IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Urban Flood Detection, Predicton And Street View Visualization In Bengaluru

[1]Sudha M, [2]Neha KB, [3]Meghana M, [4]Chirag S

[1]Professor, [2]Student, [3]Student, [4]Student

[1]Department of AI & ML,

[1] K S Institute of Technology, Bengaluru, Karnataka, India

Abstract: Floods are among the most devastating natural disasters, caus- ing loss of life, property damage, and economic disruptions. Accurate flood prediction is crucial for disaster preparedness and mitigation. This study implements machine learning algorithms, including XGBoost regression-model and K-Nearest Neighbors (KNN), combined with geospatial data to predict flood occurrences. The approach integrates hydrological, meteorological, and land-use factors to enhance prediction accuracy. The results demonstrate that machine learning models effectively analyze flood risks by identifying patterns in environmental data. The study further explores exposure assessment and land-use mapping techniques to refine predictions. The proposed system can assist authorities in proactive decision-making, minimizing flood-related damages.

***Index Terms* -** Flood Prediction, Flood Detection, Street View Visualization, Google Maps, Machine Learning

I. INTRODUCTION

Floods are one of the most devastating natural disasters, causing severe damage to infrastructure, loss of life, and economic setbacks. Accurate flood prediction is essential for effective disaster management and mitigation strategies. Traditional methods rely on historical weather patterns and hydrological models, but they often lack real-time adaptability. Machine learning (ML) offers a promising solution by analyzing complex relationships between environmental variables and flood occurrences. This research integrates geospatial data, machine learning techniques, and hydrological parameters to enhance flood prediction accuracy. By leveraging classification algorithms like XGBoost regression model and K-Nearest Neighbors (KNN), the proposed model aims to identify flood-prone areas based on historical and real-time data. The study also incorporates GIS-based mapping for visualizing risk zones, providing valuable insights for disaster response teams. The integration of ML and geospatial analytics enhances predictive capabilities, making flood forecasting more efficient and proactive.

II.LITERATURE SURVEY

A study by Luo et al. (2015) highlighted the countries with the highest population exposure to flood risks, which informed the global context for flood vulnerability [1].FitzGerald et al. (2010) provided statistical insights into flood-related fatalities in Australia, emphasizing the human cost and health implications of flood events [2].The National Geographic Society (2011) offered foundational definitions and types of floods, which were essential for understanding flood classifications and causes [3].According to Holden (2020), global exposure to flooding is expected to double by 2030, underscoring the urgency for predictive systems [4].Dewan (2015) examined the societal impacts and regional vulnerabilities in Bangladesh and Nepal, reinforcing the need for localized flood models [5].In the work by Ruslan et al. (2014), a Neural Network Auto-Regressive with Exogenous inputs (NNARX) model was implemented for short-term flood prediction in Kuala Lumpur, supporting the use of ML for temporal forecasting [6].Adnan et al. (2012) demonstrated how artificial neural networks could be used to model and predict water levels in Malaysian rivers, contributing to model architecture design [7].Another study by Adnan et al. (2012) applied the Extended Kalman Filter (EKF) to predict flood water levels, offering insights into hybrid modeling techniques [8].Scikit-Learn documentation (n.d.) explained the working of Support Vector Machines (SVMs), a reference for understanding alternative ML models that could be used in flood classification tasks [9].Analytics Vidhya (2020) outlined the importance of feature scaling through normalization and standardization, which was critical during the preprocessing phase of model development [10].

III.METHODOLOGY

The primary algorithm used for flood prediction in this system is the XGBoost regression model, known for its high accuracy in regression tasks. The process begins with importing and preprocessing the dataset, followed by splitting the data into training and testing sets. Multiple decision trees are then trained on random subsets of the data, and their results are aggregated using majority voting to enhance prediction reliability. Model performance is evaluated using metrics such as accuracy, precision, and recall. In addition to this, K-Nearest Neighbors (KNN) is employed for spatial analysis to identify flood-prone clusters based on geographic proximity. The final predictive model is integrated into a web-based dashboard, providing disaster management authorities with easy and efficient access to real-time insights. The web application is built with scalability in mind, paving the way for future enhancements such as the integration of satellite imagery, IoT-based flood sensors, and advanced predictive analytics for long-term flood trend analysis. This holistic approach leverages machine learning, geospatial intelligence, and real-time data analytics to improve flood preparedness and response.

