JOURNAL LA MULTIAPP

VOL. 05, ISSUE 06 (890-897), 2024
DOI: 10.37899/journallamultiapp.v5i6.1520

Advancing Flood predication based on integrating Traditional and Deep Learning Techniques

Muthana S. Mahdi[1], **Ishraq Khudhair Abbas**[1], **Ban M. Muhammed**[2], **Hiba Abdulrazzaq Abbas**[1]

[1]*Department of Computer Science, College of Science, Mustansiriyah University, Baghdad, Iraq*
[2]*Iraqi Commission for Computers and Informatics (ICCI), Baghdad, Iraq*

**Corresponding Author: Muthana S. Mahdi*
Email: muthanasalih@uomustansiriyah.edu.iq

Article Info

Article history:
Received 1 August 2024
Received in revised form 30 August 2024
Accepted 28 December 2024

Keywords:
Floods
Mean Squared Error (MSE)
Root Mean Squared Error (RMSE)
and R-squared (R^2) hybrid
Machine Learning (ML)
DNNs

Abstract

Flood prediction is critical for reducing the negative effects of catastrophes. This study investigates the use of classical and deep learning algorithms to accurately anticipate flood levels using past ecological information. The study takes two approaches: linear regression and deep neural networks (DNN). These models are trained and evaluated using historical records stretching years, as well as moisture, temperature, and rainfall measurements. Visualization tools, such as development/validation reduction charts projected against real floodwater level graphs, can help understand model learning behaviors and efficacy. Furthermore, a combination model method is investigated, which combines forecasts from both approaches and has the promise to improve the accuracy of predictions. Future forecasts for the year 2025 can be generated using expected atmospheric conditions, illustrating the models' usefulness in projecting eventual floodwater levels. Measures of assessment and visualization results demonstrate the effectiveness of deep learning technologies in improving flooding predictions when contrasted with classic conventional techniques. This investigation advances flooding forecasting skills by combining known statistical approaches and new deep learning algorithms, thereby helping in active catastrophe control and vulnerability planning.

Introduction

Floods pose enormous hazards to ecosystems and towns worldwide, emphasizing the crucial importance of timely and precise forecasting systems. Traditionally, flood forecasting has depended on statistical models like Linear Regression, which uses past information on storms, the flow of rivers, and surrounding factors to determine prospective flood waters (Duan et al., 2006; Madsen et al., 2014). These approaches bring beneficial insights into flood circumstances, but they may fail to identify nonlinear correlations and complicated linkages in the data. As DNNs grow more popular in applications that are mission-critical, maintaining their consistent operation becomes vital. Also, it can solve the complex problem. Traditional resilience solutions fail to compensate for the particular characteristics of DNN algorithms/accelerators, rendering them infeasible or ineffectual (Mittal, 2020; Parashar et al., 2019). DNNs have recently transformed forecasting by allowing complex designs to be retrieved from huge amounts of data (Chollet, 2017). DNNs excel at learning tiered structures for information, making them ideal for applications requiring nonlinear correlations and

ISSN: 2716-3865 (Print), 2721-1290 (Online)

component interactions, such as flood predictions (Mosavi et al., 2018; Sit et al., 2020; Dtissibe et al., 2024).

This study investigates and compares the efficacy of classic methods of statistics and current deep-learning techniques in predicting floods. The present investigation aims to assess the forecasting capacity of these algorithms through various situations and circumstances by using complete datasets that include previous records of the process of precipitation temperature, humidity, and river flow dynamics, as well as advances in mathematical methods (Yosri et al., 2024; Herath et al., 2023; Li & Jun, 2024; Karim et al., 2023).

The primary goals involve the following: a) Assessment of analytical techniques: Evaluating the accuracy and efficacy of Linear Regression models for catching historical flood dynamics and forecasting future flood levels; b) Application of Deep Learning: Implementing DNN architectures, such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), can capitalize on timing and geographical connections in data on the environment for improved forecasting precision; c) Hybrid Modeling Approaches: Investigating hybrid models that combine the strengths of statistical and deep learning methods to capitalize on both historical patterns and complicated data interactions; d) The present research aims to improve flood projection features and notify resilient approaches to disaster mitigation by using accurate assessments such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared (R^2), as well as graphics of predicted instead of reality levels of a flood.

After this introduction, the rest of the research includes related works in the second section, the research methodology in the third section, the results in the fourth section, and we end with the conclusions in the fifth section.

