



IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 Issue: VI Month of publication: June 2025 DOI: https://doi.org/10.22214/ijraset.2025.72213

www.ijraset.com

Call: 🕥 08813907089 🔰 E-mail ID: ijraset@gmail.com



A Comparative Study of Machine Learning Algorithms for Flood Risk Prediction

Dr. D. P. Gaikwad¹, Onkar Sonawane², Sujit Walunj³, Vikas Lawand⁴, Pranav Nirwane⁵ Department of Computer Engineering, AISSMS College Of Engineering, Pune, India

Abstract: Floods are one of the most frequent and devastating natural disasters, significantly affecting human life, infrastructure, and the environment. Accurate and timely flood prediction is essential for disaster preparedness and mitigation. This study presents an intelligent flood risk prediction framework using machine learning and deep learning techniques applied to four Indian cities—Pune, Nashik, Kolhapur, and Satara—each associated with major rivers such as the Mula-Mutha, Godavari, Panchganga, and Krishna, respectively. The dataset incorporates a comprehensive set of hydrometeorological and environmental features including rainfall, temperature, humidity, wind speed, water level, discharge, groundwater level, soil moisture, atmospheric pressure, evaporation rate, and historical flood events. Four algorithms—Random Forest, Support Vector Machine (SVM), XG-Boost, and Artificial Neural Networks (ANN)—were trained and evaluated to predict the flood_risk level. The model performances were compared using accuracy, precision, recall, F1-score, and ROC-AUC. The results demonstrate that the integration of multiple data sources and ensemble techniques significantly improves predictive performance. This research contributes to the development of smart, data-driven flood early warning systems tailored for regional hydrological conditions.

Keywords: Flood Prediction, machine learning, deep learning, Random Forest, SVM, XG-Boost, ANN, environmental data, flood risk.

I. INTRODUCTION

Floods are among the most devastating natural disasters, causing significant human and economic losses, especially in countries like India where monsoon variability and unplanned urban growth contribute to frequent flooding. Cities such as Pune, Nashik, Kolhapur, and Satara, located near major rivers, are increasingly vulnerable due to factors like heavy rainfall, river overflow, and inadequate infrastructure. Traditional flood forecasting methods struggle to handle the complexity and real-time nature of flood dynamics. With the rise of Machine Learning (ML) and Deep Learning (DL), predictive models have become more accurate and adaptable. These technologies excel at analysing large, diverse datasets and identifying complex patterns, making them well-suited for flood risk prediction. Their ability to process real-time environmental and hydrological data with minimal human intervention offers a significant advantage over conventional approaches.

This study proposes a machine learning framework using environmental data from four flood-prone cities in Maharashtra. Key parameters include rainfall, temperature, humidity, river discharge, groundwater levels, and past flood events. The system employs and compares four models—Random Forest, SVM, XG-Boost, and ANN—to classify flood risk levels and determine the most effective algorithm based on performance metrics like accuracy, precision, recall, F1-score, and AUC-ROC.

A. Problem Statement

Floods in Maharashtra cities like Pune, Nashik, Kolhapur, and Satara have become more frequent due to climate change, erratic monsoons, and rapid urbanization. Existing flood forecasting

systems lack accuracy, spatial resolution, and adaptability, hindering timely disaster response. This research aims to develop a machine learning-based flood risk prediction framework that analyses dynamic environmental factors to accurately classify flood risk levels. By comparing models like Random Forest, SVM, XG-Boost, and ANN, the study seeks to build a scalable, real-time system to support early warning and disaster management efforts.

B. Objectives

 To collect and integrate multi-city environmental data including rainfall, temperature, humidity, wind speed, river water levels, groundwater levels, soil moisture, pressure, evaporation rate, and historical flood events from the cities of Pune, Nashik, Kolhapur, and Satara.



ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538 Volume 13 Issue VI June 2025- Available at www.ijraset.com

- 2) To preprocess the dataset by handling missing values, normalizing features, and encoding categorical data to ensure clean input for training and testing machine learning models.
- *3)* To build and train four machine learning models—Random Forest, Support Vector Machine (SVM), XGBoost, and a Deep Learning model (Artificial Neural Network)—for classifying flood risk into levels based on environmental indicators.
- *4)* To evaluate the performance of each model using relevant metrics such as accuracy, precision, recall, F1-score, and ROC-AUC, and identify the best-performing algorithm for deployment.
- 5) To develop an interactive web-based interface using the Flask framework, enabling users (e.g., municipal bodies, disaster response teams, and citizens) to:
- *a)* Upload live data or city-specific inputs
- b) View predicted flood risk level
- c) Visualize data through dynamic charts and model insights

II. METHODOLOGY

This research follows a structured, multi-phase methodology encompassing data acquisition, preprocessing, model training, performance evaluation, and real-time deployment via a web interface. The entire pipeline is designed to convert raw hydrometeorological data into actionable flood risk predictions using machine learning techniques and an interactive user interface.



Figure 1: System Architecture

In Figure [1], the step-by-step process followed for developing the flood prediction system. It begins with data collection from environmental and hydrological sources, including rainfall, temperature, humidity, and river discharge. The collected data is then preprocessed through cleaning, normalization, and handling of missing values to ensure quality input. Next, suitable machine learning models such as Random Forest, SVM, XG-Boost, and ANN are selected and built. These models are trained using historical data and then tested on unseen data to evaluate performance using metrics like accuracy, precision, recall, and F1-score. Finally, the prediction result is shown through a user-friendly dashboard or interface to support timely monitoring and decision-making.



ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538 Volume 13 Issue VI June 2025- Available at www.ijraset.com



Figure 2: Methodology

A. Data Collection and Integration

In Figure [2], the study begins by collecting extensive hydrometeorological data from four flood-prone cities in Maharashtra: Pune, Nashik, Kolhapur, and Satara. Each city corresponds to a major river—Mula-Mutha, Godavari, Panchganga, and Krishna, respectively. The dataset includes several critical environmental and climatic parameters such as rainfall (mm), temperature (°C), humidity (%), wind speed (km/h), water level (m), discharge (m³/s), groundwater level (m), soil moisture (%), atmospheric pressure (hPa), evaporation rate (mm/day), and previous flood events. These attributes were chosen for their direct influence on flood behavior, ensuring a robust basis for risk classification.



International Journal for Research in Applied Science & Engineering Technology (IJRASET) ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538 Volume 13 Issue VI June 2025- Available at www.ijraset.com

B. Data Preprocessing

In Figure [2], to prepare the data for training, comprehensive preprocessing was conducted. Missing values were handled using interpolation and statistical imputation techniques, while outliers were detected and treated using the Z-score and IQR methods. Numerical features were normalized to maintain consistency in scale, and categorical variables such as city names and river identifiers were encoded into machine-readable formats. After cleaning and transforming the data, it was split into training and testing sets in an 80:20 ratio to evaluate model performance under unbiased conditions.

C. Model Development

In Figure [2], the study began by collecting extensive hydrometeorological data from four flood-prone cities in Maharashtra: Pune, Nashik, Kolhapur, and Satara. Each city corresponds to a major river—Mula-Mutha, Godavari, Panchganga, and Krishna, respectively.Four machine learning and deep learning algorithms were implemented to model the flood risk classification: Random Forest, Support Vector Machine (SVM), XGBoost, and Artificial Neural Network (ANN). Random Forest was used for its robustness and ability to handle non-linear interactions, while SVM provided a margin-based classifier ideal for handling multi-class data. XGBoost was selected for its accuracy and efficiency in large-scale predictions, and ANN was deployed to capture complex, deep patterns within the data. Each model was trained using the processed dataset, and hyperparameters were fine-tuned using grid search and cross-validation to improve accuracy and generalization.

D. Model Evaluation

In Figure [2], the study began by collecting extensive hydrometeorological data from four flood-prone cities in Maharashtra: Pune, Nashik, Kolhapur, and Satara. Each city corresponds to a major river—Mula-Mutha, Godavari, Panchganga, and Krishna, respectively. Once trained, all models were evaluated using multiple performance metrics including accuracy, precision, recall, F1-score, and the ROC-AUC curve. This helped in assessing each model's capability to correctly predict flood risk levels—particularly focusing on reducing false negatives, which are critical in flood forecasting. The best-performing model was selected for deployment based on its balanced performance across all metrics.

