



# Flood Prophecy Using Machine Learning Algorithms

Angel Barakka J, Joel Stelton M, Bhuvaneshwaran K, Prashana R  
Department of Computer Science and Business Systems  
Sri Sairam Engineering College, Chennai, Tamil Nadu, India.

**Abstract** —Our project revolves around the development of an advanced flood warning system aimed at significantly enhancing disaster response efforts. At its core, the system harnesses the power of machine learning to predict and mitigate the impact of potential flooding events. A pivotal aspect of our approach is the creation of a user-friendly interface accessible to the general public. Through this interface, individuals can access crucial information regarding the likelihood of flooding in their respective regions. By utilizing historical data on river flow and visualizing rainfall patterns at the sub-division level, users gain valuable insights into the potential risks they face. The methodology employed in our project places a strong emphasis on the utilization of machine learning algorithms. These algorithms analyze vast datasets to forecast future outcomes related to flooding with remarkable accuracy. Additionally, we prioritize the speed and timeliness of our predictive models, ensuring that users receive timely alerts and warnings well in advance of potential flood events. By proactively alerting communities to the possibility of flooding, our system aims to minimize the loss of life and property that often accompanies such disasters. Drawing on lessons learned from past incidents, such as the devastating floods that struck Tamil Nadu, we are committed to leveraging cutting-edge technology to create a safer and more resilient future for at-risk communities.

Keywords: Flood prediction, machine learning, forecasting floods.

## I. INTRODUCTION

Our project focuses on developing a machine learning (ML) model for flood prediction, utilizing river flow and rainfall data as primary inputs. The model is integrated into a user-friendly interface, enabling users to access and visualize predictions conveniently. By leveraging historical river flow and rainfall data, our ML algorithm generates forecasts of potential flood events with enhanced accuracy. Through iterative training and validation processes, we refine the model to capture the complex dynamics of flood occurrence, considering various contributing factors.

The user interface provides intuitive access to real-time and projected flood predictions, empowering stakeholders to make informed decisions and implement proactive measures. Users can interact with the interface to explore forecasted scenarios, assess flood risk levels, and plan mitigation strategies accordingly. Our project aims to bridge the gap between advanced data analytics and practical usability, facilitating effective flood management and disaster preparedness. By empowering users with actionable insights, we strive to enhance resilience in flood-prone regions and minimize the socio-economic impact of flooding events.

## Motivation

The escalating frequency and severity of floods in recent years highlight the pressing necessity for innovative solutions to anticipate and mitigate their profound impacts on both human settlements and natural environments. Floods not only endanger lives and property but also disrupt critical infrastructure, hamper economic activities, and exacerbate ecological degradation. Conventional flood management strategies often prove inadequate in accurately predicting and effectively responding to evolving flood risks. This underscores the imperative for novel approaches that leverage data-driven analytics and technology to deepen our comprehension of flood dynamics and enhance preparedness and response mechanisms.

Ultimately, our driving force emanates from a shared dedication to shielding lives, livelihoods, and ecosystems from the ravages of flooding. By harnessing the synergies of data and technology, we envision fostering a more resilient future wherein communities can thrive amidst the evolving challenges posed by climate variability.

## Relevance of the project

Our flood prediction project is highly relevant in light of the escalating frequency and severity of flooding events globally, exacerbated by both climate change and human activities. As urbanization encroaches into flood-prone areas and populations continue to grow, the vulnerability to flooding amplifies, presenting substantial challenges to regions across the spectrum of development. Accurate flood prediction holds profound relevance in mitigating the profound impacts of these events. By providing precise forecasts of flood occurrences, we empower communities and authorities to execute timely evacuation plans, allocate emergency resources effectively, and implement preventive measures to safeguard lives and assets.

Moreover, the significance of our project extends beyond immediate crisis response to encompass long-term planning and resilience-building endeavours. By furnishing stakeholders with dependable flood forecasts, we facilitate proactive land-use planning, infrastructure development, and disaster risk reduction strategies. This proactive stance not only mitigates the socio-economic toll of flood damage but also enhances community resilience and sustainability.

