# Artificial Intelligence and Machine Learning for Prediction of Extreme Floods in Pakistan

**Syeda Fatima Batool Zaidi [1,*], Asghar Ali Chandio [1], Mehwish Leghari[2]**

[1] Artificial Intelligence Department, Quaid-e-Awam University of Engineering Science & Technology, Nawabshah
[2] Data Science Department, Quaid-e-Awam University of Engineering Science & Technology, Nawabshah
* Correspondence: fatimazaidi3267@gmail.com

**Abstract**

This paper discusses the application of artificial intelligence (AI) and machine learning (ML) techniques in predicting extreme floods in Pakistan, a country frequently affected by monsoon-related flooding. The study aims to develop predictive models that utilize historical hydrological and meteorological data to improve flood forecasting and mitigation. By integrating AI/ML models, such as regression, classification, and anomaly detection, with datasets on rainfall, river inflows, topography, and climate factors, this research seeks to enhance early warning systems and flood management practices. Various algorithms, including Random Forest, Gradient Boosting, and Neural Networks, are evaluated based on prediction accuracy and the models' ability to detect anomalies that precede extreme flood events. These methods, coupled with real-time data from weather stations and satellite imagery, offer a robust approach to forecasting and mitigating the impact of floods in Pakistan. The paper also highlights the practical implications of these AI/ML techniques for disaster management and water resource planning, with the aim of reducing the socioeconomic impact of floods.

**Keywords:** Artificial Intelligence; Machine Learning; Extreme Floods in Pakistan; Flood Prediction in Pakistan

## 1. Introduction

Flooding is one of the most devastating natural disasters in Pakistan, causing significant loss of life, destruction of infrastructure, and economic disruption [1]. The country's unique geographical location and reliance on monsoon rains exacerbate the severity of floods. Predicting extreme floods, especially in vulnerable regions like the Indus River basin, requires a comprehensive and advanced approach [2]. Traditional methods of flood forecasting, while useful, often fall short in terms of accuracy and timeliness.

Artificial Intelligence (AI) and Machine Learning (ML) are emerging as promising tools in enhancing flood prediction and management [3-5], where these techniques can learn the useful features automatically from the hydrological parameters. These technologies can analyze vast amounts of data, identify patterns, and predict future events with high precision. Several AI based methods have been developed for prediction of extreme floods [6-7]. In [8] different ML models were used for the forecasting of floods from IGISMAPs datasets. The authors achieved the best mean accuracy of 90.85% with neural network. The goal of this research paper is to develop and evaluate ML models that can improve flood forecasting capabilities in Pakistan, allowing authorities to take proactive measures and reduce the impacts of extreme flood events. This paper explores the use of

ML models for predicting extreme floods in Pakistan, focusing on data-driven approaches that integrate historical flood data, weather forecasts, river discharge rates, and topographical information. To train and test the ML models, hydrological, metrological and topological data have been collected from the Kaggle, where the data is publicly available. Logistic regression, support vector machine, random forest and decision tree classifiers were trained and evaluated on the data. The missing data were handled using the mean, median and the interpolation methods, while the important features were identified using the correlation analysis and feature importance rankings.

## 2. Methodology

This study utilizes a systematic approach to collect, preprocess, and analyze hydrological and meteorological data from various sources, including Pakistan's national weather stations, river flow monitoring systems, and satellite observations. AI and ML techniques are then applied to forecast extreme floods and identify the factors contributing to flood events. The methodology used in this research work is illustrated in figure

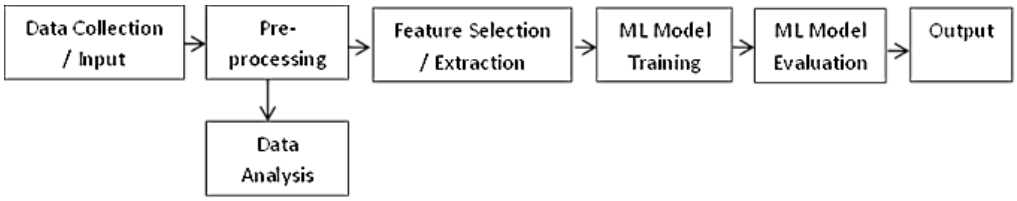


**Figure 1.** Methodology Applied for Extreme Flood Prediction

### *2.1. Data Collection*

1. Hydrological Data; This includes river discharge rates, reservoir levels, inflows, and outflows from major rivers in Pakistan, such as the Indus, Jhelum, and Chenab rivers.