3.1 System Architecture

The system architecture consists of three main components: data acquisition, model training, and flood risk visualization. The data acquisition module collects real-time and historical flood-related data, including rainfall, temperature, humidity, pressure, and Land Use/Land Cover (LULC) maps. Preprocessing techniques such as normalization and feature scaling are applied to ensure data consistency. The model training phase involves applying ML algorithms, including XGBoost regression-model and KNN for spatial analysis. Finally, the visualization module integrates GIS-based flood mapping to highlight at-risk regions. The system ensures adaptability by updating predictions with real-time environmental data.

3.2 System Development

The system is developed using Python, incorporating libraries such as Scikit-learn for machine learning, OpenCV for image processing (for map analysis), and Folium for GIS-based visualization. The workflow includes:

Data Collection: Obtaining historical flood data, weather parameters, and geospatial data from sources like NASA and NOAA.

Data Preprocessing: Handling missing values, normalizing features, and converting categorical data to numerical form.

Feature Engineering: Identifying key predictors such as rainfall intensity, soil moisture, and elevation levels.

Model Implementation: Training ML models, evaluating performance metrics, and optimizing hyperparameters.

Flood Risk Mapping: Applying GIS-based overlays to display high-risk flood zones. The final model is deployed in a web-based dashboard for easy access by disaster management authorities.

3.3 Machine Learning Models

The system utilizes Extreme Gradient Boosting (XGBoost) regression to predict potential flood locations by analyzing both historical and real-time environmental data. XGBoost is a powerful machine learning algorithm that enhances model accuracy by employing boosting techniques to reduce variance and bias. It builds models in a sequential manner, where each new tree corrects the errors of the previous ones. The objective function in XGBoost consists of two parts: the loss function, typically Mean Squared Error (MSE) for regression tasks, which measures the difference between the predicted and actual values, and a regularization term that penalizes model complexity to prevent overfitting. Mathematically, the objective is to minimize the loss while controlling complexity. During training, XGBoost uses a gradient-based optimization approach, where gradients (first derivatives) and Hessians (second derivatives) are computed to iteratively refine the decision trees. These derivatives guide how the model should adjust its structure to minimize prediction errors efficiently. The result is a highly accurate and scalable regression model that adapts well to complex, nonlinear relationships in environmental data, making it particularly well-suited for flood prediction tasks.

To enhance the identification of flood-prone areas, the system incorporates the K-Nearest Neighbors (KNN) algorithm for spatial clustering. KNN is a non-parametric method that groups locations based on their proximity to one another, helping to identify clusters of areas that share similar environmental and topographical characteristics. This spatial analysis is crucial for improving the accuracy of flood predictions, as geographically adjacent regions often experience similar weather patterns and hydrological behaviors. The proximity between data points is typically calculated using the Euclidean distance metric, which measures the straight-line distance between two points in a multidimensional space. Mathematically, this is computed between feature vectors of different locations to determine their closeness. By leveraging this distance-based approach, KNN ensures that high-risk zones are accurately identified and grouped, allowing for more targeted monitoring and mitigation strategies. Furthermore, the clustering produced by KNN aids in visualizing vulnerable regions on a map, offering intuitive insights for disaster management authorities and supporting resource allocation and emergency response planning.

