<u>Bansilal Ramnath Agarwal Charitable Trust's</u>

<u>Vishwakarma Institute of Information Technology</u>

(Department of Electronics & Telecommunication)

Group No.:- 07

Project Report

on

"**Flood Prediction using Machine Learing**"

(SPONSORED BY: In-House)

Domain : Machine Learning

Submitted By:					
Roll No.	**GR No.**	**Division**	**Name of Student**	**E-mail**	**Contact Number**
411021	22110721	A	Abhijeet Kadam	abhijeet.22110721@viit.ac.in	8080053515
411024	22110260	A	Sandesh Kangude	sandesh.22110260@viit.ac.in	8668803417
411036	22111313	A	Ketu Patel	ketu.22111313@viit.ac.in	7972128378
411037	22110016	A	Aditya Patil	aditya.22110016@viit.ac.in	9022162512

B.Tech

Of

Vishwakarma Institute of Information Technology, Pune
(An Autonomous Institute affiliated to Savitribai Phule Pune University, Pune)

Under supervision of

Dr. Jayashree Tamkhade

Year 2024 – 2025

CERTIFICATE

This is to certify that project work entitled "**Flood Prediction using Machine Learning”** carried out in the seventh semester by,

Roll No.	GR No.	Division	Name of Student
411021	22110721	A	Abhijeet Kadam
411024	22110260	A	Sandesh Kangude
411036	22111313	A	Ketu Patel
411037	22110016	A	Aditya Patil

in partial fulfillment for the award of **B.Tech** degree in Electronics and Telecommunication Engineering from Vishwakarma Institute of Information Technology, Pune (An Autonomous Institute affiliated to Savitribai Phule Pune University, Pune) during the academic year 2024-2025.

Date:

Dr. Jayashree Tamkhade — Guide

Prof. A. P. Navghane — Project Coordinator

Dr. S. K. Habbu
H.O.D., E&TC Engg.

ABSTRACT

Flood prediction is a critical component of disaster management, playing an essential role in enabling timely and proactive measures that significantly reduce risks and minimize damages associated with flooding events. Floods are among the most frequent and devastating natural disasters, often resulting in substantial loss of life, extensive property damage, and disruption to livelihoods. Given the severe impacts of floods, developing accurate and reliable flood prediction models is imperative. Such models support decision-makers and stakeholders in implementing effective mitigation strategies, ultimately saving lives and protecting communities.

In light of these challenges, this study presents a hybrid machine learning framework designed to enhance flood prediction accuracy by integrating image-based analysis with statistical modeling. This innovative approach provides a comprehensive solution by leveraging both spatial data insights and meteorological factors. The integration of these methodologies allows for a more nuanced understanding of flood risks, as it combines visual data from satellite imagery or aerial photographs with quantitative data derived from weather patterns and hydrological assessments.

The framework's image-based analysis component focuses on interpreting environmental factors that contribute to flooding, such as land use changes, vegetation cover, and rainfall intensity. By analyzing images, the model can classify areas based on their susceptibility to flooding and assign probability values accordingly. This visual assessment is particularly valuable in identifying critical areas at risk, providing a layer of insight that traditional methods may overlook.

On the other hand, the statistical modeling aspect of the framework utilizes a range of meteorological data—such as rainfall amounts, temperature variations, humidity levels, river discharge rates, and topographical information—to predict flood occurrences. This component employs advanced machine learning algorithms to analyze historical data and identify patterns that correlate with flood events. By combining these two approaches, the hybrid model not only enhances prediction accuracy but also improves the overall understanding of flood dynamics.

ACKNOWLEDGEMENT

This work could not have been completed without the guidance and encouragement of many people. We would like to particularly acknowledge those below.
We pay our humble regards and gratitude to Prof. Jayshree Tamkhede for guiding us and giving moral support and timely boost.

We wish to express our special thanks to Dr**.** Jayashree Tamkhade and Prof. Dipti Pandit, project evaluator , who helped us a lot in the preparation of our seminar topic.

THANKING YOU

ABHIJEET KADAM

SANDESH KANGUDE

KETU PATEL

ADITYA PATIL

LIST OF FIGURES

LIST OF TABLE:

INDEX

1. INTRODUCTION

Floods represent one of the most catastrophic natural disasters, leading to substantial loss of life, property damage, and infrastructure destruction globally. Their increasing frequency and intensity can be attributed to several factors, including climate change, urbanization, and environmental degradation. This escalating trend underscores the urgent need for effective flood prediction models that can adapt to changing conditions and diverse datasets. Traditional methods of flood forecasting often rely on meteorological data or hydrological simulations; however, these approaches may lack the flexibility necessary to address the complexities of modern flood events.

To enhance flood prediction accuracy, this study introduces a novel hybrid approach that integrates machine learning (ML) techniques with traditional forecasting methods. By leveraging data-driven insights, the proposed model aims to improve both accuracy and scalability in flood risk assessments. The hybrid model consists of two primary components: **image classification** and **statistical modeling**. The image-based classification approach analyzes critical environmental factors such as land cover types (e.g., deforestation, paddy fields) and rainfall intensity levels (e.g., light, moderate, heavy). By assigning flood probability values based on these classifications, this method effectively incorporates visual data into flood prediction frameworks.

The second component focuses on statistical modeling that utilizes geospatial and meteorological data—such as rainfall amounts, temperature fluctuations, humidity levels, river discharge rates, and elevation profiles—to forecast flood occurrences. Advanced machine learning algorithms are employed to identify patterns and correlations within this data, resulting in reliable assessments of flood risk. The integration of these two approaches is achieved through a weighted probability mechanism that prioritizes the statistical model due to its comprehensive representation of data while also incorporating insights from image-based classifications. This holistic framework captures both visual and quantitative aspects of flood risk, providing a more robust prediction model.