Related works

There are many related works: -

In Habibi et al. (2023) This paper presents hybrid Machine Learning (ML) models that combine ensemble techniques with Feature Selection (FS) algorithms for Flood Hazard (FH) prediction, which is critical for managing destructive natural disasters such as floods. The ideal Flood Influential Factors (FIFs) are identified using the Simulated Annealing (SA) and Information Gain (IG) FS algorithms. Random search (RS) and repeated cross-validation are used to improve group machine learning (ML) models such as AdaboostM1 (ABM), Boosted Generalized Linear Model (BGLM), and Stochastic Gradient Boosting. Rainfall, distance to river, height, and limestone are identified as key FIFs in the Sardabroud watershed, Iran. The SA-ABM model has the greatest AUC value (0.983) in ROC analysis, indicating 27% of the area with high flood hazard.

In Habibi et al. (2023). This study presents unique hybrid Machine Learning (ML) models for enhancing flash flood susceptibility (FFS) mapping in the Neka-Haraz watershed in Iran. It combines ensemble ML models (BAFDA, XGB, ROF, and BGAM) with wrapper-based factor optimization methods (RFE, Boruta) to meta-optimize super-parameters using Random Search (RS). According to ROC AUC and efficiency measures, BGAM-Boruta performs best (AUC = 0.953, Efficiency = 0.910). FFS is influenced by several elements, including distance to river, slope, rainfall, altitude, and distance to road. The models demonstrate substantial potential (up to 46% coverage) for detecting high-risk FFS areas, which can aid in effective disaster management methods.

In Anaraki et al. (2023). The present investigation uses advanced combination models (GLMBoost, RF, and BayesGLM) to improve flash-flood risk estimation in Iran, addressing the critical requirement for accurate modeling in the face of frequent natural disasters. The models obtain great performance by using simulated annealing for feature selection (accuracy

ISSN: 2716-3865 (Print), 2721-1290 (Online)

= 90-92%, Kappa = 79-84%, Success ratio = 94-96%, Threat score = 80-84%, and Heidke skill score = 79-84%). Distance from downloads, vegetation, the amount of drainage, use of land, and slope all have a substantial impact on the models, allowing for more effective hazard mapping and flood mitigation measures in data-scarce places.

In Hosseini et al. (2020). This paper presents a novel methodology for analyzing flood frequency under climate change scenarios (HadCM3, CGCM3, CanESM 2). Precipitation is classified using Multivariate Adaptive Regression Splines (MARS) and an M5 Model Tree, while the Whale Optimization Approach (WOA) trains a Least Square Support Vector Machine to model it. The Wavelet Transform (WT) separates the temperatures and precipitation, and different algorithms (LSSVM-WOA, LSSVM, KNN, ANN) reduce temperature as well as precipitation. The discharge simulation covers the present, near future, and far future eras, with flood frequency calculations finding a lower 200-year discharge across all scenarios. Uncertainty study using ANOVA and fuzzy approaches highlights hydrological model uncertainties, with HadCM3 demonstrating decreased uncertainty during high return periods.

Methods

The Linear Regression algorithm uses scikit-learn's linear regression to generate a baseline forecast based on fundamental conditions. At that point, the Deep Neural Network approach, built with Keras and TensorFlow, uses many complex layers with reduced linear unit (ReLU) events and dropout for legalization. Models are trained on a subset of data and verified on rest to determine predicted performance using MSE and R-squared (R^2). Figure 1 illustrates the d]general data flow diagram of proposed system.

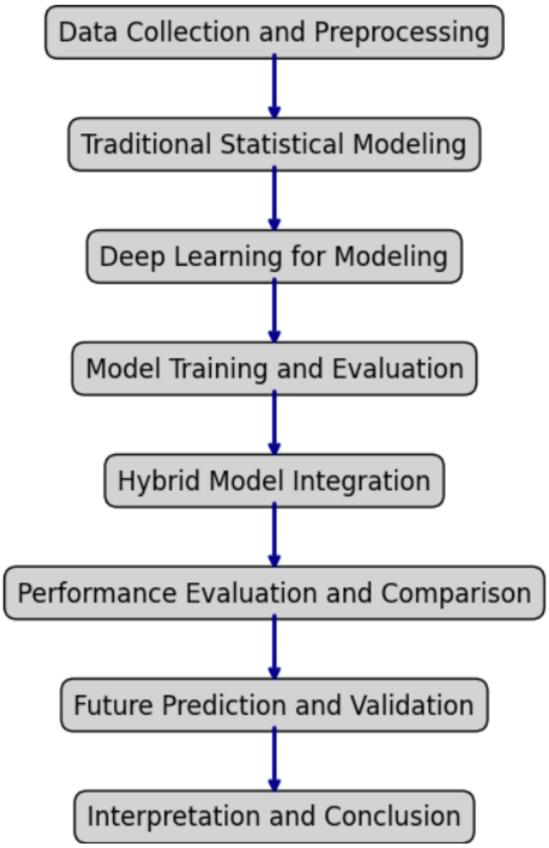

Figure 1. Data flow diagram of proposed system

The general steps are:-

ISSN: 2716-3865 (Print), 2721-1290 (Online)