## Design Methodology

The flow begins at the "Home page," which serves as the central hub, and from there, it branches into two main paths: one leading to the "Rainfall Homepage" and the other to the "Flood Home page." Here's an elaborate explanation of each path:

### Rainfall Path:

1. Rainfall Home page: Users begin by accessing the Rainfall section of the system. Here, they are prompted to provide specific inputs – namely, the year for which they want rainfall data and the geographical subdivision of interest.
2. Data Submission: Once the user enters the year and selects the sub-division, this information is sent to the backend.
3. Backend Processing: The backend algorithm processes the data. This algorithm is designed to analyze historical rainfall data and, likely using statistical or machine learning techniques, predict rainfall for the given year and sub-division.
4. Display of Results: The outcome of this process is displayed to the user in a graphical format. This visualization compares actual recorded rainfall with the rainfall predicted by the system's algorithm, providing a clear comparison between the two datasets.

## Flood Path:

1. Flood Home page: Users looking for information on floods will navigate to this section. They must provide two pieces of information: the date for which they want flood data, and the name of the river they are interested in.

2. Date Specification: The system needs to understand whether the user is interested in past data or predictions for the future. This is crucial as it determines the subsequent processing steps.

3. Past Data Processing: If the user selects "Past," the system uses the input data to retrieve and display historical flood features, such as river discharge, runoff, and the actual flow level. The predictor in the backend is responsible for analyzing past data and possibly providing insights or validation against past predictions.

4. Future Data Processing: If the user is interested in future predictions, the input is first analyzed by a time series analyzer in the backend. Time series analysis is a common approach to forecast future events based on historical data. The output from the time series analyzer is then fed into a flow level predictor, which uses the analyzed data to predict future flood characteristics for the specified river, such as discharge, runoff, and expected flow level. These predictions are then displayed to the user, providing them with an estimate of future flood conditions.



## Abridgement

Flood forecasting is very important for disaster management. Accurate prediction can be made by analyzing many factors such as rainfall patterns, water levels and historical data. Use advanced modelling techniques and real-time monitoring to predict flood events. This leads to timely warning and evacuation, reducing the risk to human life and minimizing property damage. A good flood forecast ensures preparedness and allows authorities to effectively allocate resources to reduce the impact of natural disasters.

## II. RELATED WORKS

[1] This study delves into the intricate challenge of uncertainty within flood susceptibility modeling, particularly emphasizing the spatial disparities among different machine learning (ML)-based models. While ML-powered flood susceptibility models (FSMs) have demonstrated their potential in accurately pinpointing flood-prone areas, the inconsistency in their spatial predictions poses a significant hurdle for decision-makers and planners alike. To tackle this issue, the researchers investigated four distinct ML-based FSMs: random forest (RF), k-nearest neighbor (KNN), multilayer perceptron (MLP), and a hybrid model amalgamating genetic algorithm-Gaussian radial basis function-support vector regression (GA-RBF-SVR). These models were applied to the southwest coastal region of Bangladesh, a region highly susceptible to floods.

Despite achieving commendable accuracy in predicting flood-prone areas, the researchers observed notable spatial discrepancies across the model outcomes. This indicates that while the models were proficient in identifying areas prone to flooding, they generated divergent spatial predictions for the same geographic region. This variation can be attributed to the multifaceted nature of flood dynamics and the diverse array of factors influencing flood susceptibility.

In response to these spatial disparities, the researchers proposed a framework aimed at optimizing the amalgamation of the four ML-based models. By leveraging the unique strengths of each model and minimizing classification errors, the optimized model exhibited enhanced spatial agreement compared to the individual models. This implies that integrating multiple models can enhance the reliability and robustness of flood susceptibility predictions. The implications of this study are far-reaching, particularly in the realms of risk management and disaster preparedness in flood-prone areas. By mitigating uncertainty in flood susceptibility modeling, the framework proposed in this study holds promise for informing the development of more effective risk-based development plans and bolstering early warning systems. Ultimately, such

efforts can play a pivotal role in mitigating the adverse impacts of flooding and bolstering the resilience of vulnerable communities.

[2] India faces significant challenges due to flooding, as evidenced by recent disasters such as the devastating floods in Kerala in August 2018. While previous efforts have utilized Internet of Things (IoT) and machine learning (ML) techniques to predict floods based on various parameters like rainfall, humidity, temperature, water flow, and water level, there has been a gap in considering the possibility of flood occurrence based solely on temperature and rainfall intensity.