E. Web-Based Deployment Using Flask

In Figure [2], to ensure practical usability, the final model was deployed through a Flask-based web interface. The GUI allows users to input real-time data or select city-specific parameters to receive instant flood risk predictions. The interface also includes data visualization features such as charts and graphs for monitoring key environmental indicators. This interactive system is designed to support early warning dissemination, helping authorities and citizens make proactive decisions during potential flood events. The modular nature of the framework also makes it scalable and adaptable to other geographic locations by simply retraining the models with new regional data.

F. Data Collection

In this study, data was collected from four major cities in Maharashtra, India—Pune, Nashik, Kolhapur, and Satara—each situated along prominent rivers prone to seasonal flooding: Mula-Mutha (Pune), Godavari (Nashik), Panchganga (Kolhapur), and Krishna (Satara). The dataset was compiled from authenticated sources, including regional meteorological departments, river basin authorities, and open government portals.

location	river_name	rainfall (mm)	temperature (°C)	humidity (%)	wind_speed (km/h)	water_level (m)
Pune	Mula- Mutha River	5.69	18.78	54.45	18.09	1.74
Nashik	Godavari River	13.36	22.61	52.39	11.35	3.10
Kolhapur	Panchganga River	16.67	17.32	56.48	12.46	3.02
Satara	Krishna River	11.93	15.08	43.14	15.46	1.68
Pune	Mula- Mutha River	16.78	22.71	67.33	7.20	2.30



ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538 Volume 13 Issue VI June 2025- Available at www.ijraset.com

roundwater_level (m)	soil_moisture (%)	pressure (hPa)	evaporation_rate (mm/day)	previous_flood_events	flood_risk
1.63	46.17	1015.10	7.61	0	Low
2.06	46.41	1016.27	7.32	0	Low
1.82	33.74	1019.07	5.56	0	Low
1.00	28.95	1023.14	7.37	0	Low
1.91	40.94	1023.63	5.82	0	Low

g

Figure 3: Data Collection

In Figure [3], the collected data consists of both meteorological and hydrological parameters, which are critical for accurate flood risk modeling. These parameters include:

- Rainfall (mm)
- Temperature (°C)
- Humidity (%)
- Wind Speed (km/h)
- Water Level (m)
- Discharge (m³/s)
- Groundwater Level (m)
- Soil Moisture (%)
- Atmospheric Pressure (hPa)
- Evaporation Rate (mm/day)
- Previous Flood Events (numeric indicator)

Each record is also tagged with the location (city) and associated river name to maintain geographic integrity. The target variable in this study is flood risk, which is a categorical feature representing the potential flood threat as Low, Moderate, or High.

The temporal range of the dataset spans multiple seasons to capture monsoonal variations and historical flood patterns, enabling the models to generalize well across different climatic conditions. Data consistency and relevance were ensured through rigorous preprocessing before being fed into machine learning pipelines.

III. DATA ANALYSIS

Data analysis plays a critical role in identifying meaningful patterns and relationships between environmental variables and flood risk. The dataset, which includes both meteorological and hydrological features across four cities—Pune, Nashik, Kolhapur, and Satara—was first explored using descriptive statistics, visualization techniques, and correlation matrices.

A. Flood Risk Distribution





In Figure [4], to better understand the overall class distribution of the target variable flood_risk, a visual representation was generated using a categorized count plot. The analysis revealed that the dataset contains a higher number of instances classified as



International Journal for Research in Applied Science & Engineering Technology (IJRASET) ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538 Volume 13 Issue VI June 2025- Available at www.ijraset.com

"Low" flood risk, followed by "Medium" and "High" risk levels. This indicates a moderate class imbalance, which can potentially influence the performance of classification models. The visualization clearly illustrated that low-risk events dominate the dataset, suggesting that extreme flood scenarios are relatively infrequent but critically important for early warning systems. This distribution insight was vital for applying balancing strategies such as stratified sampling during model training to ensure robust learning and fair evaluation across all risk categories.

B. Seasonal Analysis of Rainfall Trends

To analyse seasonal rainfall trends, the dataset was grouped by calendar month and the average rainfall was calculated accordingly. The resulting line plot, enriched with visual shading for clarity, illustrated a distinct monsoonal pattern, with peak rainfall occurring during the months of June to September.