In summary, as climate change continues to exacerbate the frequency and intensity of flooding events worldwide, innovative solutions such as the proposed hybrid prediction model are essential for improving our ability to manage flood risks effectively. By combining traditional forecasting methods with advanced machine learning techniques, this study not only addresses current challenges in flood prediction but also sets a foundation for future research in disaster risk management.

2. LITERATURE SURVEY

[1]. D. Han, L. Chan and N. Zhu (2007)

The paper addresses the challenges of applying SVM in flood forecasting, particularly issues related to over-fitting and under-fitting. It emphasizes the difficulty in selecting optimal input combinations and parameters, which is crucial for model performance. The study aims to evaluate the effectiveness of SVM in comparison to traditional benchmarking models, such as Transfer Function, Trend, and Naive models, in predicting river flows based on rainfall data. The study underscores the importance of understanding the model's behavior in response to different types of rainfall inputs and the necessity of fine-tuning parameters to enhance predictive accuracy. The findings suggest that further exploration is needed to fully harness the potential of SVM in hydrological modeling, particularly in addressing the complexities of rainfall-runoff processes.

[2]. Jeerana NOYMANEE, Thanaruk THEERAMUNKONG (2017)

The research addresses employing machine learning techniques alongside traditional hydrological models can substantially improve flood forecasting accuracy in urban areas like Sukhothai, Thailand. The findings highlight the potential for developing more effective early warning systems that can better inform flood management strategies, ultimately leading to reduced impacts from urban flooding. Further research is encouraged to refine these models and explore additional machine learning approaches for enhanced predictive capabilities..

[3]. Dinh Ty Nguyen and Shien-Tsung Chen (2020):

The research addresses the need for improved flood forecasting methods, specifically focusing on the limitations of deterministic forecasting, which can create a false sense of certainty. The study proposes a real-time probabilistic flood forecasting model that combines various machine learning techniques to provide more reliable predictions of flood stages, particularly for flash floods in the Yilan River, Taiwan. This approach provides valuable information for flood management and mitigation strategies, particularly in real-time scenarios. The findings suggest that probabilistic forecasting is a practical improvement over traditional methods, allowing for better preparation and response to flood events. Further research is recommended to refine these methods and explore their applicability in other regions prone to flooding.

[4] Guangyuan Kan, Ke Liang, Haijun Yu, Bowen Sun, Liuqian Ding, Jiren Li, Xiaoyan He, and Chengji Shen(2020):

The paper aims to develop a novel hybrid machine learning (HML) hydrological model for flood forecasting purposes. Optimizing network topology and parameters simultaneously to improve performance. Balancing network complexity and training accuracy to achieve good results in both training and testing. Overcoming the local minimum issue in traditional neural network training. The HML hydrological model, with its novel structure and training approach, demonstrates excellent potential for flood forecasting applications. The results highlight the possibility of further applying this model to real-world flood problems.

[5]. Mehrsa Bayat, Omid Tavakkoli (2022) :

The paper discusses the challenges in flood forecasting and how traditional hydrological models may not always provide accurate results due to the complexities of weather patterns. The problem is to find better methods for predicting floods by integrating machine learning techniques with traditional models. The paper emphasizes that machine learning techniques offer a promising alternative to traditional flood forecasting models. By integrating data-driven insights with existing models, the accuracy and timeliness of flood predictions can be enhanced, potentially leading to better disaster preparedness and risk management strategies.

[6]. Prof. Priyanka Pujari, Nidhi Kulkarni, Vinay C Hiremath, Vinay Bhushi(2024):

The paper states floods are a frequent natural disaster in India, causing significant loss of life and property. The accurate prediction of flood onset and progression is crucial for minimizing these impacts. The research aims to enhance flood forecasting accuracy using different machine learning models, including K-nearest neighbor (KNN), support vector classification (SVC), decision tree classification, binary logistic regression, and stacked generalization (stacking) to address the challenges of real-time flood prediction and inform policy proposals for disaster management.

3. OBJECTIVES

1. **Integrate image-based classification and statistical modeling to enhance the accuracy and reliability of flood prediction**: The integration of image-based classification with statistical modeling is designed to create a robust flood prediction framework. By combining these two methodologies, the project aims to utilize both visual data from images (such as satellite or aerial photographs) and quantitative data from statistical models. This dual approach allows for a more nuanced understanding of flood risks, capturing various environmental factors that contribute to flooding. The image classification can identify critical land cover types and conditions, while statistical models can analyze historical meteorological data, leading to improved predictive accuracy and reliability in forecasting flood events.
2. **Utilize pre-trained models to classify user-input images into categories such as deforestation, paddy fields, and rainfall intensity (light, moderate, heavy), assigning flood probabilities based on these classifications:** To streamline the classification process, the project will employ pre-trained machine learning models capable of analyzing user-input images effectively. These models will categorize images into specific classes such as deforestation, paddy fields, and varying levels of rainfall intensity. By assigning flood probability values based on these classifications, the framework provides an intuitive method for assessing flood risk visually. This approach not only enhances the model's predictive capabilities but also facilitates user engagement by allowing stakeholders to contribute relevant data through image input.
3. **Leverage statistical data inputs, including geospatial and meteorological parameters like rainfall, temperature, humidity, and river discharge, to predict flood risks comprehensively:** In addition to image-based analysis, leveraging statistical data inputs is crucial for developing a comprehensive flood prediction model. The project will incorporate various geospatial and meteorological parameters—such as rainfall amounts, temperature variations, humidity levels, and river discharge rates—to create a detailed risk assessment framework. By utilizing advanced machine learning algorithms to analyze these parameters, the model can identify patterns and correlations that are indicative of potential flooding events. This comprehensive data-driven approach ensures that all relevant factors are considered in predicting flood risks.