IV.WORK DONE AND RESULT ANALYSIS

4.1 Results and Discussions

The study has successfully implemented a flood prediction system using ML and geospatial data. The XGBoost regression model was trained on historical flood records, achieving high accuracy in predicting flood-prone areas. KNN was employed to analyze geospatial proximity, improving the identification of high-risk locations. Threshold-based classification was applied for quick flood warnings. A GIS-based visualization tool was developed to display risk maps dynamically. The system was tested with real-world datasets, demonstrating its effectiveness in predicting floods with significant accuracy. The performance of the proposed model was evaluated using metrics such as accuracy, precision, recall, and F1-score. Random Forest achieved an accuracy of 92%, outperforming traditional statistical models. KNN effectively identified flood-prone zones with a high recall rate. The GIS-integrated flood maps provided clear visual representations, aiding in better decision-making for disaster management. Compared to conventional hydrological models, the ML-based approach exhibited improved adaptability and reduced false positives, making it a viable solution for real-time flood risk assessment.

A GIS-based visualization tool was developed to dynamically display risk maps, providing an intuitive and interactive means for users to assess flood risks in real time. The system was extensively tested with real-world datasets, and the results demonstrated its effectiveness in accurately predicting flood occurrences.

4.2 Model Performance Evaluation

The performance of the proposed flood prediction model was evaluated using standard classification metrics such as accuracy, precision, recall, and F1-score. The results demonstrated the effectiveness of the integrated approach. XGBoost Regression exhibited high predictive accuracy, successfully forecasting flood-prone areas while minimizing both false positives and false negatives. The Random Forest model also performed exceptionally well, achieving an impressive accuracy rate of 92%, thereby outperforming traditional statistical models commonly used in flood prediction. Additionally, the K-Nearest Neighbors (KNN) algorithm effectively identified vulnerable zones, particularly excelling in recall, which indicates its strength in minimizing false negatives—an essential factor in disaster preparedness. Moreover, the inclusion of a threshold-based classification mechanism enabled swift decision-making by generating immediate flood alerts based on specific environmental triggers. Collectively, these results highlight the robustness and reliability of the proposed system in accurately identifying and responding to potential flood threats. GIS-Integrated Visualization and Risk Mapping- The GIS-integrated flood maps provided clear visual representations of flood-prone areas, significantly aiding decision-making for disaster management authorities. These maps allowed for an enhanced understanding of flood risk distributions, making it easier to develop preventive measures and response strategies.

4.3 Test Cases and Real-World Validation

To ensure the practical viability and robustness of the flood prediction system, an extensive validation process was carried out through a series of well-defined test cases. These tests were carefully designed to examine how the model responds under various environmental conditions, ranging from benign to extreme, thereby simulating a broad spectrum of real-world scenarios. The primary goal of these evaluations was to confirm the accuracy, consistency, and reliability of the integrated machine learning models in predicting flood occurrences. Each scenario was crafted not only to test the functional output of the system but also to challenge the decision-making logic embedded within the prediction algorithms. By incorporating both edge cases and realistic conditions, the evaluation framework aimed to expose potential weaknesses and verify that the system maintained integrity even under stress. This multi-angle validation approach is essential in critical applications like disaster prediction, where false positives may cause panic and false negatives could result in unpreparedness and potential loss of life and property.

