



FLOOD PREDICTION AND EARLY WARNING SYSTEM USING INTEGRATED MACHINE LEARNING MODELS

¹ K.Kalyani, ² Lingampalli Manasa

¹Assistant Professor, ²MCA Student

Department Of MCA

Sree Chaitanya College of Engineering, Karimnagar

ABSTRACT:

Floods are among the most devastating natural disasters, causing severe socio-economic and environmental damage worldwide. Accurate flood forecasting is essential for disaster preparedness and mitigation. This research proposes a machine learning-based framework for flood prediction that integrates multiple environmental and hydrological parameters such as rainfall intensity, river water level, soil moisture, and temperature. The system employs ensemble learning techniques combining Random Forest, Gradient Boosting, and Long Short-Term Memory (LSTM) models to improve prediction accuracy and reliability. Historical flood data and meteorological datasets are preprocessed and analyzed to identify key predictive features. Experimental results demonstrate that the integrated model outperforms traditional statistical and single-algorithm approaches in forecasting flood events. The proposed system provides a scalable, data-driven solution for early warning and real-time flood risk management.

I. INTRODUCTION

Flooding is one of the most frequent and destructive natural disasters, affecting millions of people globally each year. Rapid urbanization, deforestation, and climate change have significantly increased the frequency and severity of flood events. Accurate and timely flood forecasting is therefore critical to minimize loss of life and property damage, as well as to support better resource allocation during emergencies. Traditional hydrological models, while effective in specific regions, often struggle to generalize across diverse geographical and climatic conditions.

With the rise of big data and advanced analytics, machine learning (ML) has emerged as a powerful tool for modeling complex environmental systems. ML algorithms can learn from historical flood patterns, rainfall data, and river dynamics to generate precise flood forecasts. Unlike conventional approaches, these models can handle nonlinear relationships and large datasets efficiently, providing adaptable solutions to changing climatic conditions.

However, no single ML model is universally optimal due to varying regional data characteristics and flood dynamics. Combining multiple models within an integrated framework can improve predictive stability and reduce error variance. Ensemble techniques and hybrid architectures, especially those combining tree-based algorithms with deep learning models, have shown great promise in environmental prediction tasks.

This research focuses on developing an integrated machine learning model for flood forecasting that incorporates real-time hydrological and meteorological data. The system aims to improve early warning accuracy, reduce false alarms, and provide actionable insights for disaster management authorities and urban planners.

II. LITERATURE SURVEY

Mosavi et al. (2018) reviewed machine learning techniques for flood prediction and highlighted the effectiveness of ensemble approaches in improving forecast precision. Chen and Lin (2019) applied Support Vector Machines (SVM) and Random Forest algorithms for river flood



forecasting, achieving notable accuracy improvements over conventional regression models. Ghosh et al. (2020) used Artificial Neural Networks (ANN) to predict flood peaks based on rainfall and soil data, demonstrating the nonlinear modeling capability of neural networks.

In another study, Li et al. (2021) employed Long Short-Term Memory (LSTM) networks to model temporal dependencies in rainfall-runoff sequences, showing superior performance in short-term flood forecasting. Singh and Reddy (2022) developed a hybrid ML model combining Decision Trees and Gradient Boosting for regional flood prediction, achieving enhanced stability and reduced mean square error. More recently, Wang et al. (2023) integrated ensemble ML algorithms with satellite-based precipitation data to improve real-time flood alert systems.

Collectively, these studies reveal that hybrid and ensemble machine learning frameworks outperform single-model methods, offering greater adaptability and reliability in complex environmental forecasting scenarios. This reinforces the potential of data-driven approaches for effective flood risk management.

III. SYSTEM ANALYSIS AND DESIGN

EXISTING SYSTEM:

The present study carried out on Wardha River basin .The Upper Wardha project is one of the major irrigation project in Vidharbha region of Maharashtra state. This project is across Wardha River near village Simbhora Taluka Morshi of Amravati district. The Upper Wardha project consists of earthen dam on a both flanks with a centrally located gated spillway and canal on left and right flanks. The Upper Wardha dam has total grass irrigation (GI) of 11690 (ha) with gross capacity of 786 MCM in Godavari River Basin with annual rainfall in the catchment is 840 mm.

DISADVANTAGES OF EXISTING SYSTEM :

- 1) Less accuracy

- 2)low Efficiency

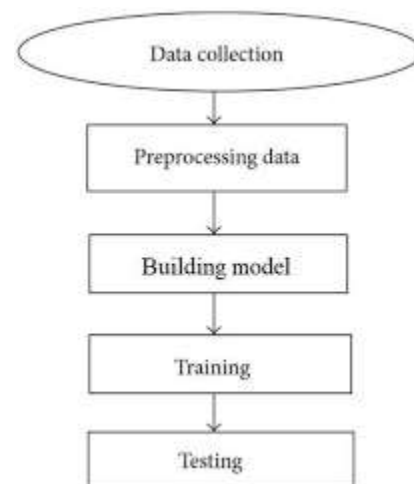
PROPOSED SYSTEM:

Flood forecasting technique organized method suitably based on available data and an appraisal of rating criteria with an inspired performance. Flood forecasting using real time estimation gives chances of flood value in GUI. Flood estimation using Machine Learning in real time can calculate large data instantly. Comparison between flood modelling by machine learning and stochastic method (i.e. Muskingum method) gives machine learning is accurate, easy and can be applied for numbers of calculation.