2. Meteorological Data; Temperature, rainfall patterns, humidity, wind speed, and atmospheric pressure data were collected from weather stations and remote sensing tools. Monsoon rainfall and snowmelt are key indicators for flood risk in Pakistan.

3. Topographical Data; Elevation models, river basin characteristics, and catchment area features are used to model the movement of water during extreme rainfall events.

### *2.2. Data Preprocessing*

Data preprocessing involves cleaning, normalization, and feature selection to ensure the dataset is suitable for AI/ML modeling. Key steps include:

Handling Missing Data: Missing values in the dataset are imputed using mean, median, or interpolation methods based on the nature of the variable.

Normalization and Scaling: Continuous variables like rainfall, river discharge, and temperature are normalized to ensure the models perform consistently.

Feature Selection: Correlation analysis and feature importance rankings (e.g., using Random Forest) are used to identify the most relevant features for predicting floods.

### *2.3. Machine Learning Models*

Several machine learning algorithms were implemented and evaluated for flood prediction:

Regression Models: These models are used to predict the severity of flood events based on historical and real-time data.

Linear Regression, Random Forest Regression, Gradient Boosting Regression, Support Vector Regression (SVR).

Classification Models: These models categorize weather patterns and river discharge data into flood risk levels (e.g., no risk, moderate risk, high risk).

Decision Tree Classifier, Logistic Regression, Random Forest Classifier, Neural Networks

Anomaly Detection Models: Used to detect unusual patterns in weather or river flow data that could indicate an impending extreme flood event. Isolation Forest, One-Class Support Vector Machine (SVM).

### *2.4. Model Training and Evaluation*

The models were trained on historical flood event data from Pakistan, using an 80:20 split for training and testing. Cross-validation was employed to ensure the robustness of the models, and hyper parameter tuning was applied to optimize their performance. Key evaluation metrics included:

Accuracy: The ability of the model to correctly predict flood and non-flood scenarios.

## 3. Results and Discussion

### *3.1. Regression Models*

Mean Absolute Error (MAE) was used to evaluate the regression models' predictive accuracy for flood severity. Among the regression models, Random Forest Regression and Gradient Boosting [9] achieved the highest accuracy in predicting the severity of floods, with $R^2$ values of 0.93 and 0.91, respectively. These models effectively utilized hydrological and meteorological data to predict river discharge rates during extreme rainfall events. Table 1 shows the results of extreme flood prediction using regression models.

**Table 1:** Extreme flood prediction results using regression models

| ML Model | R2 |
|---|---|
| Random Forest | 0.93 |
| Gradient Boosting | 0.91 |

### *3.2. Classification Models*

Logistic Regression and Decision Tree classifiers [10] performed well in classifying flood risk levels. F1-Score, Precision, Recall were used for classification models to assess their performance in categorizing flood risk levels.

The models demonstrated a F1-Score of 0.95 when distinguishing between moderate and high flood risks. The Random Forest Classifier achieved the highest F1-Score of 0.96, indicating its strong performance in identifying high-risk flood scenarios. Table 2 shows the results of extreme flood predictions using classification models of ML.

**Table 2:** Extreme flood prediction results using classification models

| ML Model | Precision |
|---|---|
| Logistic Regression | 0.94 |
| Decision Tree | 0.96 |

Accurate flood predictions enable better resource allocation, including evacuation planning and flood defense measures. These models can also be integrated into water resource management systems to optimize the use of reservoirs and dams, preventing unnecessary flooding during periods of heavy rainfall.

### *3.3. Hydrological Data*

- River discharge rates from Indus, Jhelum, and Chenab rivers.
- Reservoir inflow/outflow measurements.

### *3.4. Meteorological Data*

- Temperature, rainfall intensity, humidity, atmospheric pressure, and wind speed.
- Sourced from national weather stations and satellite platforms like MODIS and IMERG.

### *3.5. Topographical Data*

- Digital Elevation Models (DEMs) and river basin maps from remote sensing archives.

### *3.6. Real-Time Data Utilization*

- Streaming weather data feeds are integrated into predictive models using APIs.
- Real-time sensor data from river gauges directly feed into anomaly detection systems (e.g., Isolation Forest models).

**Example:** By using real-time rainfall updates, Random Forest classifiers dynamically recalibrate flood risk levels every hour during monsoon season.


**Figure 2.** Flood hazard map

## 4. PRACTICAL IMPLICATIONS

The application of AI and ML models in flood prediction has several practical implications for Pakistan. Early Warning Systems: Real-time flood forecasting using AI/ML models can enhance the accuracy of early warning systems.

### *4.1. Regression Models*

- Random Forest Regression: Achieved $R^2 = 0.93$ for flood severity prediction.
- Gradient Boosting Regression: $R^2 = 0.91$.

