IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)
An International Open Access, Peer-reviewed, Refereed Journal

# Flood Sense

An AI-Powered Flood Prediction System

[1]Mrs. Kodur Srividya, [2]Vilas V, [3]Vishal Kaman, [4]Sheetal Naik, [5]Sunidhi P
[1] Assistant Professor, [2]Student, [3]Student,[4]Student,[5]Student [1]Department of Computer Science and Engineering,
[1]K. S. Institute of Technology, Bangalore, India

***Abstract:*** Floods pose a significant threat to human life, infrastructure, and the economy. This paper presents Flood Sense, an advanced flood prediction system powered by machine learning. The system integrates historical meteorological data, real-time hydrological parameters, and satellite imagery to provide early warnings and risk assessments. Various machine learning techniques, including Decision Trees, Random Forest, and Artificial Neural Networks (ANN), are employed to enhance predictive accuracy. A web-based dashboard, built using Flask, enables real-time monitoring and alert dissemination. The goal of this system is to aid government agencies, disaster management teams, and local communities in making informed decisions to mitigate flood damage.

**Keywords:** Flood Prediction, Artificial Intelligence, Real- Time Monitoring, Disaster Management, Hydrological Analysis, Remote Sensing.

## I. Introduction

Prolonged heavy rainfall often leads to devastating flash floods, causing severe damage to the environment, economy, and livelihoods of people in affected regions. In India, where monsoon rains are a critical factor in agriculture and water resource management, the impact of floods can be particularly severe, leading to loss of life, destruction of property, and disruption of essential services. These floods not only result in immediate damage but also have long-term economic consequences, including the displacement of communities, loss of agricultural productivity, and the deterioration of infrastructure. However, despite the inevitable occurrence of floods due to natural climatic conditions, the extent of losses can be significantly mitigated through effective flood management strategies implemented before a flood occurs. The implementation of flood early warning systems is crucial in ensuring timely and proactive responses to flood events, ultimately reducing the risks and damages associated with severe flooding.

Flood early warning systems serve as an essential component in disaster management, enabling authorities and communities to prepare for and respond to impending flood events. These systems rely on real-time data monitoring and alert mechanisms, ensuring that early warnings are disseminated promptly to relevant stakeholders. The reliability and accuracy of flood warnings are of utmost importance, as they determine the effectiveness of response measures. A delay or inaccuracy in prediction can lead to erroneous decision-making, which may either cause unnecessary panic or, conversely, result in inadequate preparedness.

Accurate spatial and temporal prediction models are therefore necessary to minimize errors in forecasting flood events and to support effective crisis management. The more precise the predictions, the better the authorities and communities can prepare, reducing casualties, economic loss, and infrastructure damage.

The flood early warning system is designed to encompass four major functions: monitoring, prediction, warning, and response. These functions work in an integrated manner to ensure efficient flood management. Monitoring involves collecting real-time data from telemetry stations installed along riverbanks, which record critical parameters such as water levels, rainfall, and weather conditions. This data is then analyzed and utilized for predictive purposes to estimate potential flood events. However, one of the significant limitations of the data collected by telemetry stations is that it provides only localized measurements at the specific locations where the stations are installed. These localized readings might not always accurately represent the overall water level across an entire river basin or flood-prone area, making it challenging to assess the true extent of an impending flood. To address this issue, spatial interpolation techniques are applied to generate a comprehensive representation of water levels across a broader region. This approach allows authorities to better identify areas at risk of flooding and develop targeted response strategies accordingly.

One of the most effective ways to assess flood risk is through the creation of flood hazard maps. These maps are developed based on historical flood data, hydrological models, and real-time telemetry data. Flood hazard maps provide a spatial visualization of flood-prone areas, helping planners and emergency responders to identify vulnerable regions and take preventive measures accordingly. However, spatial data alone is not sufficient for accurately predicting floods over extended periods.

The dynamic nature of floods requires continuous updates and the integration of temporal data to improve the accuracy and effectiveness of predictions. Temporal data accounts for changes in water levels over time, allowing for a more precise and responsible flood forecasting system. By combining both spatial and temporal data, a more reliable and comprehensive flood prediction model can be developed, improving disaster preparedness and response efforts.

To enhance the accuracy of flood hazard mapping and prediction, this paper proposes a framework that integrates both spatial and temporal data to generate comprehensive flood hazard assessments. The proposed framework leverages interpolation techniques to process telemetry station data along with a temporal prediction model. This integration enables the generation of high-accuracy flood hazard maps, which are vital for flood management authorities. The core of this framework consists of two key components: the temporal prediction model and the utilization of interpolated spatial data. The temporal prediction model is based on advanced machine learning techniques, specifically the Long Short- Term Memory (LSTM) model. LSTM, a specialized type of recurrent neural network (RNN), is designed to handle sequential data and is particularly effective in time-series forecasting. This capability makes it highly suitable for predicting flood events by analyzing past water level data and identifying patterns that indicate potential future flooding.