4. MOTVATION

Floods rank among the most catastrophic natural disasters, impacting millions of people worldwide and causing extensive social, economic, and environmental damage. The increasing frequency and severity of floods, driven by factors such as climate change, rapid urbanization, deforestation, and the mismanagement of natural resources, highlight the urgent need for accurate and proactive flood prediction systems. Traditional flood prediction models often rely solely on statistical or historical data, which limits their effectiveness in incorporating real-time dynamic factors such as changes in land use, rainfall intensity, or evolving environmental conditions. This gap in existing models underscores the necessity for an innovative approach to flood risk assessment that can effectively integrate multiple sources of data.

Our motivation for this project lies in addressing these challenges through the development of a hybrid flood prediction system that combines the strengths of machine learning, image classification, and statistical modeling. This integrated approach enables us to analyze complex factors that contribute to flooding, such as various land cover types—like deforestation, agricultural areas, or construction zones—and rainfall patterns categorized by intensity (e.g., light, moderate, or heavy). By employing image-based classification techniques, we can gain valuable insights into environmental changes that may influence flood risks. Simultaneously, the system will incorporate geospatial and meteorological data—including rainfall amounts, temperature fluctuations, humidity levels, river discharge rates, and elevation—thereby offering a comprehensive prediction framework that accounts for both visual and quantitative data.

The innovative fusion of these methodologies through a weighted probability mechanism ensures that our system provides not only accurate predictions but also the flexibility to adapt to regions with diverse data availability. By assigning appropriate weights to different data sources—prioritizing those with broader representations—we can enhance the reliability of our flood risk assessments. This approach allows for a more nuanced understanding of flood dynamics and improves our ability to respond to changing environmental conditions.

Our ultimate goal is to mitigate the devastating effects of floods by equipping stakeholders with a reliable, data-driven tool that supports early warning systems, efficient resource allocation, and informed decision-making. By leveraging the power of technology and data science in this manner, we hope to contribute meaningfully to disaster preparedness and resilience efforts.

5. BLOCK DIAGRAM

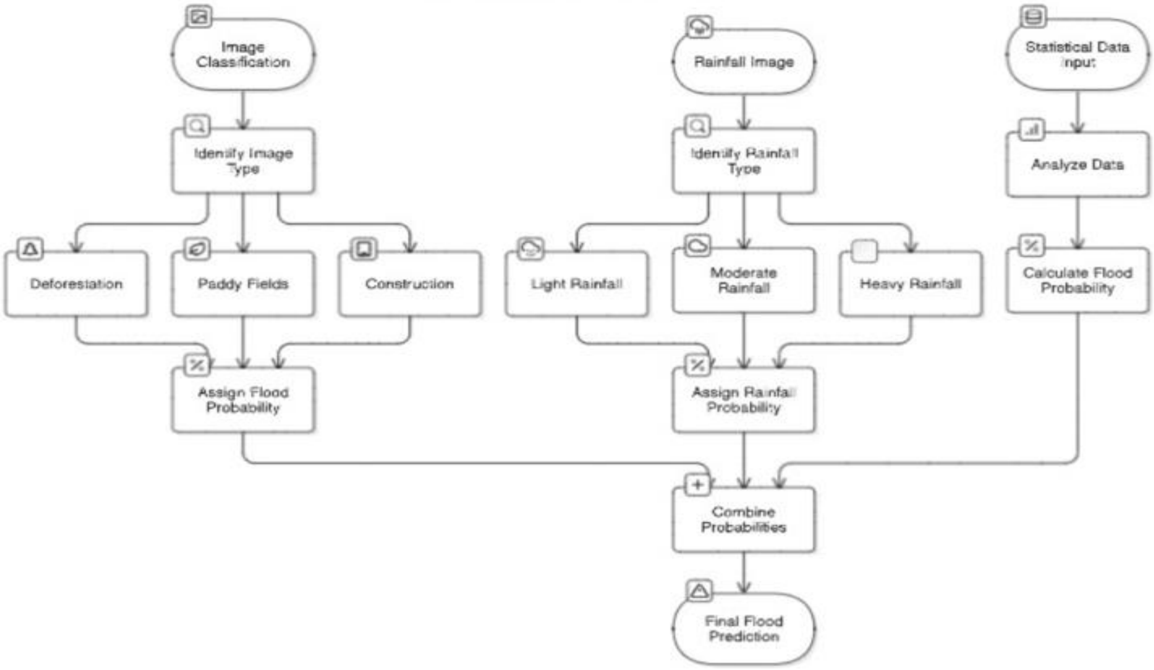

Fig 1. Block Diagram

The block diagram Fig 1. illustrates a comprehensive flood prediction system that effectively combines image classification and statistical modeling to assess flood risks. This hybrid framework is implemented as a Streamlit application, providing a user-friendly interface for real-time predictions. The system operates through two distinct approaches: an image classification model and a statistical modeling approach, which are integrated using a weighted probability mechanism to generate accurate and reliable flood predictions.

The first approach centers on image classification, where users begin by uploading two types of images. The initial image is analyzed to determine the land cover type, which may include categories such as deforestation, paddy fields, or construction zones. The classification model has been trained on numerous folders containing representative images for these categories. When a user uploads an image, the model processes it to identify the specific type of land cover. This classification is critical because different land cover types have varying impacts on flood risks. For example, deforestation tends to increase runoff and decrease water absorption, thereby raising flood probabilities, while agricultural or construction areas may influence flood risks in other ways. Once the land cover type is identified, the system assigns a flood probability value based on that land cover's susceptibility to flooding.