Figure 5: Average Monthly Rainfall(mm)

In Figure [5], these months correspond to the Southwest Monsoon season in Maharashtra, during which the study regions—Pune, Nashik, Kolhapur, and Satara—typically receive the highest volume of precipitation. In contrast, the months from November to May exhibited significantly lower rainfall, indicating dry and pre-monsoon phases. This seasonal variability plays a crucial role in flood risk prediction, as elevated rainfall during monsoon months directly impacts river water levels, soil saturation, and discharge rates. Recognizing these monthly fluctuations not only aids in identifying critical flood-prone periods but also strengthens the temporal sensitivity of the machine learning models used in the study.

C. Interactive Flood Risk Visualization

An interactive scatter plot was developed to provide a multidimensional view of the relationship between key hydrometeorological features and flood risk. In this visualization, rainfall was plotted against water level, with the flood risk level represented by color coding and discharge volume influencing the size of each data point. Additionally, hovering over individual points reveals contextual information such as the city location and associated river name, enhancing interpretability.



Figure 6: Interactive Flood Risk Analysis

In Figure [6], this interactive approach offered an intuitive exploration of how high rainfall and increased discharge correspond to elevated water levels and subsequently higher flood risk. It also allowed for easy identification of critical flood scenarios and regional variations across Pune, Nashik, Kolhapur, and Satara. Such a dynamic visualization tool proves valuable not only for researchers but also for policymakers and local authorities aiming to assess risk factors and prepare mitigation strategies in real time.

Interactive Flood Risk Analysis



ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538

Volume 13 Issue VI June 2025- Available at www.ijraset.com

D. Correlation Heatmap Analysis

To investigate the interdependence among various numerical features in the dataset, a correlation heatmap was generated.



Figure 7: Correlation Heatmap

In Figure [7], the heatmap displayed pairwise Pearson correlation coefficients between variables such as rainfall, temperature, humidity, water level, discharge, soil moisture, and others. Using a vibrant "coolwarm" color scale, the visualization highlighted both positive and negative correlations, where warmer colors represented stronger positive relationships and cooler tones indicated negative associations. Notably, strong positive correlations were observed between rainfall and water level, and between discharge and water level, affirming their combined influence on flood risk. Conversely, features like temperature and pressure exhibited weaker or inverse correlations with water-related variables. This analytical step was essential for identifying dominant predictors, reducing redundancy, and improving the feature selection strategy for the machine learning models employed in flood risk classification.

E. Time Series Analysis of Water Level Trends

A time series analysis was performed to examine the temporal fluctuations in average water levels across the four study locations— Pune, Nashik, Kolhapur, and Satara. For each city, the water level data was grouped by date and plotted to visualize the trends over time.



Figure 8: Water Level Trends Over Time

In Figure [8], the resulting line graph revealed dynamic variations in river water levels, reflecting the influence of seasonal rainfall patterns, hydrological conditions, and regional topography. Notably, the graph highlighted recurring peaks during the monsoon months, consistent with the earlier seasonal rainfall analysis. Differences in the amplitude and frequency of water level surges among the cities also underscored the localized nature of flood risks. This temporal insight is critical for developing predictive models that are sensitive to time-dependent behaviors and for implementing location-specific flood preparedness strategies.



ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538 Volume 13 Issue VI June 2025- Available at www.ijraset.com

IV. RESULT AND CONCLUSION

The study validated the effectiveness of machine learning models in predicting flood risk across Pune, Nashik, Kolhapur, and Satara using hydrometeorological data. Among the models tested, Random Forest and XG-Boost delivered the highest accuracy and robustness, effectively handling complex feature interactions. ANN showed good performance with larger datasets.

A Flask-based GUI enabled real-time flood risk predictions, while visual tools like correlation maps and time series plots enhanced model interpretability. Overall, the integration of machine learning, real-time interaction, and data visualization offers a scalable and practical solution for flood risk management and early warning systems.

A. Model Training and Evaluation

To build a robust flood risk prediction system, four different machine learning algorithms were trained and evaluated: Random Forest, Support Vector Machine (SVM), XGBoost, and a Deep Learning model based on Artificial Neural Networks (ANN). Each model was designed to classify flood risk into three categories: *Low*, *Medium*, and *High*, based on a combination of hydrometeorological input features.

The Random Forest model leveraged ensemble learning to build multiple decision trees, thereby reducing overfitting and improving prediction accuracy. The SVM classifier, with an RBF kernel, was utilized to construct high-dimensional decision boundaries for classifying complex, non-linear patterns. XGBoost, known for its speed and performance, was configured with gradient boosting to optimize classification accuracy.