The temporal prediction model is further divided into two categories: hourly and daily temporal predictions. This distinction is essential because the movement of water through a river system is not instantaneous; the time taken for water to travel from upstream to downstream locations often exceeds an hour. An hourly prediction model helps in immediate flood warnings, while a daily prediction model aids in longer-term flood preparedness and planning. This dual-layered prediction approach ensures a more comprehensive flood forecasting system, catering to both short-term emergencies and long- term flood risk management.

In addition to the temporal prediction model, the framework utilizes spatial data interpolation techniques to enhance the accuracy of flood hazard mapping. Specifically, the inverse distance weighting (IDW) method is applied to interpolate water level data from telemetry stations. The IDW method is a widely used geostatistical technique that estimates unknown values based on the values of nearby known data points, giving greater weight to closer observations. In this context, IDW is used to interpolate water levels along the river, ensuring a more complete and continuous representation of water distribution across the study area.

This interpolated spatial data is then compared with the Digital Elevation Model (DEM) of the surrounding region to assess the extent to which floodwaters may spill over riverbanks and inundate adjacent areas. The DEM provides critical topographical information, enabling the estimation of floodwater movement and the identification of regions at high risk of flooding.

By integrating temporal prediction with spatial interpolation, the proposed framework offers a robust methodology for flood forecasting and hazard mapping. The combination of LSTM-based temporal prediction and IDW- based spatial interpolation ensures that both time-dependent and geographical factors are accounted for in flood assessment. This comprehensive approach allows disaster management authorities to make informed decisions regarding evacuation plans, resource allocation, and flood mitigation strategies. Furthermore, the framework enhances the efficiency of flood early warning systems, reducing response times and improving the overall effectiveness of flood preparedness measures.

The application of this framework in real-world flood- prone areas, such as the Mahanadi River basin in India, has the potential to significantly improve flood risk management. By leveraging modern machine learning techniques and spatial analysis tools, authorities can develop highly accurate flood hazard maps that provide real-time insights into flood risks. These maps can be instrumental in guiding urban planning, infrastructure development, and emergency response initiatives. Additionally, the framework can be extended to other flood-prone regions across India, contributing to a nationwide effort to enhance flood resilience and disaster preparedness.

## II. LITERATURE SURVEY

Flood prediction is an essential component of modern disaster risk management. Historically, simulation-based hydrological models have been the core approach to predicting flood events. These models rely on mathematical representations of the water cycle, using rainfall, terrain, and river flow data to anticipate flood behaviour. However, such methods are often limited by the need for precise calibration, significant computational resources, and static input conditions. Buahin et al. [1] emphasized that while these models have been effective in structured applications, they lack the flexibility to incorporate dynamic, real-time data for immediate forecasting needs.

To complement hydrological simulation, researchers have employed spatial flood susceptibility mapping to identify regions at risk. Tehrany et al. [2] proposed an innovative model that integrates the Weights-of-Evidence technique with Support Vector Machines (SVMs), offering a robust GIS-based method for identifying flood-prone zones. This hybrid approach benefits from machine learning’s ability to handle complex variable interactions while still leveraging historical flood data and topographic indicators.

As flood forecasting evolves, the role of machine learning—particularly deep learning—has grown significantly. One of the most effective tools for handling sequential environmental data is the Long Short-Term Memory (LSTM) network, introduced by Hochreiter and Schmidhuber [4]. LSTM is known for retaining long-term dependencies in data streams, making it highly suitable for modeling temporal variables like rainfall and river discharge. In their work, Fang et al. [5] applied LSTM to hydrological forecasting and demonstrated its superior accuracy compared to traditional techniques. Their study confirmed that LSTM can adapt to continuous, time-dependent input, making it ideal for real-time flood prediction systems.

In addition to time-series analysis, spatial prediction techniques are vital for estimating the geographic spread of floodwaters. Interpolation methods help fill the gaps in data collected from limited monitoring stations. Among these, Inverse Distance Weighting (IDW) is widely used due to its simplicity and effectiveness in real-time applications. A detailed review by Li and Heap [6] compared various interpolation methods, highlighting IDW’s practicality for environmental data analysis. Supporting this, Patil and Kulkarni [7] found IDW to perform consistently well in hydrological applications where spatial continuity is needed despite limited data points.

While each of these methods contributes valuable insights and tools, they often function in isolation—focusing on either time- based prediction or spatial mapping. The Flood Sense system bridges this gap by integrating LSTM for predicting water levels over time and IDW for generating flood maps based on telemetry data. This combined approach offers both temporal accuracy and spatial clarity. Unlike traditional systems that rely on fixed models or purely statistical methods, Flood Sense adapts to real- time data inputs and presents outputs in a geographically contextualized format using digital elevation models (DEMs). As a result, it supports quicker decision-making and improved situational awareness during flood events, positioning it as a more effective and responsive solution than many existing approaches.