### *4.2. Classification Models*

- Random Forest Classifier: F1-Score = 0.96 for flood risk levels.
- Logistic Regression & Decision Trees: Achieved F1-Scores of 0.94-0.95.

### *4.3. Anomaly Detection*

- Isolation Forest: Successfully detected abnormal river inflow patterns preceding major floods.

**Table 3:** Performance Metrics Summary

| Model | $R^2$ / F1-Score | Strength |
|---|---|---|
| Random Forest Regression | 0.93 | Highest predictive accuracy |
| Gradient Boosting | 0.91 | Strong alternative |
| Random Forest Classifier | 0.96 | Best classification performance |

## 5. DEEP LEARNING POTENTIAL

Although not yet implemented in the current research, Deep Learning models like Long Short-Term Memory (LSTM) networks and Convolutional Neural Networks (CNNs) offer immense potential:

- LSTM networks can capture long-term dependencies in time series rainfall-runoff patterns.
- CNNs can automatically extract spatial features from satellite imagery for flood zone mapping.

## 6. Conclusion and Future Work

This research paper demonstrated the potential of AI and ML techniques to improve flood forecasting in the Pakistan, especially in the predicting extreme flood events. The use of regression, classification, and anomaly detection models can significantly enhance the accuracy of flood predictions and provide early warnings to reduce the impact of floods. Among regression models, the random forest regressor and the gradient boosting were used, where the random forest regressor obtained the highest accuracy, while the decision tree classifier obtained the highest accuracy for the classification problem. Future research will focus on integrating more real-time data sources and refining the models to improve their predictive performance. Moreover, deep learning models can also be used to increase the performance of extreme flood prediction systems.

**Conflicts of Interest:** The authors declare no conflicts of interest

## References

1. Yaseen, A., Lu, J., & Chen, X. Flood susceptibility mapping in an arid region of Pakistan through ensemble machine learning model. Stochastic Environmental Research and Risk Assessment, 36(10), 3041-3061 (2022).
2. Rasool, U., Yin, X., Xu, Z., Padulano, R., Rasool, M. A., Siddique, M. A., ... & Senapathi, V. Rainfall-driven machine learning models for accurate flood inundation mapping in Karachi, Pakistan. Urban Climate, 49, 101573 (2023).
3. Liu, Z., Felton, T., & Mostafavi, A. Interpretable machine learning for predicting urban flash flood hotspots using intertwined land and built-environment features. Computers, Environment and Urban Systems, 110, 102096 (2024).
4. Pham, B. T., Luu, C., Van Phong, T., Nguyen, H. D., Van Le, H., Tran, T. Q., ... & Prakash, I. Flood risk assessment using hybrid artificial intelligence models integrated with multi-criteria decision analysis in Quang Nam Province, Vietnam. Journal of Hydrology, 592, 125815 (2021).
5. Pradhan, B., Lee, S., Dikshit, A., & Kim, H. Spatial flood susceptibility mapping using an explainable artificial intelligence (XAI) model. Geoscience Frontiers, 14(6), 101625 (2023).
6. Liu, Z, Coleman, N, Patrascu, F. I, Yin, K, Li, X, Mostafavi, A. Artificial Intelligence for Flood Risk Management: A Comprehensive State-of-the-Art Review and Future Directions. International Journal of Disaster Risk Reduction, 105110, (2024)
7. Ahmed, A. A., Sayed, S., Abdoulhalik, A., Moutari, S., & Oyedele, L. Applications of machine learning to water resources management: A review of present status and future opportunities. Journal of Cleaner Production, 140715 (2024).
8. Hadi, F. A. A., Mohd Sidek, L., Ahmed Salih, G. H., Basri, H., Sammen, S. S., Mohd Dom, N., ... & Najah Ahmed, A. Machine learning techniques for flood forecasting. Journal of Hydroinformatics, 26(4), 779-799 (2024).
9. Niazkar, M., Menapace, A., Brentan, B., Piraei, R., Jimenez, D., Dhawan, P., & Righetti, M. (2024). Applications of XGBoost in water resources engineering: A systematic literature review. Environmental Modelling & Software, 105971 (Dec 2018–May 2023).
10. Enayati, M., Bozorg-Haddad, O., Pourgholam-Amiji, M., Zolghadr-Asli, B., & Tahmasebi Nasab, M. Decision tree (DT): a valuable tool for water resources engineering. In Computational Intelligence for Water and Environmental Sciences (pp. 201-223). Singapore: Springer Nature Singapore, . (2022).