FLOOD PATH PREDICTOR

GROUP MEMBERS

- **Sanou Lionel Ange Daniel**
- **Fawzeia Nasr Hussein Abdelrahman**
- **Widad Amir Abdalkhalig Abbas**

Literature review

1-INTRODUCTION

Floods are among the most destructive natural disasters, causing widespread damage to communities and infrastructure worldwide. With the increasing frequency and intensity of extreme weather events due to climate change, the need for effective flood management has become more critical. Predicting how floodwaters will move is essential for reducing their impact, protecting lives, and minimizing economic losses.
Our project focuses on flood prediction because it plays a vital role in disaster preparedness and response. Accurate predictions help identify at-risk areas, enabling authorities to issue timely warnings, organize evacuations, and allocate resources effectively. In the long term, understanding flood paths also supports urban planning and the development of more resilient communities.
As a group, we believe that reviewing the existing literature is a necessary step. It permits us to understand current approaches, explore data and tools, identify research gaps, and improve our project. This literature review is crucial to ensure that our project is well-informed and capable of making a meaningful contribution to flood risk management, ultimately helping communities better prepare for and respond to future flood events.

2-Orginization

By arranging the papers in the order they were published, we can trace the evolution of flood prediction methods over time, illustrating how research has developed:

A-Early methods

We first start the literature review with an article of (Mohamed Fofana, 2023) about hydrological models in order to get an idea of the early methods of flood prediction .

B-Recent methods

With the rise of machine learning, studies from **2018 onwards** began integrating **ML algorithms** such as Artificial Neural Networks (ANNs) and Support Vector Machines (SVMs).As our project relies on machine learning ,it was necessary to read recent papers . Thus, we continued our literature review with the articles of (A Fares Hamad Aljohani, 2023)and (Léonard Boussioux, 2022).

3- Summary and Synthesis

The paper ***"Flood Forecasting and Warning System: A Survey of Models and Their Applications in West Africa"*** by Mohamed Fofana et al. provides an overview of flood forecasting models and their practical applications in West African contexts. The study reviews traditional hydrological models, hybrid approaches, and modern technologies utilized for flood prediction and warning systems.

Key Findings: The article reviews various hydrological models and their applications in flood forecasting and warning systems in West Africa. It highlights the strengths and weaknesses of different models and emphasizes the importance of non-structural methods for flood prediction.

Methodology: The methodology employed involves conducting a systematic review of existing literature and models, with a particular emphasis on non-structural methods such as hydrological rainfall-runoff models. This approach involves rigorously searching and analyzing peer-reviewed articles, technical reports, and conference proceedings to identify relevant studies. The focus on non-structural methods allows for an in-depth evaluation of models that predict flood events based on rainfall and runoff data, without relying on physical infrastructure. The contribution of this work is significant, as it provides a comprehensive overview of the current state of flood prediction methodologies, highlighting effective practices and identifying critical gaps in the research. By pinpointing these deficiencies, the study suggests areas where future research and development can improve, thereby enhancing the accuracy and reliability of flood forecasting and warning systems.

The paper titled ***"Hurricane Forecasting: A Novel Multimodal Machine Learning Framework"*** presents an innovative machine learning model, *Hurricast*, designed for 24-hour hurricane forecasting.

Key Findings: The study introduces *Hurricast*, a novel multimodal machine learning model that integrates 3D reanalysis maps, numerical storm data, and categorical inputs to predict hurricane intensity and displacement. The findings show that *Hurricast* outperforms traditional methods in forecasting 24-hour hurricane metrics.

Methodology: The framework employs a combination of gradient-boosted decision trees (XGBoost) and multimodal learning. It incorporates storm-specific engineered features and leverages historical data to improve predictive accuracy.

Contribution to the Field: This research advances hurricane forecasting by integrating multiple data modalities, enhancing accuracy, and addressing limitations in traditional forecasting methods.

The paper ***"Flood Prediction using Hydrologic and ML-based Modeling: A Systematic Review*** **by Vishal H. Goyal et al** "systematically reviews recent advancements in flood prediction methods, focusing on the integration of machine learning (ML) with traditional hydrological models. It highlights the strengths, limitations, and applications of these approaches in various regions.

Key Findings: The review underscores the growing shift toward hybrid models, which combine ML and physical simulations to improve flood forecasting accuracy, particularly in areas with sparse data. Challenges like model generalization and performance evaluation are also highlighted.