To address this gap, researchers have employed Deep Neural Network (DNN) models to forecast the occurrence of floods based specifically on temperature and rainfall intensity. This approach represents a novel application of deep learning techniques in flood prediction. Moreover, the study compares the performance of the DNN model with other traditional machine learning models such as Support Vector Machine (SVM), K-nearest neighbor (KNN), and Naïve Bayes. The comparison is made in terms of accuracy and error rates to assess the effectiveness of the DNN model in flood forecasting. The findings suggest that the DNN model exhibits the highest accuracy among the compared models when predicting floods based solely on monsoon parameters such as temperature and rainfall intensity. This indicates the potential of deep learning techniques to enhance flood forecasting capabilities, particularly in regions like India where flooding poses a significant threat to communities and infrastructure. By leveraging advanced machine learning techniques like DNN, researchers aim to improve the accuracy and reliability of flood forecasts, thereby enabling better preparedness and response strategies to mitigate the impacts of flooding in vulnerable regions. This research contributes to the ongoing efforts to develop more effective tools and methodologies for flood risk management and disaster resilience.

[3] Floods pose significant threats, causing loss of life and extensive property damage. Many cities, particularly those influenced by monsoons, face this recurring disaster annually. Timely notification of impending flood incidents is crucial for authorities and the public to implement preventive measures, plan evacuations and rescue operations, and provide relief to flood victims. Geographical locations and severity levels of affected areas are pivotal factors in flood management. However, existing flood forecasting tools often rely on manual data input and preparation, which are laborious and unsuitable for real-time forecasts. Moreover, these tools often underutilize the vast amount of data available in big data platforms. To address these challenges, this paper proposes a novel flood forecasting system that integrates meteorological, hydrological, geospatial, and crowd-sourced big data using an adaptive machine learning framework. Advanced learning strategies drive data intelligence, allowing the system to make accurate predictions. Both subjective and objective evaluations validate the system's capability to forecast flood incidents in specific areas and timeframes.

Benchmarking experiments revealed that the system configured with a Multilayer Perceptron Artificial Neural Network (MLP ANN) provided the most effective predictions. It achieved impressive metrics including a correct percentage of 97.93%, Kappa coefficient of 0.89, Mean Absolute Error (MAE) of 0.01, and Root Mean Square Error (RMSE) of 0.10. These results underscore the system's reliability and accuracy in flood forecasting, demonstrating its potential to significantly enhance early warning systems and disaster management efforts

[4] In recent years, significant strides have been made in weather forecasting and flood prediction, largely facilitated by Early Warning Systems (EWSs). However, despite ongoing efforts by scientific and user communities to innovate, EWSs predominantly focus on forecasting hazards, often overlooking information on potential consequences and impacts.

To address this gap, this study introduces a methodology for quantitative real-time impact assessment of flash floods. The approach utilizes a multi-model ensemble technique and incorporates soil moisture uncertainty. Furthermore, it extends the flood forecasting chain beyond merely providing discharge probabilities to include a fully 2D hydraulic model and a damage estimation model. This extension enables the quantitative assessment of impacts in terms of economic losses and the number of people affected. The efficacy of the methodology was tested using recent flood events in Genoa, located in north western Italy. The paper explores the potential challenges and opportunities presented by this approach within the decision-making workflow in an operational context.

By integrating impact assessment into the flood forecasting process, this methodology enhances the utility of EWSs by providing decision-makers with comprehensive information on potential consequences. This enables more informed and proactive decision-making, facilitating timely responses to mitigate the impacts of flash floods. Furthermore, the use of multi-model ensembles and consideration of soil moisture uncertainty enhance the reliability and robustness of impact assessments, further improving the effectiveness of flood management strategies. Overall, this approach represents a significant advancement in flood risk assessment and management, offering valuable insights for decision-makers and stakeholders involved in disaster preparedness and response efforts.

[5] The advent of the fourth industrial revolution, characterized by components such as cloud computing technology, artificial intelligence, big data, and the Internet of Things (IoT), has significantly impacted flood disaster mitigation strategies worldwide. Indonesia, in particular, has experienced a notable increase in flood events, nearly tripling over the past fifteen years. This literature review delves into the advancements in applying artificial intelligence, specifically machine learning, in flood mitigation within Indonesia. The study highlights that while machine learning techniques have been employed to some extent in flood disaster mitigation in Indonesia, their application has not yet been widespread. Existing applications primarily focus on predicting rainfall, river water levels, and discharge levels. However, the use of machine learning for flood forecasting and early warning systems remains limited.