Lastly, the ANN model was constructed using a multi-layer architecture with ReLU activation in the hidden layers and a softmax function in the output layer. It was trained using the Adam optimizer and sparse categorical cross-entropy as the loss function.

Each model was trained on the same training dataset and tested using a common test set. Their performance was evaluated primarily based on classification accuracy. The best-performing model—based on the highest accuracy on the test set—was automatically saved for deployment. The ANN model was saved in .h5 format, while the traditional models were stored using .pkl files via the joblib library.

This rigorous training and evaluation approach ensured that the final system incorporated the most accurate and generalizable model for predicting flood risk levels.

Model	Accuracy (%)
Random Forest	100.00
Support Vector Machine (SVM)	100.00
XGBoost	100.00
Deep Learning (ANN)	100.00

Table No. 1: Model Training and Evaluation

B. Model Performance Comparison

To assess the predictive capabilities of the proposed flood risk detection framework, four machine learning models—Random Forest, Support Vector Machine (SVM), XGBoost, and Deep Learning (ANN)—were evaluated using classification accuracy as the primary metric. After training and testing each model on the same dataset, their individual performance scores were collected and plotted to visualize comparative effectiveness.

The resulting bar chart (Figure X) illustrates the accuracy achieved by each model. The Random Forest and XGBoost classifiers demonstrated high predictive performance due to their ability to capture non-linear relationships and handle feature importance effectively. The SVM classifier, while slightly less accurate, still provided consistent results on the structured





International Journal for Research in Applied Science & Engineering Technology (IJRASET) ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538

Volume 13 Issue VI June 2025- Available at www.ijraset.com

dataset. The Artificial Neural Network (ANN) model also performed competitively, particularly benefiting from its multi-layer architecture that allowed it to learn complex patterns in the data.

This comparative analysis highlights the relative strengths of ensemble and deep learning approaches in flood risk prediction. The most accurate model was selected and stored for integration into the system's prediction pipeline. The graphical representation not only validates the effectiveness of the models used but also aids in the transparent selection of the optimal model for deployment.

C. Performance Analysis and Visualization

The proposed flood risk prediction system was evaluated using multiple machine learning and deep learning models, including Random Forest, Support Vector Machine (SVM), XGBoost, and a Deep Learning-based Artificial Neural Network (ANN). Each model was trained and tested on standardized environmental and hydrological features to ensure fair evaluation.

The results demonstrated that all models performed well, with the ANN and tree-based models (Random Forest and XGBoost) showing particularly strong accuracy and generalization. The models were assessed using various metrics such as accuracy, classification reports, and confusion matrices, which provided detailed insights into the models' ability to correctly classify different levels of flood risk (low, moderate, high).

Additionally, ROC curves and AUC scores were used for binary class scenarios to further evaluate model performance. Among all, the best-performing model was saved automatically along with the corresponding encoders and scaler for real-time deployment. Overall, the experimental results confirm the effectiveness and practicality of using AI-driven models for accurate flood risk prediction, offering a reliable tool for early warning systems and proactive disaster management.

D. Random Forest Classification Report

Class	Precision	Recall	F1-	Support
			Score	
High	1.00	1.00	1.00	287
Low	1.00	1.00	1.00	371
Mediu	1.00	1.00	1.00	219
m				
Accura			1.00	877
су				
Macro	1.00	1.00	1.00	877
Avg				
Weight	1.00	1.00	1.00	877
ed Avg				

Table No. 2: Random Forest Classification Report



Figure 10: Random Forest - Confusion Matrix



E. SVM Classification Report

Class	Precision	Recall	F1-Score	Support
High	1.00	1.00	1.00	287
Low	1.00	1.00	1.00	371
Medium	1.00	1.00	1.00	219
Accuracy			1.00	877
Macro Avg	1.00	1.00	1.00	877
Weighted Avg	1.00	1.00	1.00	877

Table No. 3: SVM Classification Report



Figure 11: SVM - Confusion Matrix

F. Deep Learning (ANN) Classification Report

Class	Precision	Recall	F1-Score	Support
High	1.00	1.00	1.00	287
Low	1.00	1.00	1.00	371
Medium	1.00	1.00	1.00	219
Accuracy			1.00	877
Macro Avg	1.00	1.00	1.00	877
Weighted Avg	1.00	1.00	1.00	877