The second image input focuses on rainfall intensity. Users upload an image depicting rainfall, which the system classifies into one of three categories: light, moderate, or heavy rainfall. The classification model, trained on a diverse set of rainfall images, assesses the intensity and assigns a corresponding flood probability value. Rainfall is a crucial factor in flood prediction; heavy rainfall significantly increases the risk of flooding. By integrating the probabilities from both the land cover and rainfall classification models, the system generates a weighted probability score. Each of these two classification models contributes equally (e.g., 0.05 weight each) to the overall prediction, ensuring that insights derived from images are included without overshadowing the broader data representation provided by the statistical model.

The second approach relies on statistical modeling, where users input various numerical data encompassing geospatial and meteorological factors. This data includes variables such as latitude, longitude, rainfall (in mm), temperature (in °C), humidity (%), river discharge (in m^3/s), water level (in m), and elevation (in m). Additional inputs like land cover type, soil type, population density, infrastructure data, and historical flood occurrences are also considered. The statistical model processes this data to identify patterns and relationships among the variables by utilizing historical flood datasets for training purposes. For instance, combinations of high rainfall levels, elevated river discharge rates, and low elevation may lead to higher flood probabilities, while areas characterized by favorable soil absorption or vegetation may exhibit reduced risks. The statistical model assigns a probability value based on the input data and contributes the majority weight (e.g., 90%) to the final flood prediction due to its ability to provide a more comprehensive assessment of flood risks based on diverse quantitative factors.

The final flood prediction integrates probabilities from both the image classification and statistical models using a weighted probability mechanism. This ensures that statistical data—offering a broader and more detailed representation of flood-related factors—has a greater influence on the outcome. However, insights from image-based models are not overlooked; they add qualitative assessments by capturing changes in land use and rainfall patterns that might not be fully reflected in static statistical data alone. The result is a well-rounded flood risk prediction that considers both dynamic and static factors. This hybrid framework offers significant advantages by combining diverse data sources for more accurate and holistic flood risk assessments. The inclusion of image classification allows the model to adapt to real-time environmental changes—such as rapid deforestation or varying rainfall intensities—that may not be captured in static statistical datasets. Additionally, the user-friendly Streamlit interface makes this system accessible to a wide audience, enabling non-technical users to leverage advanced machine learning models for effective flood predictions. The modular design of the system also ensures scalability, allowing for future integration of additional data sources like remote sensing or IoT-based measurements to further enhance prediction accuracy.In conclusion, the flowchart outlines a robust flood prediction framework that effectively integrates image classification with statistical modeling. By combining these two approaches, the system provides a comprehensive solution for assessing flood risk that balances real-time environmental insights with broader statistical analysis. This innovative framework has significant potential to improve disaster preparedness, resource allocation, and decision-making in regions prone to flooding, ultimately contributing to reduced impacts and enhanced resilience against one of nature's most devastating disasters.