Stage 1. Data Collection and Preprocessing

The study begins with the collection of extensive datasets that include historical records of environmental factors important for flood prediction. These variables usually include: a) Weather information includes annual rainfall, relative humidity, temperature, and velocity of wind records; b) Hydrological statistics include discharge from rivers, the amount of water, & the rate of flow; c) Data preparation includes cleaning up information eliminating outliers, addressing lost numbers, and assuring data uniformity; d) Feature selection is the process of identifying key characteristics that have a substantial impact on flood dynamics; d) Normalization involves scaling computational characteristics to guarantee homogeneity and improve model performance.

Stage 2. Traditional Statistical Modeling

Linear Regression: Uses scikit-learn's LinearRegression to estimate the link amongst prior atmospheric variables (such as rainfall and temp) and the extent of flooding. Uses characteristics engineering to choose appropriate markers and build a model of linearity based on past patterns.

Stage 3. Deep Learning for Modeling

Establishes a model that is sequential employing Keras and TensorFlow as the backend. Creates multi-layered patterns with dense layers, using functions of nonlinear activation like ReLU to capture intricate connections. Uses regularization for dropouts to avoid overfitting and improve applicability.

Stage 4. Model Training and Evaluation

The dataset is divided into subsets to be trained and tested using cross-validations techniques. The efficiency of the model is evaluated using measures like MSE, RMSE, and R-squared (R^2). Shows anticipated and real flood tides to evaluate model reliability and preciseness. Separates the dataset across testing, validation, and training sets for the DNN model. Monitoring training progress using validation loss to avoid overfitting and optimize hyperparameters.

Stage 5. Hybrid model integration

Investigates hybrid modeling techniques that combine forecasts from Linear Regression and DNN models. Outputs are averaged or combined to make use of each model type's advantages and improve overall prediction precision.

Stage 6. Performance evaluation and comparison

Each model's efficacy is evaluated through quantitatively specified metrics. Conducts statistical investigations to evaluate the predicting skills of traditional statistical approaches versus advanced machine learning techniques. Creates detailed visualizations, such as loss curves, scatter graphs of anticipated versus actual results, and comparison analyses, to successfully assess and explain discoveries.

Stage 7. Future Prediction and Validation

Flood forecasts are simulated for possible futures (e.g., 2025) using expected weather conditions. Model stability and reliability are validated via sensitivity analysis and scenario testing. Continually improves models depending on feedback along with performance data.

Stage 8. Interpretation and conclusion

Outlines the major results of the study, insights, and consequences. The suggestions for flood models for prediction are discussed, including their merits, boundaries, and possible applications.

ISSN: 2716-3865 (Print), 2721-1290 (Online)

Makes suggestions for improving flood predictions by combining statistical and machine learning technologies.

Results and Discussion

In the context of machine learning, an "epoch" is defined as one full trip over the whole set of data utilized in training. Throughout each epoch, the algorithm is developed on chunks of data, and its success indicators, such as the loss function, are assessed.

Each epoch is a cycle in which the model learns from the data and modifies its weights in order to reduce the coefficient of loss. The objective was to train the algorithms until the error function merges to the least, demonstrating that it is capable of recognizing the basic patterns in the data. The assessment of loss is important because it monitors the model's function on unknown information, thereby minimizing overfitting.

As example:-

Epoch 1/150: The model started with a training loss of 2529111.0000 and a validation loss of 1374155.3750 after processing 4 batches.

Epoch 2/150: The model's training loss increased to 2613328.0000, but the validation loss decreased to 1253333.2500.

Epoch 3/150: The training loss decreased to 2401501.2500, and the validation loss continued to decrease to 1145032.3750.

Epoch 4/150: Both training and validation losses decreased further.

Linear Regression - MSE: 342678.6358668617, R^2: -8.220144690470374

DNN - MSE: 80775.56675841555, R^2: -1.1733552518771475

Hybrid Model - MSE: 158831.95172309814, R^2: -3.273547934064216

Tabel I illustrates the Strategy of Epoch that can be used

Table 1. Description of Strategy of Epoch

Strategy	Description	Benefits
Concept of Epoch	A complete pass over the dataset during each training iteration, updating model weights based on calculated loss.	Improves model accuracy by iterating over the data and updating weights continuously.
Early Stopping	Stops training when performance on the validation set no longer improves after a certain number of epochs.	Prevents overfitting, improves prediction accuracy, and saves time and resources.
Performance Monitoring	Tracks metrics such as loss and accuracy to adjust the number of epochs based on performance.	Helps enhance performance and avoid underfitting or overfitting.
Cross-Validation	Evaluates the model across different data subsets to determine the optimal number of epochs.	Enhances generalization and reduces errors due to overfitting.