Table No. 4: Deep Learning (ANN)



Figure 12: Deep Learning (ANN) - Confusion Matrix



ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538 Volume 13 Issue VI June 2025- Available at www.ijraset.com

G. Graphical User Interface (GUI)

The Graphical User Interface (GUI) for the flood risk prediction system was developed using Flask, HTML, CSS, and JavaScript to provide a user-friendly and interactive platform for end-users. The primary objective of the GUI is to simplify the interaction between the user and the machine learning models by allowing real-time data input and immediate prediction results. Through the web interface, users can input environmental and hydrological parameters such as location, river name, rainfall (mm), temperature (°C), humidity (%), wind speed (km/h), water level (m), discharge (m³/s), groundwater level (m), soil moisture (%), and atmospheric pressure (hPa). The categorical inputs like 'location' and 'river name' are handled through dropdown menus pre-encoded with label encoders to maintain consistency with the trained models.

← → C (0 127.0.0.15	000;home		# 19 1 문 🔕 :
FloodGuard Pro Hom			onkarsonawane2003@gmail.com •
NAVIGATION	Flood Prediction Dash	nboard	Live Temperature: 26.99°C
/ Manual Prediction	Logged in successfully!	-	×
Batch Prediction	A Manual Input Prediction		
ACCOUNT			River Name:
옷 Profile	Choose location		Choose river
[→ Logout	Rainfall (mm):	Temperature ("C):	Humidity (%):
			eg, 70
1	Wind Speed (km/h):	Water Level (m):	Discharge (m ¹ /s):
			eg, 150
	Groundwater Level (m):	Soil Moisture (%):	Pressure (hPa):Activate Windows
1			Go to Settings to activate Windows. e.g., 1012

Once the user submits the form, the input data is preprocessed using saved encoders and a standard scaler, ensuring that the data format matches the model's training conditions. The system then loads the best-performing model—either a traditional machine learning model stored in .pkl format or a deep learning model saved in .h5 format—and provides a real-time prediction of the flood risk level. The output is presented in a clear and color-coded format, indicating the risk category (e.g., low, moderate, or high).

loodGuard Pro Home				
🖉 Manual	Groundwater Level (m):	Soil Moisture (%):		
Prediction				
Batch Prediction			Clear Form	4 Predict Risk
0.0.0				
(-) Logout	Batch Prediction (Uploan	d CSV)		
	Select CSV File:			
	Choose File No file chosen			
			b 14	load & Predict Batch
		where the softward from all physics in some of	Go to Si	ettings to activate Windows.

Additionally, the GUI includes navigation options to access visual analytics such as model accuracy comparisons, confusion matrices, and correlation heatmaps. Overall, the GUI acts as a vital interface that bridges technical model deployment and practical usability, making the flood risk prediction system accessible to researchers, policymakers, and disaster management professionals.

V. FUTURE SCOPE

The developed flood risk prediction system shows strong potential for future expansion and real-world use. It can be enhanced through real-time data integration from weather stations, IoT devices, and satellite imagery for continuous monitoring. Adding GIS data and multilingual support can improve spatial accuracy and user accessibility.

Technically, advanced deep learning models like RNNs, LSTMs, and CNNs can further improve time-series and spatial analysis. Integration with disaster management systems and continuous model refinement using real-world feedback will boost accuracy and impact. With institutional support, this AI-driven system can become a vital tool for flood preparedness and disaster resilience.



ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538 Volume 13 Issue VI June 2025- Available at www.ijraset.com

VI. CONCLUSION

The study presents the successful development of an AI-based flood risk prediction system using machine learning and deep learning models. Utilizing a rich dataset of environmental and hydrological parameters—such as rainfall, temperature, humidity, river discharge, and groundwater level—the system effectively classifies flood risk levels across multiple cities. Among the models tested (Random Forest, SVM, XG-Boost, and ANN), the Artificial Neural Network (ANN) achieved the highest accuracy, demonstrating its strength in capturing complex nonlinear patterns. Model performance was validated using metrics like accuracy, classification reports, and confusion matrices. The system also incorporated time series analysis to uncover water level trends, offering deeper insights into flood dynamics. A user-friendly GUI was developed to enable real-time interaction with model outputs, enhancing accessibility and practical application. This intelligent, data-driven approach holds significant promise for improving early warning systems and supporting disaster mitigation efforts in both urban and rural settings.