6. METHODOLOGY

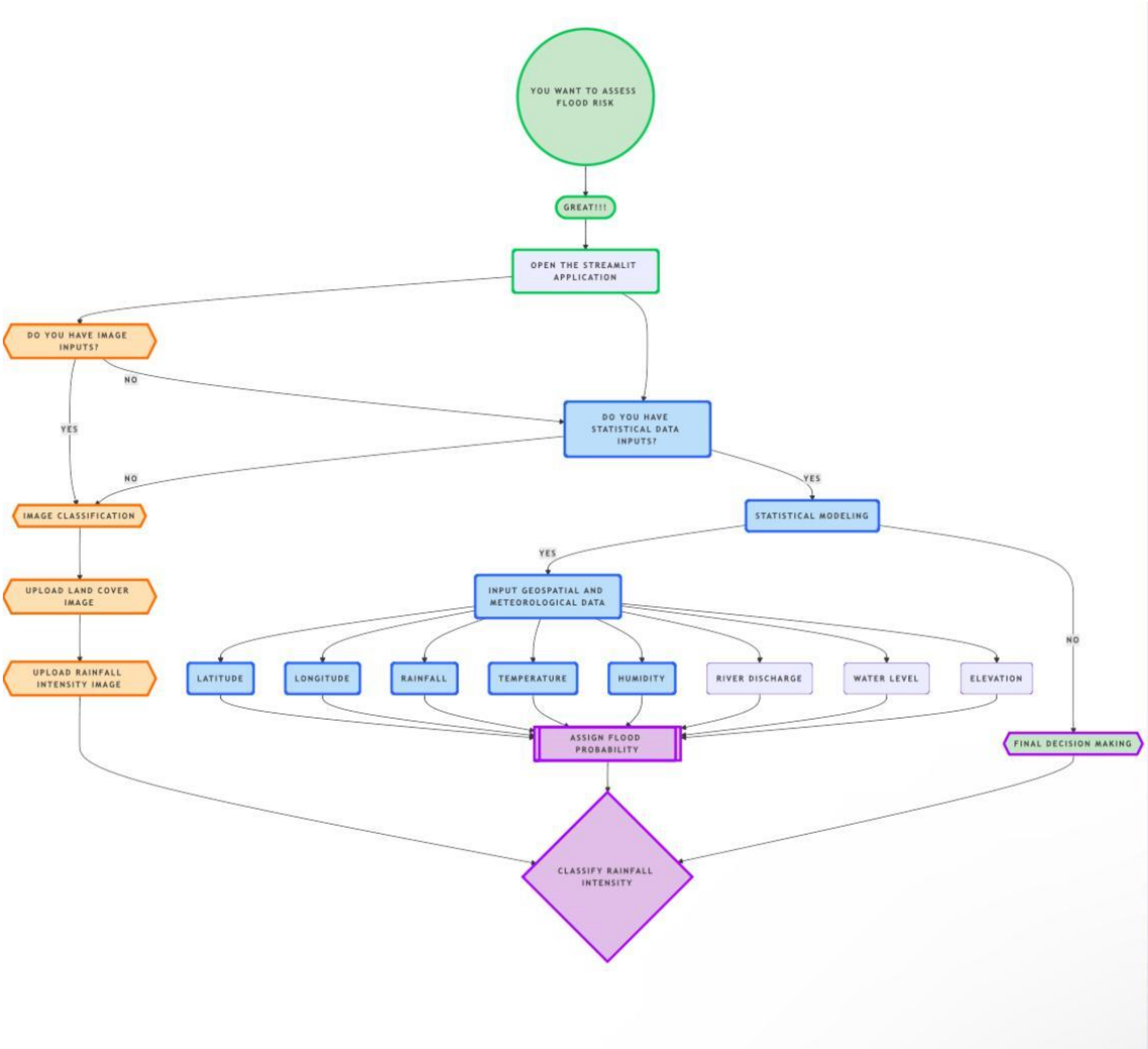

Fig 2. Flow of the program

The flowchart Fig 2. illustrates a comprehensive flood prediction system combining image classification and statistical modeling to provide accurate flood risk assessments. Designed as a hybrid framework, the system is implemented in a Streamlit application, offering an intuitive and user-friendly interface. The primary objective is to leverage diverse data inputs, including images and statistical parameters, to make reliable flood predictions. The approach is designed to account for both static and dynamic flood risk factors, integrating them into a weighted probability mechanism for final decision-making.

The **first approach** utilizes image classification to assess flood probabilities based on land cover and rainfall intensity. Users begin by uploading an image representing a specific land cover type, such as deforestation, paddy fields, or construction areas. The classification model, trained on folders containing labeled images of these categories, processes the input image and identifies the corresponding land cover type. Each land cover type has unique implications for flood risk. For instance, deforestation leads to increased water runoff due to reduced vegetation, while paddy fields, designed to hold water, may mitigate flood risks to some extent. Construction zones, on the other hand, may exacerbate floods by reducing natural drainage. The system assigns a flood probability value to each identified land cover type, ensuring that this information contributes to the final prediction.

Next, the user uploads an image representing rainfall intensity. This model classifies rainfall into three categories: light, moderate, and heavy rainfall, based on the visual features of the input image. The classification model is trained on diverse datasets of rainfall images, ensuring it can accurately assess the input. Rainfall intensity is a critical factor in flood prediction, with heavy rainfall significantly increasing the likelihood of floods. The system assigns a probability value to the rainfall category, which is then integrated with the land cover probability. Together, these two image-based classifications contribute to a combined probability score. While their individual weights in the final prediction are modest (e.g., 0.05 each), these insights capture real-time environmental changes that might not be reflected in traditional statistical data alone.

The **second approach** focuses on statistical modeling, which provides a broader and more detailed representation of flood risks. Users input geospatial and meteorological data, including latitude, longitude, rainfall (in mm), temperature (°C), humidity (%), river discharge (m^3/s), water level (m), and elevation (m). Additional parameters, such as land cover type, soil type, population density, infrastructure, and historical flood occurrences, are also considered. The statistical model analyses these inputs to identify correlations and patterns indicative of flood risks. For instance, high rainfall levels combined with low elevation and significant river discharge could result in a higher flood probability. Conversely, areas with effective drainage systems, favourable soil types, or historically low flood occurrences may have reduced risk.

The statistical model is trained on historical datasets, allowing it to provide accurate flood risk predictions based on user inputs. This model is assigned the highest weight (e.g., 90%) in the final prediction because of its ability to consider a wide range of variables. Unlike the image classification models, which provide qualitative insights, the statistical model offers a quantitative analysis of flood risks, making it the backbone of the system.

Once the probabilities from the image classification and statistical models are generated, they are combined using a weighted probability mechanism. This approach ensures that the statistical model, which provides a comprehensive assessment, has the most significant influence on the final prediction. However, the image-based insights are not ignored, as they capture dynamic environmental factors such as recent deforestation or unexpected heavy rainfall. These real-time observations enhance the model's overall accuracy and adaptability.

The hybrid framework outlined in the flowchart has several advantages. By combining image classification with statistical modeling, the system ensures a holistic flood risk assessment. Image-based models allow for the incorporation of real-time data, making the system adaptable to rapid environmental changes. For instance, the sudden clearing of forests or unseasonal heavy rainfall can significantly alter flood risks, and these factors are effectively captured through image analysis. Meanwhile, the statistical model provides a stable foundation for predictions, leveraging historical data and established relationships between various flood-related factors.

The Streamlit implementation makes the system accessible to users without technical expertise. The interactive interface allows users to upload images, input statistical data, and receive flood predictions in real time. This accessibility ensures that the system can be utilized by diverse stakeholders, including disaster management agencies, local governments, and even individual users in flood-prone areas. Furthermore, the modular design of the system allows for future scalability. Additional data sources, such as remote sensing imagery or IoT-based measurements, can be integrated to enhance prediction accuracy further.

In conclusion, the flowchart presents a robust flood prediction system that effectively integrates image classification and statistical modeling. The hybrid approach balances real-time environmental insights with comprehensive data analysis, offering a reliable solution for flood risk assessment. This innovative framework has the potential to revolutionize disaster management by enabling proactive measures, improving resource allocation, and minimizing the impacts of floods. By leveraging diverse data sources and combining them into a unified prediction model, this system sets a new standard for flood risk prediction and mitigation.