ISSN: 2716-3865 (Print), 2721-1290 (Online)

Figure 2 illustrates the prediction result

```
=== Predictions for 2025 ===
Expected floods in 2025 (Linear Regression): 1303.5520968114433 cubic meters
Expected floods in 2025 (DNN): 1427.3739013671875 cubic meters
Expected floods in 2025 (Hybrid Model): 1365.4629990893154 cubic meters
```

Figure 2. prediction results

As shown in Figure 2. The training and validation loss plots demonstrate a steady reduction in loss, indicating effective model formation and stable converging. The scatter figure contrasts projected and actual flood levels, with the majority of predictions being close to those of actual numbers but with a few differences. Overall, the model functions well, but precision should be improved.

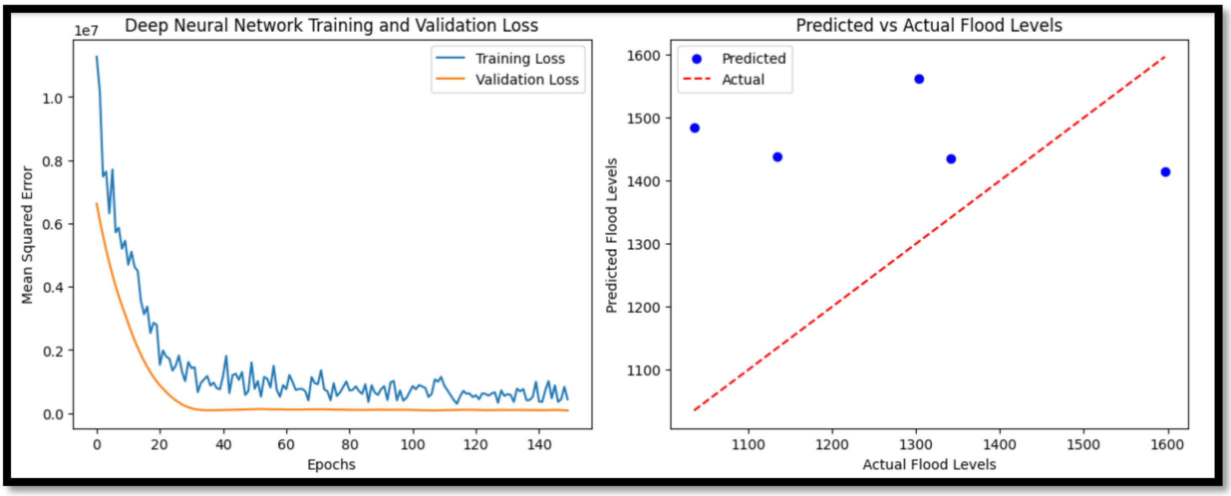

Figure 3. visualization results for training and an actual model.

Tabel 2 and table 3 illustrate the description of the figure 3

Table 2. Deep Neural Network Training and Validation Loss

Epochs	Training Loss (Mean Squared Error)	Validation Loss (Mean Squared Error)
0	High	High
~10	Significant decrease	Decrease
~50	Plateau	Slight decrease
~100	Small fluctuations around a low value	Stabilization
~140	Stabilized at low value	Stabilized at low value

Table 3. Predicted vs Actual Flood Levels

Actual Flood Levels	Predicted Flood Levels	Deviation from Actual (Predicted - Actual)
1100	~1400	Positive
1200	~1450	Positive
1300	~1350	Slightly Positive
1400	~1450	Positive
1500	~1500	0
1600	~1500	Negative

ISSN: 2716-3865 (Print), 2721-1290 (Online)

Conclusion

The study intended to forecast flood waters using an advanced neural network. Validation and training loss lines demonstrate that the system efficiently learned themes given the data since both losses dropped and steadied over epochs. Nevertheless, a scatter map comparing projected and actual flood levels shows differences, demonstrating that even when the model catches trends, its accuracy fluctuates for specific data points. This shows that the mathematical framework is a strong starting point, but it could benefit from more tuning, feature engineering, or importing more diverse data to improve prediction precision as well as reliability.

Acknowledgments

The authors thank the Department of Computer Science, College of Science, Mustansiriyah University, for supporting this work.