Looking ahead, the study suggests potential future improvements in flood mitigation efforts in Indonesia. Specifically, it advocates for the integration of machine learning methods into comprehensive flood prediction and early warning systems. Such systems would encompass predictions for rainfall, river water levels, discharge levels, and flood occurrences. Moreover, they would incorporate estimations of the spatial extent of flooding areas. Additionally, the study proposes leveraging the Internet of Things for efficient dissemination of early warning alerts. By integrating machine learning techniques into flood prediction and early warning systems, Indonesia could enhance its capacity to mitigate the impacts of flooding. These advancements would enable more accurate and timely responses to flood events, ultimately reducing the associated risks to lives and properties. Furthermore, leveraging the capabilities of the Internet of Things for early warning dissemination could ensure that critical information reaches vulnerable communities swiftly, further improving overall disaster preparedness and response efforts.

[6] Floods pose significant challenges due to their complex nature, making them difficult to accurately model. Over the past two decades, research into flood prediction models has played a crucial role in mitigating risks, informing policy decisions, and reducing the loss of human life and property damage associated with floods. Machine learning (ML) methods have emerged as powerful tools in advancing flood prediction systems, offering improved performance and cost-effective solutions.

The appeal of ML in hydrology has soared, as researchers recognize its potential benefits. By introducing novel ML methods and hybridizing existing ones, researchers aim to develop more accurate and efficient flood prediction models. The primary objective of this paper is to showcase the state-of-the-art ML models in flood prediction and provide insights into the most suitable approaches. The paper extensively reviews literature where ML models were benchmarked through qualitative analyses of robustness, accuracy, effectiveness, and speed. This investigation offers a comprehensive overview of the various ML algorithms utilized in flood prediction. Through performance comparisons of ML models, the paper delves into different techniques, providing an in-depth understanding within a framework of comprehensive evaluation and discussion.

The findings highlight the most promising prediction methods for both long-term and short-term floods. Additionally, the paper examines major trends aimed at improving the quality of flood prediction models. Hybridization, data decomposition, algorithm ensemble, and model optimization emerge as effective strategies for enhancing ML methods. Overall, this survey serves as a valuable guideline for hydrologists and climate scientists in selecting the appropriate ML method for specific prediction tasks. By leveraging the insights provided, researchers can further advance flood prediction capabilities, ultimately contributing to better disaster preparedness and response strategies.

[7] Floods, exacerbated by unusual rainfall patterns and global climate change, have become increasingly prevalent worldwide. These catastrophic events wreak havoc on property, infrastructure, and human lives. In response to the need for effective flood prediction and mitigation, machine learning models have been

developed. Creating accurate predictive models for floods is inherently challenging due to the complexity of the phenomenon. In this system, rainfall data serves as the input for four distinct machine learning models. Before this process, the data undergoes cleaning and pre-processing to ensure its quality and suitability for analysis. The dataset is then divided into a Training set and a Test set in a 7:3 ratio to facilitate model training and evaluation.

Each machine learning model's accuracy is assessed, and confusion matrix parameters are utilized to further evaluate and analyze their performance. By comparing the accuracy metrics of each model, the best-performing one is selected for deployment. Overall, this approach leverages machine learning techniques to improve flood prediction capabilities, providing valuable insights for disaster preparedness and response efforts. Through rigorous evaluation and analysis, the most effective model can be identified, offering a valuable tool for mitigating the impacts of floods on communities and infrastructure.

### III. EXISTING AND PROPOSED

#### A. Existing system:

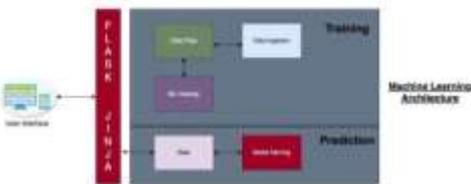
[2] This study presents a technique called RAPIDS, which stands for Radar Flood Identification of Drainage Systems. The aim is to predict floods in multiple urban areas using multiple generators, using the city's water supply system and region-specific rainfall data. The results show that a neural network can provide a good flood prediction, but this can be further improved by pruning hidden nodes and analyzing each part of them. Additionally, different performance metrics should be used to check the accuracy of the most common weather conditions.

[3] The current framework is used as a software application on various platforms. First of all, an ASP.NET framework-based mobile application written in C# was created. The software uses the HEC-HMS algorithm to predict flood events based on data stored in the MySQL database. However, this is limited to the Android operating system only. Additionally, a more in-depth analysis of flood events was created via the web platform. The proposed system was developed by estimating precipitation using data obtained from TMD big data using atmospheric and groundwater models as well as GLOFAS and ML models to predict floods in specific areas. The analysis is based on atmospheric, oceanic, geospatial and population data.