In addition, the models require detailed data that are difficult to collect for large areas. Non-deterministic methods are the statistical method and machine learning method.

The model applied historical flood to determine areas of flooding. the method consists of the water level rainfall and environmental data on flooding. The statistical method, For example, [8] logistic regression (LR), and [9] frequency ratio (FR). Machine learning methods, For example [8] artificial neural network (ANN), [5] long short-term memory (LSTM). In summary, two reasons prove why the performance of machine learning models is better than the statistical models [10]. First, machine learning models can perform complex tasks with limited information; second, machine learning models can represent large complex nonlinear systems in a computationally efficient. The principles of LSTM are designed to effectively capture and learn from sequenced data. This paper applies the long short-term memory (LSTM) method in the temporal prediction model.

The resulting temporal data also requires conversion into spatial data using spatial interpolation methods. As mentioned previously, there are several spatial interpolation models such as IDW, Spline, and Kriging. [11] comparison of interpolation methods for depth to groundwater is presented. The results show that IDW has the lowest Relative Error Coefficient compared to other techniques. The IDW technique is simple and effective, especially when the number of data points is small. Thus, this study applied the IDW to interpolate water level data in a river line in terms of spatial data.

## III. FLOOD BASICS

### A. Flood Early Warning Systems

Flood early warning systems provide real-time data from telemetry stations to provide timely and effective warnings, and to help risk areas in preparing for and responding to floods. Water level data are considered an important factor in the flood prediction model because they are the basic data for estimating future water levels. Monitored against predefined warning and critical thresholds which water levels the issuance of flood warnings when these thresholds are breached [2].

Moreover, the lag time of river water levels between two stations is another significant factor in the prediction model. The lag represents the time takes for the water to travel from the upstream peak to the downstream. This calculation helps determine the duration it takes for the water to reach its destination. The water levels can be combined with other data (e.g., discharge, rainfall, and flood history) to enhance the accuracy of the prediction model in terms of temporal data.

### B. Flood Hazard Maps

Flood hazard maps are maps that show the flood event displayed on a geolocation map. Generally, the flood hazard mapping methods can be divided method into two categories:

(1) deterministic methods and (2) non-deterministic methods. Deterministic methods are used by fundamental principles to simulate and predict flooding through a set of mathematical equations. In [7], the model was used to simulate floodplain inundation. However, this method was usually designed under specific assumptions and cannot fit other real-world conditions.

In addition, the models require detailed data that are difficult to collect for large areas. Non-deterministic methods are the statistical method and machine learning method.

The model applied historical flood to determine areas of flooding. the method consists of the water level rainfall and environmental data on flooding. The statistical method, For example, [8] logistic regression (LR), and [9] frequency ratio (FR). Machine learning methods, For example [8] artificial neural network (ANN), [5] long short-term memory (LSTM). In summary, two reasons prove why the performance of machine learning models is better than the statistical models [10]. First, machine learning models can perform complex tasks with limited information; second, machine learning models can represent large complex nonlinear systems in a computationally efficient. The principles of LSTM are designed to effectively capture and learn from sequenced data. This paper applies the long short-term memory (LSTM) method in the temporal prediction model.

The resulting temporal data also requires conversion into spatial data using spatial interpolation methods. As mentioned previously, there are several spatial interpolation models such as IDW, Spline, and Kriging. [11] comparison of interpolation methods for depth to groundwater is presented. The results show that IDW has the lowest Relative Error Coefficient compared to other techniques. The IDW technique is simple and effective, especially when the number of data points is small. Thus, this study applied the IDW to interpolate water level data in a river line in terms of spatial data.

## IV. SPATIAL TEMPORAL FLOOD MAPPING FRAMEWORK

This paper proposes a framework to apply the temporal and spatial data to generate flood hazard mapping using the interpolation of telemetry station data and a temporal prediction model. The framework integrates the temporal prediction model and the spatial data to generate flood hazard mapping. The prediction model consists of hourly and daily temporal prediction models. These models are applied to predict multiple outputs of the future water at each telemetry station.

The water level verifies the warning level from the telemetry station at each location. When the water level is higher than the Fig.

1. The Proposed Spatial Temporal Flood Hazard Mapping Framework warning level, the daily temporal prediction model is applied to predict the water level for the next day. The flood hazard map is created from the prediction values that are converted to spatial data with the IDW interpolation model, and it includes a digital elevation model for predicting flood events in this area next 3 days represented on the map, as shown in Fig. 1.

### 1. The Temporal Prediction Model

The LSTM which is a type of recurrent neural network that can learn long-term dependencies in time series data has been applied in the proposed framework. The LSTM consists of cell states and gates. The cell state is stored as the state of the memory cell. The gate is the controller of the data flow which is analogous to values control when to read, write, or forget. Thus, The LSTM is the temporal prediction model which is the process of predicting the future multiple output water level. The temporal prediction model consists of hourly and daily temporal prediction models, as shown in Fig. 1.