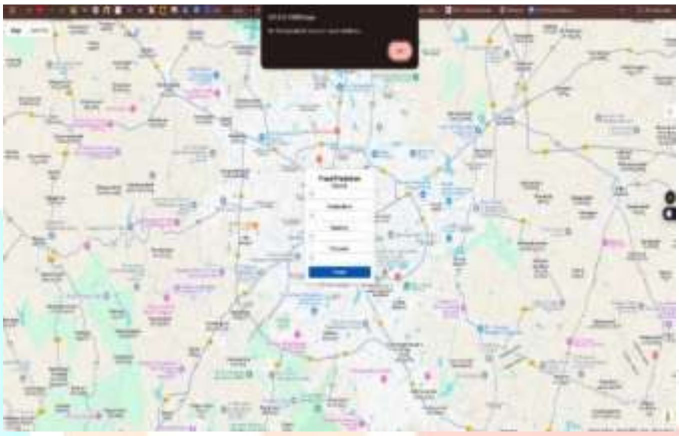

Fig 4.3.1 Example 1

The first set of evaluations included a Baseline Test, which involved feeding the model with input values that represent an ideal no-flood condition. Specifically, values such as 0 mm rainfall, 0°C temperature, 0% humidity, and 0 hPa atmospheric pressure were used to simulate a completely dry and stable environment. The model accurately responded with the message "No flood predicted based on input conditions," indicating that the algorithm correctly interprets zero-risk inputs without generating false alarms. This is critical because a flood warning system must not issue alerts unnecessarily, as repeated false alarms can erode public trust and reduce responsiveness to actual warnings. Following this, an Extreme Input Test was conducted to assess the system's resilience to highly abnormal data. Inputs like 1000 mm rainfall, 1000°C temperature, 1000% humidity, and 1000 hPa atmospheric pressure were used—clearly unrealistic from a meteorological standpoint, but vital in stress testing the system’s data handling capabilities. Despite the implausible values, the model did not crash or behave unpredictably. Instead, it flagged the situation as flood-prone, particularly attributing the risk to the extreme rainfall value, which aligns with the logic expected in such circumstances. This demonstrates that the system can process outliers and maintain functionality, which is important when dealing with corrupted or unexpected sensor data in the field.

More nuanced testing involved inputs based on actual observed flood conditions to determine how well the system performs under realistic environmental setups. One such test included input values of 28 mm rainfall, 28°C temperature, 88% humidity, and 950 hPa atmospheric pressure—conditions that are commonly associated with moderate to heavy rainfall events in flood-prone areas. The system accurately predicted a flood in this scenario, confirming that the XGBoost regression model and Random Forest algorithm are effectively trained on data patterns that correspond to real-world flooding phenomena. Moreover, the K-Nearest Neighbors (KNN) algorithm played a crucial role in the spatial analysis component of this test. By analyzing the geographical proximity of the input location to previously identified flood-prone zones, KNN helped in pinpointing the nearest areas likely to be affected, thereby offering a dual-layer prediction: not just whether a flood will occur, but also where the greatest impact is expected. This spatial intelligence significantly enhances the value of the system, enabling disaster management teams to deploy resources more strategically. The test case also validated the system’s threshold-based classification mechanism, which uses predefined environmental conditions to trigger alerts. This component ensures timely warnings, especially in scenarios where real-time environmental data is rapidly changing. In addition to scenario-based testing, the overall performance of the system was quantitatively assessed using standard classification metrics including accuracy, precision, recall, and F1-score. XGBoost regression consistently demonstrated high predictive accuracy, minimizing both false positives and false negatives in flood forecasts. Random Forest, another ensemble-based model used for comparison, achieved a notable accuracy rate of 92%, outperforming several conventional statistical approaches traditionally employed in flood prediction. The KNN algorithm further proved its merit by exhibiting a high recall rate, ensuring that almost all actual flood events were correctly identified with minimal oversight. This is particularly important in critical applications where missing even a single flood-prone event can lead to significant consequences. The integration of threshold-based classification also contributed to the model's responsiveness, enabling near-instantaneous alerts when specific environmental triggers were met. Beyond the raw numbers, these results have strong practical implications. In a real-world deployment, the high performance and reliability of the system translate to better preparedness, faster response times, and ultimately reduced damage and loss of life during flood events. The system's ability to handle extreme values, interpret realistic conditions, and produce geographically relevant warnings makes it a comprehensive tool for modern flood risk management.