Methodology: The authors analyze a wide range of studies, comparing their methodologies, geographic applications, and outcomes to identify trends and gaps in the field.

Contribution to the Field: By emphasizing the advantages of hybrid approaches, this paper provides a roadmap for enhancing flood prediction systems, offering valuable insights for future research and practical implementation.

4- Conclusion

The literature reveals a progression from traditional hydrological models to machine learning and hybrid approaches, significantly improving flood prediction accuracy and adaptability. Traditional methods laid the groundwork but struggled with real-time accuracy, while ML and hybrid models now enhance forecasting by integrating diverse data sources and techniques.
Our research builds on these advancements, aiming to refine flood prediction systems further. By addressing existing limitations, it seeks to provide innovative tools for flood risk management, improving community resilience and disaster preparedness.

DATA RESEARCH

1-INTRODUCTION

Floods are complex natural disasters influenced by various climatic, geographic, and environmental factors. To develop an effective **Flood Predictor**, a thorough exploration of the available dataset is essential to understand the relationships between variables like precipitation, temperature, and wind speed and their connection to flood occurrences. This research identifies patterns, addresses data quality issues, and highlights gaps, which are critical for improving prediction accuracy. By thoroughly analyzing the dataset, we lay the foundation for building a robust machine learning model that can aid in proactive flood management and better community preparedness.

2-ORGANIZATION

In order to collect data for our capstone project ,we use the dataset from Indian flood inventory dataset that we combine with the historical data of open-meteo.
firstly,we use the dataset of Indian flood inventory which contains geographical data of flooded areas and the period where events happened .
In the second part,we use the API of open-meteo to have precise hydrological data about area flooded in addition with meteorological data

3-DATA DESCRIPTION

The dataset contains 371 entries with the following columns:

Start Date & End Date: Start and end dates of the analyzed events.

Latitude & Longitude: Geographical coordinates of the studied locations.

temperature_2m_mean: Average temperature at 2 meters during the period.

precipitation_sum: Total precipitation (in mm) over the period.

rain_sum: Total rainfall recorded (in mm).

wind_speed_10m_max: Maximum wind speed at 10 meters (in m/s).

Flooded: Binary indicator (1 = flood occurred, 0 = no flood).

The dataset is a combination of data from Indian flood inventory which provided us with historical data on the location of floods as well as data from the api of open meteo which provided us with historical data on hydrological and meteorological factors .The dataset is on a CSV format, contains 371 rows and 9 columns. We selected our dataset for the following reasons :

A-Key Climatic Parameters:

Precipitation and **Rainfall**: Floods are strongly correlated with the amount of rainfall received over a given period.
Wind Speed: Can influence evaporation or exacerbate conditions during severe storms.

Temperature: Impacts evaporation and soil saturation, which are crucial for hydrological analysis.

B-Geographical Coordinates:

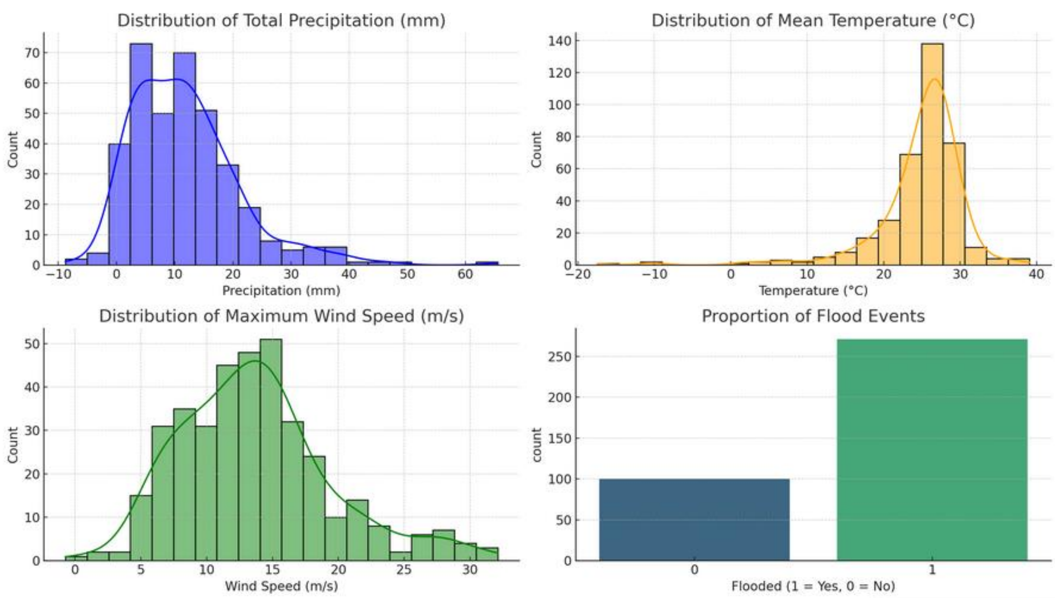
These variables help localize events and identify flood-prone areas.