The hourly temporal prediction model takes data inputs from two sources: the water level from the telemetry station outside the watershed area which is the upstream station and within the watershed area. It is applied the data from the previous three hours, represented as, where Up is the upstream telemetry station, N represents the number of telemetry stations in the watershed area, and i is the previous time. The outputs consist of water level values for the next three hours from the telemetry stations within the watershed area. The output of this method is taken as one of the inputs in daily temporal prediction models.

The daily temporal prediction model takes data inputs from the telemetry stations within the watershed area. The inputs of the model are the water level of the last three values to predict the next three days. The input is the output from the hourly temporal prediction model which is higher than the warning level of the telemetry station at each location. Then, the output is converted to spatial data to generate the flood hazard map.

2. The Spatial Module Evaluation

Flood hazard maps are generated using the outputs of the daily temporal prediction model. The water level at each telemetry station is interpreted as the water level along the river to create spatial data for the river water levels within the watershed. In this paper, the Inverse Distance Weighting (IDW) model is applied for interpolation. The model is interpolated and estimates unknown values based on its distance from known values points [12]. The unknown values are calculated with the weighted average of the values at the known points. The weights are determined by the inverse of the distance between the unknown and known points. Thus, the river will know the water level values for all points, even if there is no telemetry station installed at that point. The data will be converted into spatial data. The DEM values are compared with the water level values to identify potential flood events. Finally, the flood hazard maps represent the flood events in this area on the map.

## V. SOLUTION FLOW

The machine learning pipeline for flood prediction consists of several interconnected stages, each contributing to the development of an accurate and reliable model. The stages as illustrated in Fig 1 are detailed as follows:

### 4.1 Data Collection

The pipeline begins with the collection of data from four primary sources: historical data, sensor data, weather forecasts, and hydrological data. Historical data includes past records of rainfall, river levels, and flood occurrences, providing a baseline for pattern recognition. Sensor data, collected in real-time, captures current environmental conditions such as water levels and soil moisture. Weather forecasts provide predictive insights into upcoming meteorological events, while hydrological data encompasses information about water flow, drainage systems, and terrain characteristics. These diverse inputs are aggregated to form a comprehensive dataset for analysis.

### 4.2 Data Preprocessing

The collected data undergoes preprocessing to ensure quality and consistency. This stage involves cleaning the data by handling missing values, removing outliers, and correcting inconsistencies. Normalization and standardization techniques are applied to scale the data, making it suitable for machine learning algorithms. Additionally, data from different sources is aligned temporally and spatially to enable seamless integration.

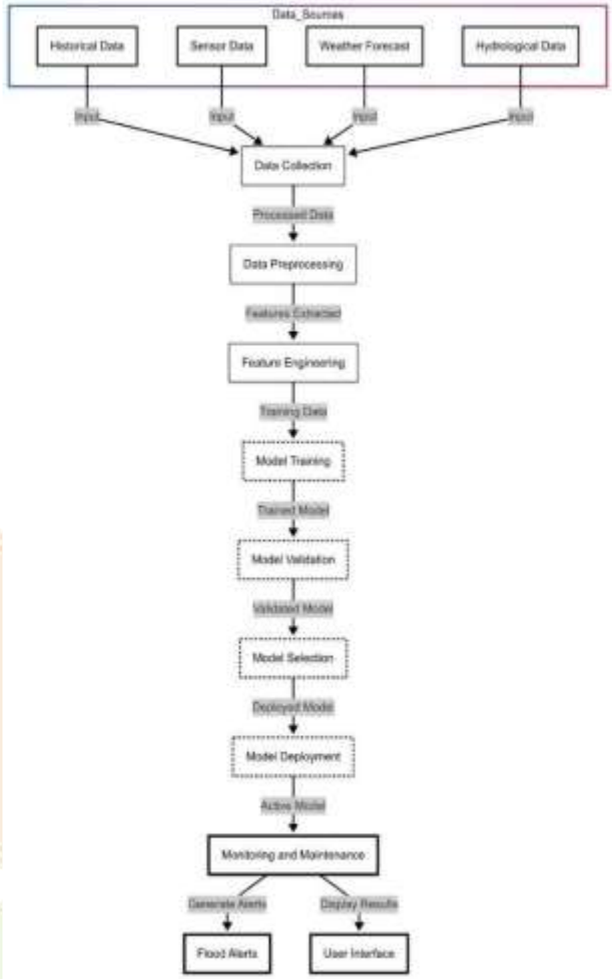



Fig 1: Solution Architecture

### 4.3 Feature Engineering

Feature engineering transforms the preprocessed data into meaningful features that enhance the predictive power of the model. This stage involves extracting relevant variables, such as rainfall intensity, river flow rates, and soil saturation levels, which are critical for flood prediction. Feature selection techniques, such as correlation analysis and principal component analysis (PCA), are employed to reduce dimensionality and eliminate redundant or irrelevant features, thereby improving model efficiency. Fig 2 shows an aspect of feature selection.