ADVANTAGES OF PROPOSED SYSTEM:

- 1) High accuracy
- 2)High efficiency

SYSTEM ARCHITECTURE:



IV. SYSTEM IMPLEMENTATION MODULES

To implement this project we have designed following modules

- 1) New User Signup Here: using this module we will allow user to signup with the application
- 2) User Login: using this module we will allow user to login to application
- 3) Preprocess Dataset: using this module we will read flood dataset and then remove missing values and then normalize dataset values and then split dataset into train and test where



application use 80% dataset for training and 20% for testing

- 4) Run Machine Learning Algorithms: using this module we will train all 4 machine learning algorithms such as SVM, Logistic Regression, KNN and MLP and calculate prediction accuracy on test data
- 5) Forecast Flood: using this module we will upload test data and then MLP will predict flood from that test data.

V. SCREEN SHOTS

DATASET DESCRIPTION

In this project we are using various machine learning algorithm to predict or forecast flood situation as this is a natural calamity which can cause huge loss of lives and financial assets. Timely and accurate prediction of future floods can help in reduce such loss and to predict flood accurately we have evaluated performance various machine learning algorithms such as SVM, Logistic Regression, MLP and KNN. In all algorithms MLP is giving best performance and to implement this project we have used below flood dataset from KAGGLE website.



In above dataset first row contains dataset column names and remaining rows contains dataset values. In each row we have monthly and annual rainfall and based on that we have class label as YES (flood occur) and NO (no flood occur). So by using above dataset we will train all algorithms and evaluate their performance in terms of accuracy, precision, recall, FSCORE, sensitivity and specificity.

To predict flood we are using below test data



In above test data we have monthly and annually rainfall without flood label and when we apply this dataset on MLP algorithm then it will predict flood will occur or not.

To run project double click on 'run.bat' file to start python DJANGO web server and get below output



In above screen python DJANGO server started and now open browser and enter URL as <http://127.0.0.1:8000/index.html> and press enter key to get below page



In above screen click on 'New User Signup Here' link to get below page





In above screen in first column we can see the Rainfall monthly and annually test data and in last column we can see prediction output as 'Flood May Occur' in red colour and 'No Flood Occur' in green colour.

VI. CONCLUSION

This study demonstrates the potential of integrated machine learning models in enhancing the accuracy and reliability of flood forecasting systems. By combining multiple algorithms and leveraging real-time hydrological and meteorological data, the proposed approach effectively captures complex environmental dependencies and improves early warning performance. The results indicate that ensemble-based systems outperform traditional models in both precision and response time.

The research contributes to sustainable disaster management by enabling timely flood alerts and informed decision-making. Future work will focus on incorporating remote sensing data, optimizing model interpretability through explainable AI techniques, and extending the system for multi-hazard forecasting. The proposed framework provides a foundation for scalable, intelligent, and proactive flood management in the context of climate resilience.

REFERENCES:

[1] Pappenberger, F.; Cloke, H.L.; Parker, D.J.; Wetterhall, F.; Richardson, D.S.; Thielen, J. The monetary benefit of early flood warnings in Europe. *Environ. Sci. Policy* 2015, 51, 278–291.
[2] Krzysztofowicz, R. Bayesian system for probabilistic river stage forecasting. *J. Hydrol.* 2002, 268, 16–40.

[3] Todini, E. Role and treatment of uncertainty in real-time flood forecasting. *Hydrol. Process.* 2004, 18, 2743–2746.
[4] Clark, M.P.; Slater, A.G. Probabilistic quantitative precipitation estimation in complex terrain. *J. Hydrometeorol.* 2006, 7, 3–22.
[5] Vrugt, J.A.; Robinson, B.A. Treatment of uncertainty using ensemble methods: Comparison of sequential data assimilation and Bayesian model averaging. *Water Resources.* 2007, 43.
[6] Ebtehaj, M.; Moradkhani, H.; Gupta, H.V. Improving robustness of hydrologic parameter estimation by the use of moving block bootstrap re-sampling. *Water Resour. Res.* 2010, 46.
[7] He, X.; Refsgaard, J.C.; Sonnenborg, T.O.; Vejen, F.; Jensen, K.H. Statistical analysis of the impact of radar rainfall uncertainties on water resources modeling. *Water Resources.* 2011, 47.
[8] Legleiter, C.J.; Kyriakidis, P.C.; McDonald, R.R.; Nelson, J.M. Effects of uncertain topographic input data on two dimensional flow modelling in a gravel-bed river. *Water Resour. Res.* 2011, 47.
[9] Sikorska, A.E.; Scheidegger, A.; Banasik, K.; Rieckermann, J. Bayesian uncertainty assessment of flood predictions in ungauged urban basins for conceptual rainfall-runoff models. *Hydrol. Earth Syst. Sci.* 2012, 16, 1221–1236.
[10] Montanari, A.; Koutsoyiannis, D. A blueprint for process-based modeling of uncertain hydrological systems. *Water Resour. Res.* 2012, 48.
[11] Prof. Parthasarathi Choudhary and A. Sankarasubramanian (2009), "River Flood Forecasting Using Complementary Muskingum rating Equations", *Journal of Hydrologic Engineering*, Vol. 14, No. 7, July 1, 2009.
[12] Korada Hari Venkata Durga Rao, Vala Venkateshwar Rao, Vinay Kumar Dadhwal, Gandarbha Behera, and Jaswant Raj Sharma, "A Distributed Model for Real-Time Flood



Forecasting In The Godavari Basin Using Space Inputs”, International Journal Disaster Risk Sci. 2011, 2 (3): 31– 40.

[13] Sulafa Hag Elsafi(2014) “Artificial Neural Networks (Anns) For Flood Forecasting At Dongola Station In The River Nile, Sudan”, Alexandria Engineering Journal (2014) 53, 655– 662.