[5] The findings show that all three machine learning systems perform well in flood prediction, with ANN outperforming SVM and RF. The authors also integrated machine learning models into an early warning website that provides flood forecasts and warning times to citizens and city officials.

[7] The results show that both ANN and DT are effective in flood prediction and ANN is better than DT. The authors also compared the performance of machine learning models with traditional statistical models and found that machine learning models outperformed statistical models.

#### B. Proposed system:



The architectural illustration presents a sophisticated machine learning system intricately linked with a user interface, meticulously designed to forecast vital parameters such as river discharge flow and rainfall patterns. This system is intelligently divided into two key domains: Training and Prediction, which collectively form the robust backbone of the Machine Learning Architecture. Complementing these functional cores is the User

Interface, strategically positioned to facilitate seamless interaction between users and the underlying predictive capabilities of the system.

##### 1. User Interface:

- Technological Infrastructure: Built on advanced technologies like Flask and Jinja, chosen for their effectiveness in web development within the Python ecosystem, the User Interface embodies contemporary web design principles to ensure optimal user engagement and satisfaction.

- Data Interaction and Visualization: Central to its role is seamless integration with the underlying machine learning processes, enabling users to effortlessly engage with historical datasets of river discharge flow and rainfall patterns. Through intuitive visualization tools and interactive features, users gain valuable insights into past trends and patterns, empowering informed decision-making and strategic planning.

## 2. Machine Learning Architecture:

### - Training Phase:

- Data Acquisition and Integration: The Training Phase begins with the crucial task of data ingestion, where diverse datasets containing historical records of river discharge flow and rainfall patterns are meticulously collected and consolidated from various sources. These sources may include meteorological agencies, monitoring stations, satellite imagery, and other relevant repositories.

- Data Pre-processing and Refinement: Following data acquisition, a meticulous process of data preparation ensues, involving thorough cleaning, normalization, and segmentation techniques. This preparatory phase is critical in ensuring the integrity and quality of the data, laying the groundwork for subsequent model training endeavours.

- Model Training and Algorithm Selection: Utilizing advanced machine learning algorithms such as Linear Discriminant Analysis (LDA) and Prophet from the `fbprophet` library, the system embarks on model training and refinement. LDA excels at identifying linear combinations of features for optimal class separation, complementing Prophet's proficiency in time-series forecasting, effectively handling seasonal variations and anomalies.

### - Prediction Phase:

- Model Deployment and Real-Time Inference: Equipped with a range of trained models, including those sculpted by LDA and Prophet, the Prediction Phase orchestrates the seamless deployment of these models within a dynamic serving environment. Here, models are ready to engage in real-time inference, processing incoming data streams with precision and efficiency.

Data Input and Feedback Loop: As the prediction engine springs into action, fueled by current river discharge and rainfall data, the system initiates a continuous feedback loop, channeling predictions and relevant data back to the User Interface. This iterative exchange enriches user experiences, fostering deeper insights and informed decision-making in disaster management and mitigation efforts..

## MACHINE LEARNING MODEL

We have used Machine Learning model for predicting the value of flood: -

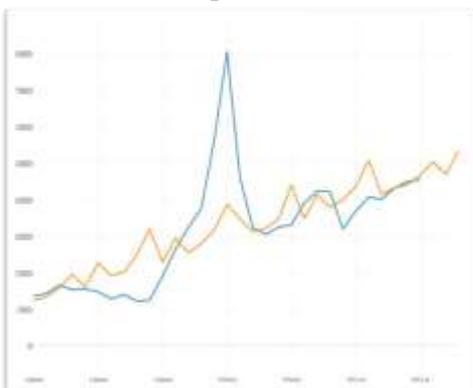
- LINEAR DISCRIMINANT ANALYSIS (LDA)

For predicting following values, we have used these Models: -

- Predicting Future -> LINEAR DISCRIMINANT ANALYSIS (LDA) AND FBPROPHET TOOL (For time series)

- Analysis of past -> LINEAR DISCRIMINANT ANALYSIS (LDA)

To develop a system for predicting future floods and analyzing past flood occurrences for rivers such as Godavari, Krishna, Mahanadi, Cauvery, and Son, we employ Flask to facilitate user input and model interaction. Users provide the intended date and the river name for prediction.