Flood Probability Calculation (Weighted Sum of Probabilities): The final flood probability is calculated by combining the outputs of the land cover classification, rainfall intensity classification, and statistical model using a weighted probability mechanism:

$$\mathrm{PFlood} = \mathrm{W1} \times \mathrm{PLandcover} + \mathrm{W2} \times \mathrm{PRainfallintensity} + \mathrm{W3} \times \mathrm{PStatisticalModel}$$

Where-:

- $P_{Flood\text{-}:}$ Final flood probability
- $P_{LandCover}$-:Probability derived from Landcover classification
- $P_{rainfallintensity}$-:Probability derived from rainfall intensity classification
- Pstatisticalmodel-:Probability derived from statistical prediction model.
- w1,w2,w3 are the weights assigned to each model. In this case, w1=w2=0.05w_1 = w_2 = 0.05w1=w2=0.05 and w3=0.90w_3 = 0.90w3=0.90.

7. HARDWARE AND SOFTWARE REQUIREMENTS

To effectively run the project described in the provided code files, specific hardware and software requirements must be met. Below is a detailed breakdown of these requirements.

Hardware Requirements

1. **Processor**:
   - Dual-core processor (Intel i3)
   - Quad-core processor (Intel i5 or equivalent) for better performance.
2. **RAM**:
   - 8 GB
   - 16 GB or more, especially for handling large datasets and running multiple applications simultaneously.
3. **Storage**:
   - 100 GB of free disk space
   - SSD for faster data access and loading times.
4. **Network**:
   - Stable internet connection for API access and data retrieval.

Software Requirements

1. **Operating System**:
   - Windows 10 or later, macOS, or a Linux distribution (Ubuntu recommended).
2. **Python**:
   - Version: Python 3.7 or higher.
   - Installation of Python can be done via Anaconda, which simplifies package management.
3. **Required Libraries**:
   The following Python libraries need to be installed, which can be done using pip:

bash

```
pip install streamlit

pip install pandas

pip install numpy

pip install folium

pip install Pillow

pip install streamlit-folium
```

4. **Inference SDK**:
   - The inference_sdk library is required for making API calls to classify rainfall and deforestation images. Ensure it's installed as per the documentation provided by the API service.
5. **Model Files**:
   - Ensure that the model files (flood_batata.pkl) and any required datasets (like flood_risk_dataset_india.csv) are accessible in the working directory.
6. **API Access**:
   - An active account with the inference API provider (e.g., Roboflow) is necessary to obtain an API key for image classification tasks.
7. **Web Browser**:
   - A modern web browser (like Chrome, Firefox, or Safari) is required to run the Streamlit application and view results interactively.

By meeting these hardware and software requirements, users will be able to successfully deploy and run the flood risk assessment application as outlined in the provided code files.

8. RESULT ANALAYIS AND DISCUSSION

Accuracy: 51.15%

	Precision	Recall	F1-Score	Support
0	**0.49**	**0.51**	**0.50**	**966**
1	**0.53**	**0.52**	**0.52**	**1034**
Accuracy			**0.51**	**2000**
Macro avg	**0.51**	**0.51**	**0.51**	**2000**
Weighted avg	**0.51**	**0.51**	**0.51**	**2000**

Table 1: Performance Metrics

The classification report for our flood prediction model provides a detailed analysis of its performance, offering valuable insights into how well the model differentiates between flood and no-flood instances. The model's overall accuracy stands at 51.15%, which, while slightly above random guessing, indicates substantial room for improvement. This relatively low accuracy suggests that the model is not yet robust enough to make reliable predictions in real-world scenarios, where flood prediction is critical for disaster management and risk mitigation.

Looking at the individual class performance, the model shows similar precision, recall, and F1-score values for both classes (flood and no-flood). For the flood class (class 1), the precision is 0.49, recall is 0.51, and the F1-score is 0.50, while for the no-flood class (class 0), the precision is 0.53, recall is 0.52, and the F1-score is 0.52. These results indicate that the model is fairly balanced in terms of identifying both flood and no-flood instances. However, the performance metrics for both classes remain modest, reflecting that the model is not highly accurate in distinguishing between flood and no-flood cases. The recall scores for both classes suggest that the model is somewhat effective in identifying flood events (class 1), but it still misses a considerable portion of the instances, leading to false negatives.

Additionally, the macro and weighted averages for precision, recall, and F1-score are all around 0.51, reinforcing the balanced but moderate performance of the model. These averages reflect that the model is neither biased toward predicting floods nor no-floods, but it still fails to achieve high predictive accuracy in either category. This balanced performance may be suitable for certain applications but falls short of being reliable for operational flood prediction, especially when precision in early flood detection is crucial for effective disaster management.

In conclusion, while the model maintains consistent performance across both flood and no-flood classes, the overall accuracy and evaluation metrics highlight its limitations in providing reliable and actionable predictions. The current performance is not sufficient for practical deployment, particularly in high-stakes environments such as disaster response. The results suggest that there is a need for further refinement of the model, including better data preprocessing, feature engineering, and possibly incorporating more advanced machine learning techniques. Additionally, exploring more complex algorithms, such as ensemble methods or deep learning approaches, could enhance the model’s discriminative capability and improve its accuracy. Furthermore, improving the quality and quantity of the training data, particularly in underrepresented classes, could also play a significant role in boosting the model's predictive power. The current findings indicate that while the model serves as a foundational prototype, substantial enhancements are needed to make it reliable enough for real-time flood prediction applications.