C-Flood Label:

Provides a target variable for training a supervised machine learning model to predict floods.

These data points directly support the project's goal of modeling flood behavior. They enable the linking of climatic conditions to flood occurrences, forming the basis for predicting future flood paths.

4-**Data Analysis and Insights**



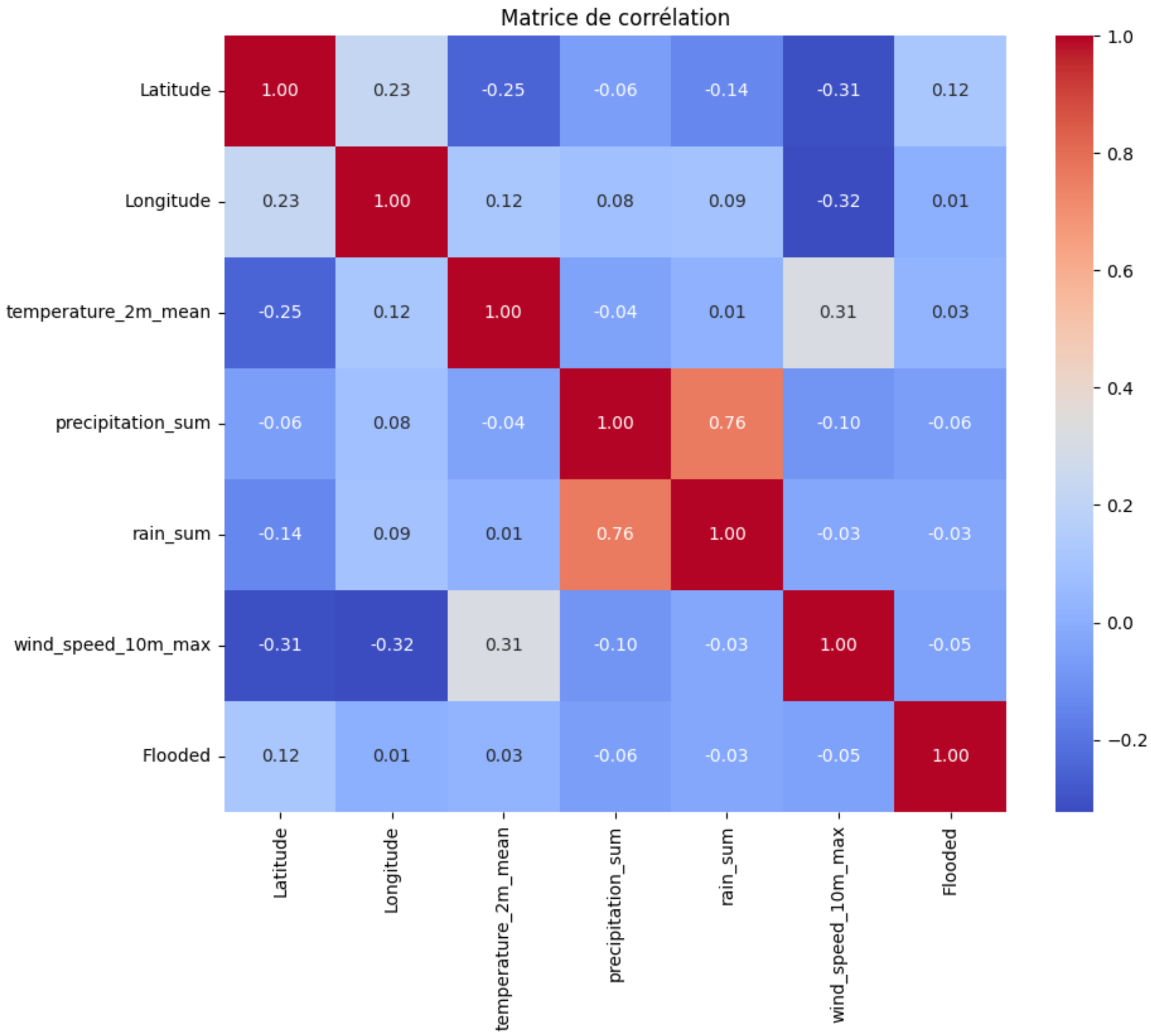
VISUAL INSIGHTS

1. **Distributions**:

    - Precipitation and rainfall show positively skewed distributions, indicating frequent low-precipitation events with occasional extreme values.
    - Wind speed has a roughly normal distribution centered around 13 m/s.
    - Temperature mostly falls between 20–30°C, with few extreme outliers.

2. **Flood Occurrence**:

    A significant portion of the dataset corresponds to flood events (Flooded = 1), which may influence the balance of the machine learning model and require addressing (re-sampling or weighting).



Correlation of features with flood occurrence (Flooded):

- **Latitude** shows the strongest positive correlation (+0.12), indicating that certain latitudinal regions are more prone to flooding.

- **Precipitation**, **Rainfall**, and **Wind Speed** show weak or slightly negative correlations, suggesting that the current feature representation might not fully capture flood dynamics.

Also The average precipitation is 11.75 mm, with a range from -8.77 mm to 65.52 mm, indicating possible data entry issues due to negative values.

5-CONCLUSION

The analysis identifies key insights, noting that certain latitudinal regions are more prone to flooding due to geographical and climatic factors, while emphasizing the importance of correcting negative precipitation values to avoid model distortion. With 73% of instances labeled as floods, there is a risk of overfitting, necessitating resampling or class weight adjustments during training. The presence of outliers in temperature, precipitation, and wind speed, indicative of extreme events, underscores the need for their proper incorporation in flood prediction models. This research is crucial for developing a robust machine learning model by highlighting data gaps, such as missing terrain and soil moisture data, and enhancing predictive accuracy

TECHNOLOGIE REVIEW

1-INTRODUCTION