Our approach involves training a machine learning model on existing river datasets, which include parameters like discharge and flood runoff to determine flood occurrences. The model selection includes Linear Discriminant Analysis (LDA), a method widely used in statistics and machine learning for feature characterization and classification.

Linear Discriminant Analysis (LDA) is a technique akin to Fisher's linear discriminant, utilized in pattern recognition and dimensionality reduction. It identifies linear combinations of

features that distinguish between different classes of events or objects. In our context, LDA aids in predicting future flood values and analyzing past flood occurrences. For predicting future flood values, we employ both LDA and the FBProphet tool, leveraging time series analysis. This combination enhances the accuracy of our predictions over time.

#### IV. CONCLUSION

Floods in India caused a fifth of the world's deaths, making India the world's hardest hit after Bangladesh. The rainy season in India is generally from June to September and accounts for approximately 75% of India's annual rainfall. Many studies have been done using machine learning algorithms such as ANN for flood prediction. Most systems use electronic inputs with a hidden layer to predict floods and include parameters such as precipitation, temperature, water flow, water level and humidity. A system has been developed using deep neural networks that takes water flow into account in flood prediction. The challenge with all these systems is that, most of the time, using normal power supply, with the advent of deep neural networks we can more accurately predict the probability of flooding based on rainfall and temperature. This research uses deep neural networks to predict flood events based on temperature and precipitation. Based on possible flood forecasts, evacuation warnings can be issued to save lives and property. The database contains a lot of observed and stored rainfall data, such as minimum and maximum temperature and flood occurrence. Comparing the accuracy achieved by four different algorithms, the results show that deep neural networks can be used effectively for flood prediction. The study here clearly shows that DNN provides better accuracy with 89.71% accuracy compared to other algorithms such as SVM, KNN, and Naive Bayes. A limitation of this research project is that we only have data from 1990 to 2002 to validate the machine learning model for flood prediction. The latest information is not available, which would make it more accurate than possible, but this is not a big problem. Many aspects of future work can be addressed from this research. First, the unevenness of data distribution is not taken into account. According to the prediction, further studies can be carried out primarily on neural networks, which can achieve greater accuracy by taking into account the latest information. Secondly, other factors, including floods, can be taken into account to estimate flood risk in various cities/states of India. Finally, by combining machine learning models, the results become open to further investigation and the development of new methods.

#### V. FUTURE ENHANCEMENT

Machine learning algorithms such as ANN, KNN and SVM have been successfully used in flood prediction. KNN also works on a similar principle. That is, identify the "k" information points closest to the new information and classify the closest information as the most common class. Artificial neural networks are complex systems that can model relationships between inputs and outputs, learn from data, and improve accuracy over time. SVM is a binary classification algorithm that divides data into different classes based on hyperplanes to ensure separation of classes. All of these algorithms can learn historical data such as rainfall, water flow, humidity, and other factors in the environment to accurately predict flood occurrence. These algorithms can be used to generate accurate flood forecasts and provide timely warnings to vulnerable communities so they can take appropriate precautions and reduce flooding. To predict future floods, we will use the ML model developed by combining all three of KNN, ANN and SVM to predict floods.

#### REFERENCES

1. Adnan, Mohammed Sarfaraz Gani, et al. "A novel framework for addressing uncertainties in machine learning-based geospatial approaches for flood prediction." *Journal of environmental management* 326 (2023): 116813.
2. Sankaranarayanan, Suresh, et al. "Flood prediction based on weather parameters using deep learning." *Journal of Water and Climate Change* 11.4 (2020): 1766-1783.
3. Puttinaovarat, Supattra, and Paramate Horkaew. "Flood forecasting system based on integrated big and crowd source data by using machine learning techniques." *IEEE Access* 8 (2020): 5885-5905.
4. Silvestro, F., et al. "Impact-based flash-flood forecasting system: Sensitivity to high resolution numerical weather prediction systems and soil moisture." *Journal of Hydrology* 572 (2019): 388-402.

5. Riza, Hammam, et al. "Advancing Flood Disaster Mitigation in Indonesia Using Machine Learning Methods." 2020 International Conference on ICT for Smart Society (ICISS). IEEE, 2020.

6. Mosavi, Amir, Pinar Ozturk, and Kwok-wing Chau. "Flood prediction using machine learning models: Literature review." *Water* 10.11 (2018): 1536.

7. Kashyap, K., A. Karthik, and G. S. ShivaKumar. "Flood prediction using machine learning." *Int Res J Eng Technol (IRJET)* 8.7 (2021): 4.