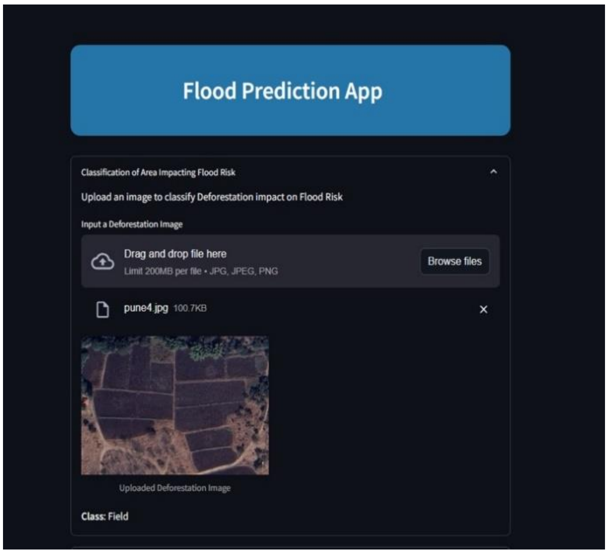

Fig 3. Output Image 1

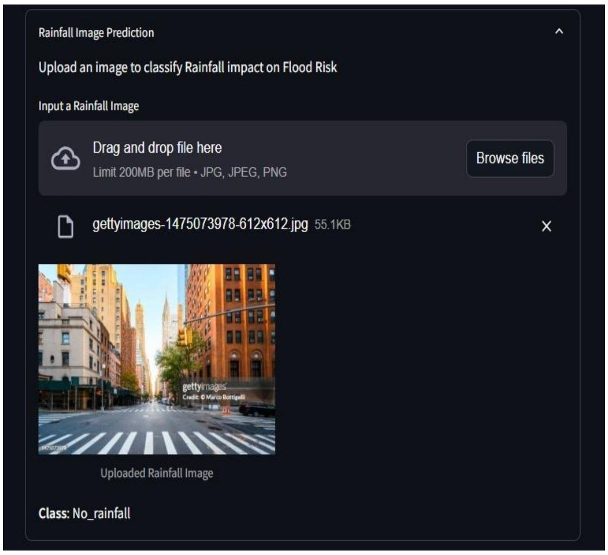

Fig 4. Output Image 2

These are the following output figures through which user will enter his image and based on which we will complete our first phase that is inserting images by user and assigning probabilities.

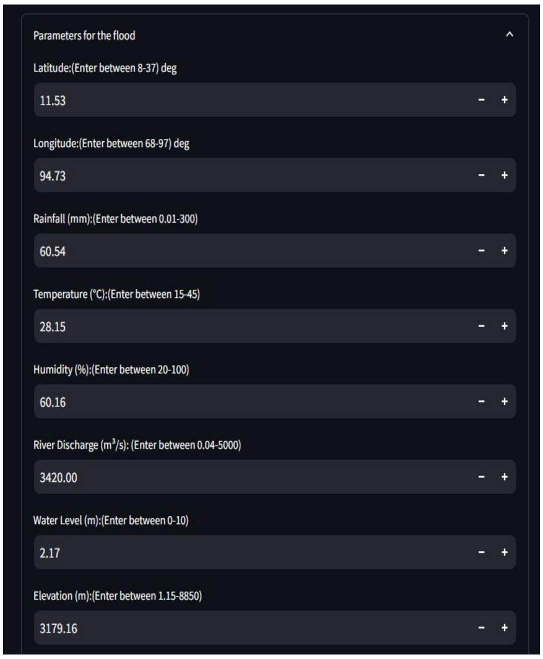

Fig 5. Output Image 3

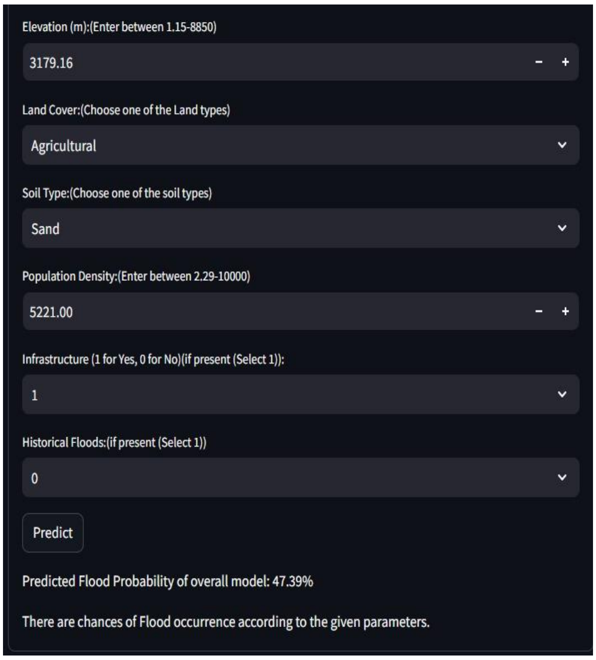

Fig 6. Output Image 4

In this study, we developed a comprehensive flood prediction model deployed on a Streamlit platform, utilizing two distinct approaches: image-based classification and statistical data analysis. In the first approach, users input images into the system, which classifies them into predefined categories. For instance, images of landscapes such as deforested areas, paddy fields, industrial zones, and construction sites are categorized to assess their impact on flood probability. This classification is performed using multiple datasets organized into folders containing representative images for each category. Based on the identified category, a flood probability value is assigned to the input image. Similarly, the user is prompted to upload rainfall-related images, which are classified into one of three rainfall intensity categories—light, moderate, or heavy rainfall—using another image dataset. A corresponding flood probability is also assigned to this input.

In the second approach, statistical data is collected from the user, including parameters such as latitude, longitude, rainfall (mm), temperature (°C), humidity (%), river discharge (m^3/s), water level (m), elevation (m), land cover, soil type, population density, infrastructure, and historical flood records. This data is used to predict flood occurrence through a statistical machine learning model trained on the aforementioned features.

To integrate these approaches, probability values from the two classification models (land type and rainfall classification) and the statistical model are combined. Probabilities of 0.05 are assigned to each classification model, while the remaining weight is allocated to the statistical model. This weighted approach enhances prediction accuracy by leveraging both qualitative image analysis and quantitative statistical insights, ensuring a holistic and robust flood prediction mechanism.