Fig 2: Accuracy score and feature selection

Fig 3 illustrates confusion matrix feature extraction in the web application.


Fig 3: Confusion matrix

### 4.4 Model Training

The engineered features are used to train a machine learning model. The training data, a subset of the processed dataset, is fed into the model to learn patterns and relationships between input features and flood occurrences. Common algorithms for flood prediction include decision trees, random forests, and neural networks, selected based on their ability to handle complex, non- linear relationships in environmental data.

### 4.5 Model Validation

The trained model is evaluated using a separate validation dataset to assess its performance. Metrics such as accuracy, precision, recall, and the F1-score are calculated to measure the model's ability to predict floods accurately. Cross- validation techniques, such as k-fold cross-validation, are employed to ensure the model generalizes well to unseen data and is not overfitting to the training set.

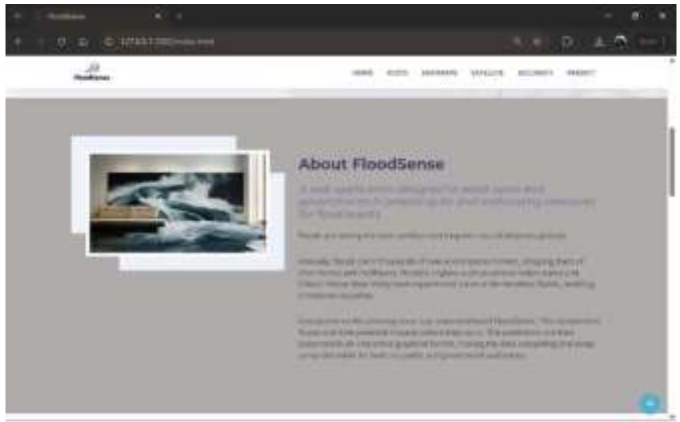
### 4.6 Model Selection

Based on the validation results, the best-performing model is selected for deployment. This stage involves comparing multiple models or configurations to identify the one with the highest predictive accuracy and robustness. Hyperparameter tuning, using techniques like grid search or random search, may be performed to optimize the model's performance further.

### 4.7 Model Deployment

The selected model is deployed into a production environment, where it becomes an active model capable of making real-time predictions. The deployment process involves integrating the model with a system that continuously feeds it with new data from the same sources used during training. The active model generates predictions about potential flood events based on incoming data.
Fig 4 demonstrates the about page of flood sense web application present on the scrollable homepage.

Fig 4: About Flood Sense




### 4.8 Monitoring and Maintenance

Continuous monitoring is essential to ensure the model's performance remains consistent over time. This stage involves tracking the model's predictions against actual outcomes to detect any degradation in accuracy, often caused by data drift or changes in environmental patterns. Maintenance activities, such as retraining the model with updated data or fine-tuning its parameters, are performed to sustain its effectiveness.

### 4.9 Output Generation

The pipeline culminates in two primary outputs: flood alerts and a user interface. The active model generates alerts when it predicts a high likelihood of flooding, enabling timely interventions by authorities or communities. Simultaneously, the predictions and relevant data are displayed through a user interface, providing stakeholders with an accessible and interpretable visualization of the results.

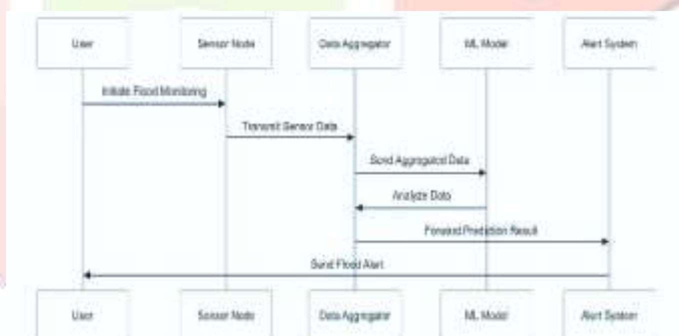



Fig 5: Sequence diagram

Fig 5 shows the sequence diagram of the web application flood sense.
The flood monitoring system operates through a series of interconnected stages, each performed by a specific component. The workflow is designed to ensure seamless data flow and accurate predictions, as detailed below.

### 4.10 Initiation of Flood Monitoring

The process begins with the user, who initiates flood monitoring. This step typically involves activating the system through a user interface, such as a mobile application or a web portal. The user's action triggers the sensor nodes deployed in the target area, signaling the start of data collection. This initiation step is crucial, as it allows the system to operate in real-time, responding to user demands or predefined schedules for monitoring flood-prone regions.