In recent years, flooding has emerged as a significant global challenge, causing widespread devastations and economic loss. To mitigate the impact of these natural disasters, accurate and timely flood prediction is crucial. This project aims to develop a Floods Path Predictor, a tool that leverages advanced machine learning techniques to forecast potential flood zones. By analyzing historical weather data, river flow information, and topographical features, this tool will provide valuable insights to communities and authorities, enabling proactive measures to safeguard lives and property. This project aligns with several United Nations Sustainable Development Goals (SDGs), including Climate Action, Sustainable Cities and Communities, No Poverty, and Good Health and Well-being.

2-TECHNOLOGY OVERVIEW

Machine Learning (ML) and Deep Learning (DL) are AI subsets that enable systems to learn from data and make decisions without explicit programming. Key ML features include supervised learning for predictions, unsupervised learning for pattern discovery, and reinforcement learning for decision-making. DL features include neural networks inspired by the human brain, CNNs for image analysis, RNNs for sequential data, and GANs for data generation. ML use cases span healthcare (disease diagnosis), finance (fraud detection), and marketing (customer segmentation), while DL excels in computer vision (object detection), NLP (text generation), and autonomous vehicles (self-driving cars). In flood prediction, ML can predict water levels and classify risk areas, while DL, via CNNs and RNNs, can analyze satellite imagery and model time-series data for forecasting. The choice between ML and DL depends on data quality, problem complexity, computational resources, and time constraints. A hybrid approach, using ML for data preprocessing and DL for predictions, can effectively leverage both techniques to enhance flood prediction accuracy.

3-RELEVANCE

Machine Learning (ML) and Deep Learning (DL) are crucial for the Flood Path Predictor project, enabling complex data analysis and accurate predictions. ML techniques enhance model performance through feature engineering and DL, particularly Convolutional Neural Networks (CNNs), can identify flood-prone areas from satellite imagery. Regression and classification models in ML predict flood water levels and risk areas, while Recurrent Neural Networks (RNNs) forecast future flood events using time-series data. ML and DL can process real-time data for timely alerts and early warning systems, improving decision-making and resource allocation. By providing data-driven insights, these technologies guide policymakers in flood mitigation strategies, ensuring effective resource distribution to high-risk areas.