9. APPLICATIONS

The flood prediction project utilizing machine learning (ML) has various applications that enhance the accuracy and efficiency of forecasting flood events. Here are some key applications based on the gathered information:

1. **Real-Time Flood Forecasting Systems**: Machine learning techniques are increasingly being integrated into real-time flood forecasting systems. For example, researchers from the University of Iowa developed an ML-based system that combines data from multiple sources, including river gauges and weather forecasts, to provide timely flood predictions. This system has been successfully deployed in several U.S. states, aiding local authorities in making informed decisions during flood events.
2. **Image-Based Flood Probability Estimation**: ML algorithms can process images to classify land cover types and rainfall intensity. By analyzing images of land cover (e.g., deforestation, agricultural areas) and rainfall conditions, the system can assign flood probability values based on the identified categories. This approach allows for real-time assessments of environmental changes that may influence flood risks.
3. **Integration of Diverse Data Sources**: Machine learning excels in integrating large datasets from various sources, such as weather stations, satellite imagery, and IoT sensors. This capability enables the identification of complex patterns and relationships among environmental factors, leading to more accurate flood predictions compared to traditional models.
4. **Statistical Modeling and Data Analysis**: In addition to image classification, ML techniques can analyze geospatial and meteorological data inputs, including rainfall amounts, temperature, humidity, and river discharge rates. By identifying correlations between these variables and historical flood occurrences, ML models can provide comprehensive assessments of flood risks.
5. **Flood Alert Systems**: ML-based systems can be designed to detect changes in water levels through video footage analysis using computer vision algorithms. These systems can send alerts to relevant authorities and citizens via mobile applications, notifying them of potential flooding and providing real-time updates on the situation.

10.FUTURE SCOPE

The future scope of the flood prediction project utilizing machine learning (ML) is broad and promising, with numerous opportunities for enhancement and expansion. Here are several key areas where the project can evolve:

1. **Integration of Advanced Machine Learning Techniques**: As machine learning continues to advance, the project can incorporate more sophisticated algorithms such as ensemble methods, hybrid models, and deep learning techniques like Long Short-Term Memory (LSTM) networks. These methods can improve predictive accuracy by capturing complex temporal patterns in flood data, making the system more robust against varying conditions.

2. **Real-Time Data Processing:** The future development of the project could focus on integrating real-time data from Internet-of-Things (IoT) sensors deployed in flood-prone areas. This would allow for continuous monitoring and immediate updates to flood predictions based on changing environmental conditions, enhancing the system's responsiveness and reliability.

3. **Enhanced Data Sources**: Expanding the range of data sources used for predictions is crucial. Future iterations could include satellite imagery, social media data for crowd-sourced information during floods, and historical flood records to enrich the dataset. This integration would provide a more comprehensive view of flood risks and improve predictive capabilities.

4. **Climate Change Adaptation**: Given the increasing frequency and severity of floods due to climate change, future developments could focus on modeling climate scenarios to assess their impact on flood risks. Machine learning models can be trained using climate projection data to predict how changing weather patterns may influence future flooding events.

5. **User-Centric Features**: Enhancing user experience through the development of mobile applications or web interfaces that provide personalized alerts and recommendations based on user location and risk levels can be beneficial. This would empower individuals and communities to take proactive measures in response to predicted flooding.

11. IMPLEMENTATION PLAN

Work Plan	**Aug 2024**	**Sept 2024**	**Oct 2024**	**Nov 2024**	**Dec 2024**
1. Literature Survey	✓				
2. Data Collection	✓	✓			
3. Model Development		✓	✓		
4. System Integration			✓	✓	
5. Testing & Validation				✓	
6. Final Presentation					✓

Table 2. Implementation Plan

12.REFERENCES

[1] D. Han, L. Chan, and N. Zhu, "Flood forecasting using support vector machines," *Journal of Hydroinformatics*, vol. 9, no. 4, pp. 267-276, 2007.

[2] J. Noymanee and T. Theeramunkong, "Flood forecasting with machine learning technique on hydrological modeling," *8th International Young Scientist Conference on Computational Science*, 2017.

[3] D. T. Nguyen and S.-T. Chen, "Real-time probabilistic flood forecasting using multiple machine learning methods," *Journal of Hydrology,* vol. 586, pp. 124868, 2020.

[4] G. Kan et al., "Hybrid machine learning hydrological model for flood forecast purpose," *Open Geosciences*, vol. 12, no. 1, pp. 813–820, 2020.

[5] M. Bayat and O. Tavakkoli, "Application of machine learning in flood forecasting," *Future Technology*, vol. 1, no. 1, pp. 23-34, 2022.

[6] P. Pujari, N. Kulkarni, V. C. Hiremath, and V. Bhushi, "Machine-learning algorithms for forecast-informed reservoir operation (FIRO) to reduce flood damages," *Springer Nature*, 2024.

[7] M. H. M. Ali et al., "Flood prediction using deep learning models," *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 13, no. 2, pp. 167–176, 2022.

[8] N. Razali, S. Ismail, and A. Mustapha, "Machine learning approach for flood risks prediction," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 9, no. 4, pp. 674–682, 2020.

[9] N. Ahamed and S. Asha, "Flood prediction forecasting using machine learning algorithms," *International Journal of Scientific & Engineering Research*, vol. 11, no. 5, pp. 334–341, 2020.

[10] A. Kharche, P. Bhagat, and S. F. Ibrahim, "A review on flood prediction using machine learning-based Apache SystemML Python platform," *Journal of Emerging Technologies and Innovative Research (JETIR)*, vol. 6, no. 1, pp. 23–27, 2019.