### 4.11 Transmission of Sensor Data

Upon activation, the sensor nodes collect environmental data, such as water levels, rainfall intensity, and soil moisture, from the monitored area. These nodes, often deployed in a network across rivers, floodplains, or urban drainage systems, transmit the collected data to a data aggregator. The transmission is typically performed using wireless communication protocols, such as LoRa or Zigbee, ensuring efficient and reliable data transfer even in remote or challenging environments.

### 4.12 Data Aggregation

The data aggregator receives sensor data from multiple nodes and consolidates it into a unified dataset. This stage involves preprocessing tasks, such as filtering noise, handling missing values, and aggregating data over time or space to reduce redundancy. The aggregator ensures that the data is structured and ready for analysis, minimizing the computational load on the subsequent ML model. This step is essential for improving the quality of input data, which directly impacts the accuracy of flood predictions.

### 4.13 Data Analysis Using Machine Learning

The aggregated data is sent to the ML model, which analyzes it to predict potential flood events. The ML model, typically trained on historical and real-time data, employs algorithms such as decision trees, support vector machines, or deep neural networks to identify patterns indicative of flooding. The analysis involves processing features like rainfall trends, water level changes, and geographical factors to generate a prediction result. This stage is the core of the system, as it transforms raw data into actionable insights, enabling the system to forecast flood risks with high accuracy.

### 4.14 Forwarding Prediction Results

Once the ML model completes its analysis, it forwards the prediction result to the alert system. The prediction result typically includes a probability or classification of flood risk (e.g., low, medium, or high) along with relevant metadata, such as the predicted time and location of the flood. This step ensures that the alert system receives the necessary information to generate appropriate notifications, bridging the gap between data analysis and user communication.

Fig 6 displays the location safety information of a location (i.e. Bangalore in this case) with respect to floods.

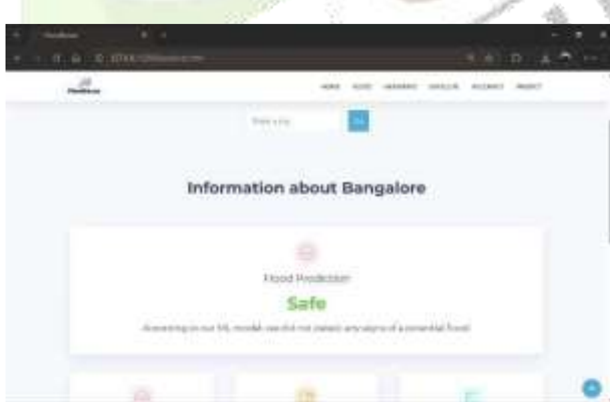



Fig 6: Location safety check

# IV. RESULTS AND DISCUSSION

## 4.1 Study Area

The study area is the Ping River basin covering six districts in Chiang Mai province, Thailand. The river length is 118 km. The watershed covers an area is 2,146 km2. The river is the main river of Chiang Mai province, and it flows through economic and community areas. Flooding can have a significant impact on the province. Fig. 2 shows the study area and the position of the telemetry stations. The water level monitoring in this area is conducted by four telemetry stations and one upstream station, named P.1, P.67, P.75, and P.20, respectively. The lag time between P.1 to P.20 takes approximately 24 hours. The telemetry has the hourly observed water level data.

Fig 7 highlights the accuracy precision and F-1 score of the existing and proposed system.

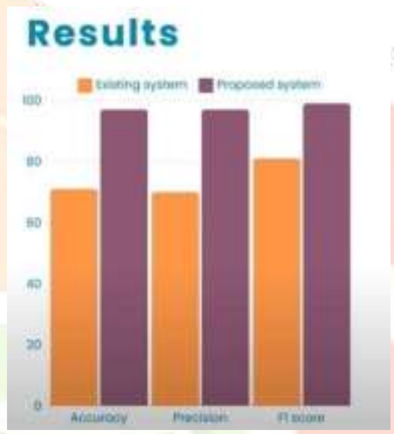



Fig 7: Existing vs Proposed system

Figure 7 provides a detailed comparative assessment of the predictive performance of the existing flood forecasting approaches versus the proposed Flood Sense system.

The evaluation is conducted using three standard classification metrics: accuracy, precision, and F1-score, each reflecting different aspects of model effectiveness in real-world flood detection scenarios.
The existing system, typically reliant on traditional statistical models or simpler machine learning algorithms, demonstrates moderate effectiveness. These models often operate on limited input features, are less adaptive to real-time variations, and lack the integration of both spatial and temporal data. As a result, their predictive capabilities are constrained, leading to increased instances of false positives and false negatives. This is reflected in their comparatively lower accuracy and F1-scores, indicating reduced reliability and balance in prediction. The lower precision values further suggest that many of the flood alerts generated by existing methods are inaccurate, which can lead to unnecessary panic or inefficient deployment of emergency resources.

