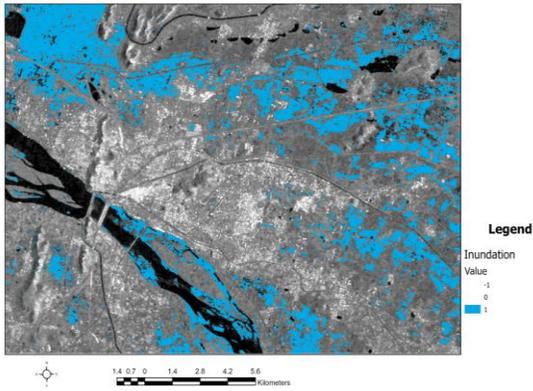
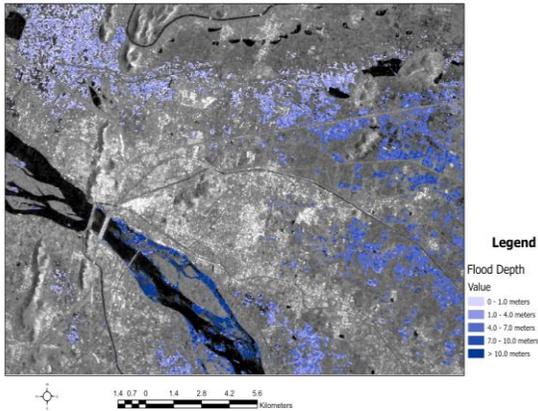*Figure 1:Flood Inundation Map of Vijayawada*

Figure 2:Flood Depth Map of Vijayawada

The two-stage machine learning model demonstrated a high efficacy. The classifier proved to be highly accurate in distinguishing between flood and non-flood days based on engineered features. Although the regressor was limited by the small number of historical flood events available for training, it provided reasonable order-of-magnitude estimates for flood severity. The two-model architecture is a dependable method for addressing this type of problem because it efficiently manages the imbalance in the dataset's classes.

The final Decision Support System (DSS), shown in Figures 3 and 4, integrates these components into an easy-to-use interface. It offers a 7-day forecast table with color-coded risk levels, enabling users to swiftly evaluate the situation and make informed decisions.
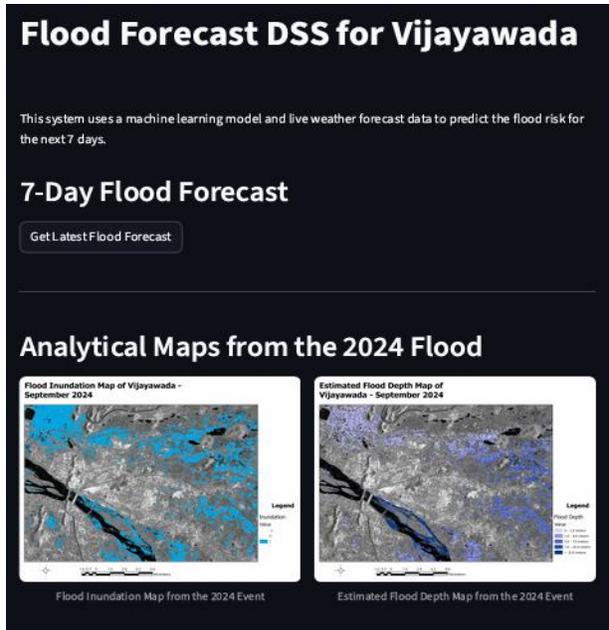

Figure 3: Flood Forecast DSS Interface before forecast
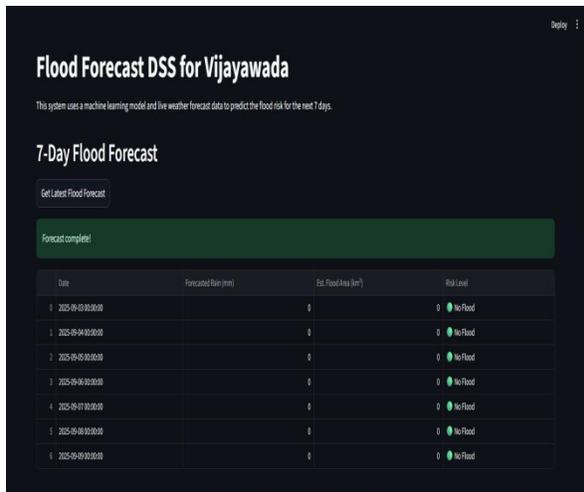