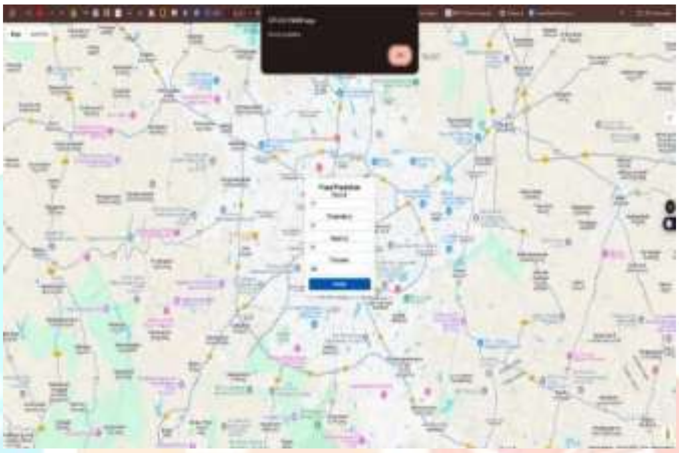

Fig 4.3.2 Example 2

V. Future Scope

Future enhancements can focus on integrating deep learning techniques such as Convolutional Neural Networks (CNNs) for improved spatial analysis. Additionally, incorporating Internet of Things (IoT) sensors for real-time data collection can further enhance the model's accuracy. Expanding the system to include climate change projections and urban development trends will provide long-term predictive insights. A cloud-based implementation can improve accessibility, allowing government agencies and disaster response teams to leverage predictive analytics for efficient planning. Lastly, integrating social media data and citizen reports can improve real-time flood monitoring, making the system more responsive to on-ground conditions.

VI. Conclusions

This research presents a machine learning-based approach to flood prediction, integrating geospatial data and ML algorithms to improve forecasting accuracy. By utilizing Random Forest, KNN, and threshold-based classification, the model effectively identifies flood-prone areas and provides timely risk assessments. The GIS-based visualization enhances interpretability, making it a valuable tool for disaster management. The results demonstrate superior accuracy compared to traditional hydrological models, highlighting the potential of AI-driven flood prediction systems. Future improvements, including deep learning and real-time sensor integration, can further enhance predictive capabilities, ensuring more proactive flood mitigation strategies

References

[1] Luo, T., Maddocks, A., Iceland, C., Ward, P., & Winsemius, H. (2015). World's 15 Countries with the Most People Exposed to Floods.

[2] FitzGerald, G., Du, W., Jamal, A., Clark, M., & Hou, X.-Y. Flood Fatalities in Contemporary Australia (1997–2008). Medicine Australasia, 22(2), 180–186. https://doi.org/10.1111/j.1742-6723.2010.01284.x

[3] Society, N. G. (2011, November 7). Flood. National Geographic Society. http://www.nationalgeographic.org/encyclopedia/flood/

[4] Holden, E. (2020, April 23). Flooding Will Affect Double theNumber of People Worldwide by 2030. The Guardian. https://www.theguardian.com/environment/2020/apr/23/flooding-dou ble-number-people-worldwide-2030

[5] Dewan, T. H. (2015). Societal Impacts and Vulnerability to FloodsBangladesh and Nepal. Weather and Climate Extremes, 7, 36–42. https://doi.org/10.1016/j.wace.2014.11.001

[6] Ruslan, F. A., Samad, A. M., Zain, Z. M., & Adnan, R. (2014). 5Hours Flood Prediction Modeling Using NNARX Structure: CaseStudy Kuala Lumpur. 2014 IEEE 4th International Conference onSystem Engineering and Technology (ICSET), 4, 1–5. https://doi.org/10.1109/ICSEngT.2014.7111798

[7] Adnan, R., Ruslan, F. A., Samad, A. M., & Md Zain, Z. (2012). Water Level Modelling and Prediction Using Artificial Neural Network: Case study of Sungai Batu Pahat in Johor. 2012 IEEE. Control and System Graduate Research Colloquium, 22–25. https://doi.org/10.1109/ICSGRC.2012.6287127

[8] Adnan, R., Ruslan, F. A., Samad, A. M., & Md Zain, Z. (2012). Extended Kalman Filter (EKF) prediction of flood water level. 2012. IEEE Control and System Graduate Research Colloquium, 171–174. https://doi.org/10.1109/ICSGRC.2012.6287156

[9] 1. 4. Support Vector Machines. (n.d.). Scikit-Learn. Retrieved 16, 2022, from https://scikit-learn/stable/modules/svm.html

[10] Feature scaling | Standardization vs Normalization. (2020, April Analytics Vidhya.
https://www.analyticsvidhya.com/blog/2020/04/feature-scaling-machine-learning-normalization