In contrast, the proposed Flood Sense system incorporates a hybrid architecture that addresses these limitations by combining Long Short-Term Memory (LSTM) networks for capturing temporal patterns in water level fluctuations and Inverse Distance Weighting (IDW) for generating continuous spatial data from discrete telemetry inputs. This dual-layered approach enables the model to account for both time-based trends and geographical variation in river systems, resulting in more informed and context-aware flood predictions.

The integration of historical and real-time data enhances the temporal accuracy, while spatial interpolation ensures that the absence of direct sensor data at certain locations does not compromise the system's performance. Consequently, the proposed model achieves notably higher accuracy, indicating a better match between predicted and actual flood occurrences. Its improved precision demonstrates that the model is more selective in issuing flood warnings, minimizing false alarms. The higher F1- score underscores its balanced handling of both false positives and false negatives, making it more dependable in operational settings.

Overall, the results displayed in Figure 7 underscore the superiority of the proposed system in terms of prediction reliability, responsiveness, and practical applicability. The use of deep learning and spatial modeling together offers a robust solution to the complexities inherent in flood forecasting, enabling timely alerts and more effective disaster response planning.

## 4.2 Performance Metrics and Parameter Settings

The performance evaluation of the prediction model is the accuracy based on the mean absolute percentage error (MAPE). It calculates the mean of the absolute percentage error difference between predicted values and observed values error as depicted Eq. (1).

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{O_i - P_i}{O_i}\right| * 100$$

Eq. (1): Mape calculation formula

Where Oi represents the observed water level value from the telemetry station, Pi represents the predicted water level value from the telemetry station and n is the number of data. The performance evaluation of the spatial data applies the confusion matrix technique. The technique summarizes the performance of a classification model consisting of the number of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) of the prediction model. The confusion matrix technique can calculate the metrics including accuracy, precision, recall, and F1-score.

The dataset of the case study covers the period from April 2016 to March 2023 with 61,345 hourly records and 2,557 daily records. The training dataset is 80% of the data with 49,077 hourly and 3,196 daily records. The testing dataset is 20% remaining 20%, with 12,268 hourly records and 640 daily records. The LSTM model parameters include 16 inputs for the hourly model, 9 inputs for the daily model, and 9 outputs for both. The hidden layers consist of 64 neurons for each model, with each model trained between 100 - 500 iterations.

## 4.3 Temporal Model Evaluation

The temporal prediction model is evaluated using the MAPE. The evaluation is conducted at iterations of 100, 300, and 500, with results as shown in Table I. The table shows the metrics including the minimum, mean, and maximum of the training and testing process. The MAPE of hourly and daily temporal prediction models for the training processes have accuracy are 3.17% and 4.88%. The testing processes have accuracy are 3.48% and 4.72%, respectively.

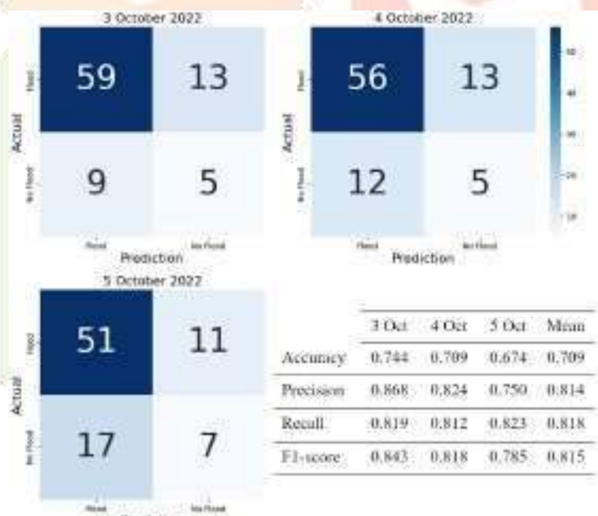
Each station has four line graphs representing the water level predictions for the next three hours and the observed values. The yellow line represents the warning level, and the red line indicates the critical level. These graphs shows the water level peak in October 2022.

## 4.4 Map Evaluation

In Fig. 8, the flood hazard map represents the Ping River in Chiang Mai province from 3 to 5 October 2022.The output exceeded the warning level on 2 October 2022 from the daily temporal prediction model. During this period, Chiang Mai province suffered extensive damage from Typhoon Noru. The observation map is data from the Geo-Informatics and Space Technology Development Agency (GISTDA) in

Thailand, which only provides monthly-scale data. The accuracy of the flood hazard map was evaluated at the sub-district level by predicting flood occurrence covering 86 sub-districts, as follows in Fig 9. The overall prediction accuracy for October is 70.90%. However, the F1-score is 81.50%, as follow Fig 9.