References

Anaraki, M. V., Farzin, S., Mousavi, S. F., & Karami, H. (2021). Uncertainty analysis of climate change impacts on flood frequency by using hybrid machine learning methods. *Water Resources Management*, *35*, 199-223. https://doi.org/10.1007/s11269-020-02719-w

Chollet, F. (2017). Deep Learning with Python. Manning Publications.

Dtissibe, F. Y., Ari, A. A. A., Abboubakar, H., Njoya, A. N., Mohamadou, A., & Thiare, O. (2024). A comparative study of Machine Learning and Deep Learning methods for flood forecasting in the Far-North region, Cameroon. *Scientific African*, *23*, e02053. https://doi.org/10.1016/j.sciaf.2023.e02053

Duan, Q., Sorooshian, S., & Gupta, V. K. (2006). Effective and Efficient Global Optimization for Conceptual Rainfall-Runoff Models. *Water Resources Research*, *32*(4), 101-110. https://doi.org/10.1029/96WR03180.

Habibi, A., Delavar, M. R., Nazari, B., Pirasteh, S., & Sadeghian, M. S. (2023). A novel approach for flood hazard assessment using hybridized ensemble models and feature selection algorithms. *International Journal of Applied Earth Observation and Geoinformation*, *122*, 103443. https://doi.org/10.1016/j.jag.2023.103443

Habibi, A., Delavar, M. R., Sadeghian, M. S., Nazari, B., & Pirasteh, S. (2023). A hybrid of ensemble machine learning models with RFE and Boruta wrapper-based algorithms for flash flood susceptibility assessment. *International Journal of Applied Earth Observation and Geoinformation*, *122*, 103401. https://doi.org/10.1016/j.jag.2023.103401

Herath, M., Jayathilaka, T., Hoshino, Y., & Rathnayake, U. (2023). Deep machine learning-based water level prediction model for Colombo flood detention area. *Applied Sciences*, *13*(4), 2194. https://doi.org/10.3390/app13042194

Hosseini, F. S., Choubin, B., Mosavi, A., Nabipour, N., Shamshirband, S., Darabi, H., & Haghighi, A. T. (2020). Flash-flood hazard assessment using ensembles and Bayesian-based machine learning models: Application of the simulated annealing feature selection method. *Science of the total environment*, *711*, 135161. https://doi.org/10.1016/j.scitotenv.2019.135161

Karim, F., Armin, M. A., Ahmedt-Aristizabal, D., Tychsen-Smith, L., & Petersson, L. (2023). A review of hydrodynamic and machine learning approaches for flood inundation modeling. *Water*, *15*(3), 566. https://doi.org/10.3390/w15030566

ISSN: 2716-3865 (Print), 2721-1290 (Online)

Li, L., & Jun, K. S. (2024). Review of machine learning methods for river flood routing. *Water*, *16*(2), 364. https://doi.org/10.3390/w16020364

Madsen, H., Lawrence, D., Lang, M., Martinkova, M., & Kjeldsen, T. R. (2014). Review of trend analysis and climate change projections of extreme precipitation and floods in Europe. *Journal of Hydrology*, *519*, 3634-3650. https://doi.org/10.1016/j.jhydrol.2014.11.003

Mittal, S. (2020). A survey on modeling and improving reliability of DNN algorithms and accelerators. *Journal of Systems Architecture*, *104*, 101689.

Mosavi, A., Ozturk, P., & Chau, K. W. (2018). Flood prediction using machine learning models: Literature review. *Water*, *10*(11), 1536. https://doi.org/10.3390/w10111536

Parashar, A., Raina, P., Shao, Y. S., Chen, Y. H., Ying, V. A., Mukkara, A., ... & Emer, J. (2019, March). Timeloop: A systematic approach to dnn accelerator evaluation. In *2019 IEEE international symposium on performance analysis of systems and software (ISPASS)* (pp. 304-315). IEEE. https://doi.org/10.1109/ISPASS.2019.00042

Sit, M., Demiray, B. Z., Xiang, Z., Ewing, G. J., Sermet, Y., & Demir, I. (2020). A comprehensive review of deep learning applications in hydrology and water resources. *Water Science and Technology*, *82*(12), 2635-2670. https://doi.org/10.2166/wst.2020.369

Yosri, A., Ghaith, M., & El-Dakhakhni, W. (2024). Deep learning rapid flood risk predictions for climate resilience planning. *Journal of Hydrology*, *631*, 130817. https://doi.org/10.1016/j.jhydrol.2024.130817

ISSN: 2716-3865 (Print), 2721-1290 (Online)