REFERENCES

- X. Wang, G. Huang, and J. Liu, Projected increases in intensity and frequency of rainfall extremes through a regional climate model ing approach, J. Geophys. Res., vol. 119, no. 23, pp. 13271 13286, Dec. 2014.
- [2] Climate Change 2014 Synthesis Report. Accessed: Jun. 6, 2020. [Online]. Available:<u>https://archive.ipcc.ch/pdf/assessment-report/ar5/syr/SYR_AR5_FINAL_full_zh.pdf</u>
- [3] L. Ling, Z. Yusop, and M. F. Chow, Urban ood depth estimate with a new calibrated curve number runoff prediction model, IEEE Access, vol. 8, pp. 1091510923, Jan. 2020.
- [4] M. Nayeb Yazdi, M. Ketabchy, D. J. Sample, D. Scott, and H. Liao, An evaluation of HSPF and SWMM for simulating streamow regimes in an urban watershed, Environ. Model. Softw., vol. 118, pp. 211225, Aug. 2019.
- [5] H. C. Winsemius, J. C. J. H. Aerts, L. P. H. van Beek, M. F. P. Bierkens, A. Bouwman, B. Jongman, J. C. J. Kwadijk, W. Ligtvoet, P. L. Lucas, D. P. van Vuuren, and P. J. Ward, Global drivers of future river ood risk, Nature Climate Change, vol. 6, no. 4, pp. 381385, Dec. 2015.
- [6] L. Al eri, L. Feyen, and G. Di Baldassarre, Increasing ood risk under climate change: A pan-European assessment of the bene ts of four adap tation strategies, Climatic Change, vol. 136, nos. 34, pp. 507521, Mar. 2016.
- [7] O. M. Rezende, A. B. R. da Cruz de Franco, A. K. B. de Oliveira, A. C. P. Jacob, and M. G. Miguez, Aframeworkto introduce urban ood resilience into the design of oodcontrolalternatives, J. Hydrol., vol. 576, pp. 478493, Sep. 2019.
- [8] H. Wang, C. Mei, J. Liu, and W. Shao, A new strategy for integrated urban water management in China: Sponge city, Sci. China Technol. Sci., vol. 61, no. 3, pp. 317329, Jan. 2018.
- [9] U. C. Nkwunonwo, M. Whitworth, and B. Baily, A review of the current status of ood modelling for urban ood risk management in the develop ing countries, Sci. Afr., vol. 7, Mar. 2020, Art. no. e00269.
- [10] N. Singh and S. R. Mohanty, Short term price forecasting using adaptive generalized neuron model, Int. J. Ambient Comput. Intell., vol. 9, no. 3, pp. 419428, Jul. 2018.
- [11] S. Gupta, M. Khosravy, N. Gupta, H. Darbari, and N. Patel, Hydraulic system onboard monitoring and fault diagnostic in agricultural machine, Brazilian Arch. Biol. Technol., vol. 62, pp. 16784324, Oct. 2019.
- [12] A.Sorjamaa, J.Hao, N.Reyhani, Y.Ji, and A.Lendasse, Methodologyfor long-term prediction of time series, Neurocomputing, vol. 70, nos. 1618, pp. 28612869, Oct. 2007.
- [13] C. Hu, L. Bi, Z. Piao, C. Wen, and L. Hou, Coordinative optimization control of microgrid based on model predictive control, Int. J. Ambient Comput. Intell., vol. 9, no. 3, pp. 5775, Jul. 2018.
- [14] C.Ding, X.Cao, and C.Liu, Howdoesthestation-area built environment in uence metrorail ridership? Using gradient boosting decision trees to identify non-linear thresholds, J. Transp. Geography, vol. 77, pp. 7078, May 2019.
- [15] Z. Wu, Y. Zhou, H. Wang, and Z. Jiang, Depth prediction of urban ood under different rainfall return periods based on deep learning and data warehouse, Sci. Total Environ., vol. 716, May 2020, Art. no. 137077.











45.98



IMPACT FACTOR: 7.129







INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 🕓 (24*7 Support on Whatsapp)