4-Comparison and Evaluation

For the Flood Path Predictor project, Machine Learning (ML) techniques such as Random Forest, Gradient Boosting Machines (GBM), and Support Vector Machines (SVM) offer distinct advantages and challenges. Random Forest is highly effective with large datasets and provides feature importance, making it suitable for diverse data types, but it can be computationally expensive and complex to interpret. GBM offers high predictive accuracy and handles missing values well, though it requires careful hyperparameter tuning to avoid overfitting and is computationally intensive. SVM is excellent in high-dimensional spaces and effective for both linear and nonlinear data, but it demands significant preprocessing and is sensitive to feature scaling. Python, with its extensive libraries like Scikit-learn, offers versatility and robust performance, making it the preferred tool. While Random Forest and GBM are powerful for achieving high accuracy, they require substantial computational resources. SVM provides precision in classification but at a higher computational cost. Considering factors like cost, ease of use, scalability, and performance, a hybrid approach that leverages the strengths of these ML algorithms can effectively enhance flood prediction accuracy.

5-USE CASES AND EXAMPLES

The application of machine learning (ML) techniques to predict natural disasters, particularly floods, has gained significant traction in recent years. For instance, the National Oceanic and Atmospheric Administration (NOAA) employs advanced ML models to predict flood risks in the United States, analyzing historical weather data, river flow data, and topographic information to generate flood forecasts. Similarly, the European Union's Copernicus Emergency Management Service (EMS) uses satellite imagery and ML to monitor and predict flood events across Europe, providing early warnings by analyzing data from various sensors. In urban flood modeling, the Los Angeles County Flood Control District (LACFCD) utilizes hydrodynamic models coupled with ML to simulate urban flooding, considering factors like rainfall intensity, drainage infrastructure, and land use. The UK Centre for Ecology & Hydrology (UKCEH) employs a combination of hydrological models and ML to forecast river flow and flood risk in the UK, incorporating real-time data on rainfall, soil moisture, and river levels. Additionally, NASA's Flood Mapping System uses satellite imagery and ML to map global flood extent, identifying flooded areas and assessing the severity of flooding through radar data analysis.

6- Identify Gaps and Research Opportunities

While machine learning (ML) has shown significant promise in flood prediction, several challenges and areas for improvement remain. Ensuring access to high-quality, real-time data from diverse sources, such as meteorological stations, satellite imagery, and hydrological models, can be challenging, particularly in regions with limited infrastructure. Inconsistent data formats and quality also hinder model performance, highlighting the need for standardized data formats and quality control procedures. Deep learning models, despite their power, are often complex and difficult to interpret, limiting their transparency and trust in decision-making processes. Ensuring that models generalize well to different regions and climate conditions is crucial, necessitating the development of models that can adapt to varying environmental factors. Efficiently incorporating real-time data into models is essential for accurate and timely predictions, and quantifying the uncertainty associated with predictions helps decision-makers assess potential risks. Ethical considerations, such as mitigating data biases and ensuring equitable social impact, are critical. Future research directions include developing hybrid models that combine physical-based and ML approaches, ensembling multiple models for enhanced accuracy, making AI models more interpretable, integrating data from diverse sources to improve prediction accuracy, and designing user-friendly interfaces for effective dissemination of flood information. Addressing these gaps will lead to more accurate, reliable, and impactful flood prediction systems.

7-CONCLUSION

In conclusion, the effective implementation of machine learning techniques, particularly those from the supervised learning family, holds immense potential for developing robust and accurate flood prediction models. By leveraging advanced algorithms like Random Forest, Gradient Boosting Machines, and Support Vector Machines, we can analyze historical data to identify patterns and trends that can inform future predictions.

The choice of Python as the primary programming language, coupled with powerful libraries like NumPy, Pandas, and Scikit-learn, provides a versatile and efficient framework for data analysis, model development, and deployment.

However, it is essential to acknowledge the limitations and challenges associated with these technologies. Data quality, model interpretability, and real-time prediction capabilities are critical areas that require ongoing research and development. By addressing these challenges and exploring innovative approaches, we can further enhance the accuracy and reliability of flood prediction systems