Figure 4: 7-Day Flood Forecast DSS Interface

VI. FUTURE SCOPE

The limitations of existing system is lack of previous flood event data to use in training regression models. The model is effective but is limited by the limited data high-magnitude events on which it has been trained. The future work should be aimed at improving the training datasets. This is possible by using the same satellite analysis workflow on other flood events in the area. The hyper parameters of the models can be further optimized to achieve the optimal model performance, and as dataset sizes grow,

more sophisticated time-series models like LSTMs, RNNs, and so on can be explored to obtain a better fit to the underlying time-dependent relationships. IOT devices can be added to improve the model by gathering sensor-based real-time data to forecast more accurate predictions of floods.

VII. CONCLUSION

This study successfully developed a DSS (Decision Support System) for flood forecasting for Vijayawada. This study effectively demonstrated the power of integrating two distinct but complementary methodologies: a detailed satellite-based geospatial analysis of a past event and a forward-looking, time-series-aware machine learning model. The key achievement of the geospatial phase was the successful reconstruction of the September 2024 flood, which yielded an inundation area map and a corresponding flood depth map. This provided a high-fidelity, quantitative "ground truth" measurement that is often missing in traditional flood prediction studies. By incorporating this satellite-derived data into a comprehensive historical dataset, we were able to train a robust two-stage predictive pipeline. The RandomForestClassifier proved highly effective at identifying the conditions leading to a flood, while the specialist RandomForestRegressor provided a quantitative estimate of the potential magnitude. The final DSS application that process live weather forecast data, represents a major improvement over the traditional alerts, to a practical and responsive forecasting tool. The framework presented in this paper provides a powerful and scale-able model for flood forecasting capable of utilizing satellite and weather information to gain information that can help to mitigate the impact of an imminent flood event.

REFERENCE

[1] N. Gauhar, S. Das, and K. S. Moury, "Prediction of Flood in Bangladesh using k-Nearest Neighbors Algorithm," in *2021 2nd International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST)*, 2021.

[2] Z. K. Lawal, H. Yassin, and R. Y. Zakari, "Flood Prediction Using Machine Learning Models: A Case Study of Kebbi State Nigeria," in *2021 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE)*, 2021.

[3] N. U. Reddy, P. V. Kumar, N. Kapileswar, J. Simon, and P. P. Kumar, "A Prediction Model for Minimization of Flood Effects using Machine Learning Algorithms," in *2022 Sixth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, 2022.

[4] S. S. L, S. P, C. P, U. C, and H. S. MS, "A Hybrid Flood Forecasting Approach Using Hidden Markov Model and Machine Learning Techniques," in *2023 7th International Conference on Computing, Communication, Control and Automation (ICCUBEA)*, 2023.

[5] D. Myrchiang, M. Ado, and K. Amitab, "Machine Learning-Based Flood Prediction of Assam," in *2023 International Conference on Computational Intelligence, Networks and Security (ICCINS)*, 2023.

[6] Ashok. P, K. J. S, S. N. K, and V. P. V, "Flood Prediction in the Area of Tamil Nadu and Andhra Pradesh Using Machine Learning Technique," in *2022 1st International Conference on Computational Science and Technology (ICCST)*, 2022.

[7] M. M. A. Syeed, I. Ishrar, M. Farzana, M. H. Nushra, I. Namir, and T. Rahman, "Flood Prediction Using Machine Learning Models," in *2022 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)*, 2022.

[8] M. Vimala, S. Ramasamy, P. R. Kumar, N. N. Sahi, A. Rajesh, and S. Siddharth, "Flood Prediction using Supervised Machine Learning Algorithms," in *2024 5th International Conference on Smart Electronics and Communication (ICOSEC)*, 2024.