Fig 8: Flood Map with Typhoon Noru on October 2022




| | 3 Oct | 4 Oct | 5 Oct | Mean |
|---|---|---|---|---|
| Accuracy | 0.744 | 0.709 | 0.674 | 0.709 |
| Precision | 0.868 | 0.824 | 0.750 | 0.814 |
| Recall | 0.819 | 0.812 | 0.823 | 0.818 |
| F1-score | 0.843 | 0.818 | 0.785 | 0.815 |

Fig 9: Confusion Matrix to Flood Map Evaluation on October 2022

## VI. CONCLUSION

This paper proposes to integrate the spatial and temporal data to generate flood hazard mapping using interpolation telemetry station data and a temporal prediction model. The case study is in the Ping River basin, Chiang Mai province, which is an economically significant area in the province. The evaluation results of the proposed model show the MAPE values of the hourly and daily temporal prediction model using LSTM are 3.17% and 4.88% in the training process.

Moreover, the accuracy result of the flood hazard map is 70.90% and the F1-score is 81.50% of the model from the confusion matrix technique. The flood hazard map result shows low accuracy because of the monthly cumulative nature of the observation data, whereas the proposed model operates at a daily scale, leading to high bias. Future work will further enhance the correlation parameters with temporal and spatial data, and adjust the prediction model to support variation of parameters.
In conclusion, the integration of spatial and temporal data through advanced machine learning and geospatial analysis techniques represents a significant advancement in flood forecasting and hazard mapping. The proposed framework addresses the limitations of traditional flood prediction methods by combining real-time telemetry data, machine learning-based temporal forecasting, and geostatistical interpolation. By doing so, it provides a more reliable and actionable approach to flood risk assessment, ultimately helping to mitigate the devastating impacts of floods on communities, economies, and ecosystems.

## REFERENCES

[1] H. Hapuarachchi, Q. Wang, and T. Pagano, "A review of advances in flash flood forecasting," Hydrological processes, vol. 25, no. 18, pp.2771–2784, 2011.
[2] A. Wannachai, S. Aramkul, B. Suntaranont, Y. Somchit, and P. Champrasert, "Hero: Hybrid effortless resilient operation stations for flash flood early warning systems," Sensors, vol. 22, no. 11, p. 4108, 2022. Fig. 5. Confusion Matrix to Flood Map Evaluation on October 2022
[3] S. Zhong, C. Wang, Z. Yu, Y. Yang, and Q. Huang, "Spatiotemporal exploration and hazard mapping of tropical cyclones along the coastline of china," Advances in Meteorology, vol. 2018, 2018.
[4] C. Hu, Q. Wu, H. Li, S. Jian, N. Li, and Z. Lou, "Deep learning with a long short-term memory networks approach for rainfall- runoff simulation," Water, vol. 10, no. 11, p. 1543, 2018.
[5] Z. Fang, Y. Wang, L. Peng, and H. Hong, "Predicting flood susceptibility using lstm neural networks," Journal of Hydrology, vol. 594, p. 125734, 2021.
[6] F.-W. Chen and C.-W. Liu, "Estimation of the spatial rainfall distribution using inverse distance weighting (idw) in the middle of taiwan," Paddy and Water Environment, vol. 10, pp. 209–222, 2012.
[7] W. Li, K. Lin, T. Zhao, T. Lan, X. Chen, H. Du, and H. Chen, "Risk assessment and sensitivity analysis of flash floods in ungauged basins using coupled hydrologic and hydrodynamic models," Journal of Hydrology, vol. 572, pp. 108–120, 2019.
[8] M. Rahman, C. Ningsheng, M. M. Islam, A. Dewan, J. Iqbal, R. M. A. Washakh, and T. Shufeng, " Flood susceptibility assessment in bangladesh using machine learning and multi-criteria decision analysis,"Earth Systems and Environment, vol. 3,
pp. 585–601, 2019.
[9] M. S. Tehrany, B. Pradhan, and M. N. Jebur, "Flood susceptibility analysis and its verification using a novel ensemble support vector machine and frequency ratio method, " Stochastic environmental research and risk assessment, vol. 29, pp. 1149–1165, 2015.
[10] A. B. Massada, A. D. Syphard, S. I. Stewart, and V. C. Radeloff, " Wildfire ignition-distribution modelling: a comparative study in the huron–manistee national forest, michigan, usa," International journal of wildland fire, vol. 22, no. 2, pp. 174–183, 2012.
[11] Y. Sun, S. Kang, F. Li, and L. Zhang, "Comparison of interpolation methods for depth to groundwater and its temporal and spatial variations in the minqin oasis of northwest china," Environmental Modelling & Software, vol. 24, no. 10, pp. 1163– 1170, 2009.
[12] G. Y. Lu and D. W. Wong, "An adaptive inverse-distance weighting spatial interpolation technique," Computers & geosciences, vol. 34, no. 9, pp. 1044–1055, 2008.