JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

FLOOD FORECASTING IN UPPER MAHANADI RIVER BASIN USING DEEP LEARNING MODELS

[1]Sheelavathy K V, [2]Akhil C, [3]Abhishek Holla, [4] Ayesha Shifana, [5] Kushal Achar
[1]Professor, [2]Student, [3]Student, [4]Student, [5]Student
[1]Computer Science and Engineering,
[1]Reva University, Bengaluru, India

Abstract: This study presents the development and evaluation of an independent Artificial Neural Network, (ANN) model tailored for daily streamflow forecasting within the upper Mahanadi River basin, spanning a lead time of 1 to 3 days. Utilizing hydro- meteorological data from 2001 to 2011, including daily discharge measurements from five stations, inflow data from the Hirakud Reservoir, and spatial average rainfall, temperature data, the ANN model is meticulously calibrated and evaluated. The methodology involves a meticulous selection of input time steps (lags) for each variable, employing both correlation-based and constant lag methodologies. Performance evaluation system of measurement, such as Nash-Sutcliffe Efficiency (NSE), Ratio of root mean square error to standard deviation of measured data, (RSR), Mean Absolute Error, (MAE), and Error in volume, (Evol), are working to assess the model's predictive capability.

The discoveries indicate that the ANN model consistently demonstrates commendable performance, particularly for lead times up to 2 days. However, its accuracy moderates for a 3-day lead time forecast. Analysis of input variables reveals that incorporating discharge, rainfall, temperature, and inflow data, either individually or in combination, yields the best outcomes for runoff forecasting up to two days lead time. Furthermore, input time step selection significantly influences the model's predictive accuracy, with correlation-based lag times generally enhancing performance compared to constant lag times. Based on these findings, leveraging correlation-based lag times is recommended for developing ANN-based daily streamflow forecasting models, providing valuable insights for flood prediction applications.

Keywords: **Streamflow forecasting, Artificial Neural Network (ANN), Mahanadi River basin, Lead time, Hydro-meteorological data, Input variables, Time steps, Correlation- based lag, Constant lag, Performance evaluation, Flood prediction.**

I. INTRODUCTION

Globally, floods annually inflict widespread devastation on lives and property. The requirement for precise and timely flood forecasts is paramount in curtailing these damages. Recent strides in flood prediction hinge on machine learning (ML) and deep learning algorithms, lauded for their heightened efficacy compared to traditional methods. Within this study, a streamflow forecasting model, specifically the Artificial Neural Network, is developed solely for the upper Mahanadi River basin.

This investigation scrutinizes various input variables sourced from 2001 to 2011, encompassing daily discharge measurements from five stations, inflow data from the Hirakud Reservoir, as well as spatial average rainfall, temperature data from the upper Mahanadi basin. Meticulous selection of input time steps (lags) for each variable is undertaken, leveraging both correlation-based and constant lag methodologies.

Subsequently, the ANN model is deployed to forecast inflows to the Hirakud Reservoir, extending up to 3-days ahead. Performance evaluation is conducted utilizing metrics such as Nash-Sutcliffe Efficiency, (NSE), Ratio of root mean square error to standard deviation of measured data, (RSR), Mean Absolute Error, (MAE), and Error in volume, (Evol).

The ANN model consistently exhibits commendable performance across various scenarios, signifying its prowess in flood forecasting applications.

II. LITERATURE SURVEY

Numerous studies have delved into flood forecasting methodologies, each addressing specific challenges and employing diverse techniques. Tiwari and Chatterjee (2011)proposed the Wavelet–Bootstrap–ANN, (WBANN) hybrid model for daily discharge-forecasting in the Mahanadi Riverbasin, leveraging wavelet and bootstrapping techniques to enhance the accuracy and reliability of the ANN model. Rahman et al (2012) developed a flood forecasting system for the Jamuneswari river catchment in Bangladesh, integrating MIKE 11 river- modeling software modules and European Centre for Medium-Range Climate Forecasting (ECMCF) rainfall data to extend forecast lead-time.

Sahoo et al. (2019) investigated the suitability of an LSTM- RNN model for low-flow time series forecasting, demonstrating its superiority over traditional methods. Gao et al. (2020) proposed new data-driven flood forecasting models using LSTM and GRU networks, showcasing improved routine associated to ANN models, especially with improved time steps. Van et al (2020) introduced a novel 1D convolutional neural network, (CNN) for rainfall-runoff modeling, highlighting its effectiveness in capturing nonlinear relationships and outperforming traditional models.

Lee et al. (2021) assessed the effectiveness of DL models, particularly LSTM-based models, in predicting streamflow, demonstrating their capability even in the presence of upstream dams and reservoirs. Zhang et al. (2021) evaluated the influence of input variables on deep RNN models' predictions for daily runoff forecasting, emphasizing the importance of selecting relevant climate factors to enhance model accuracy.

Each of these studies contributes valuable insights into flood forecasting methodologies, showcasing the versatility and efficacy of machine learning (ML) and deep learning techniques in addressing complex hydrological challenges.

II.1. Studies on the Mahanadi River basin

Many studies have been directed on the Mahanadi River basin, shedding light on different aspects of flood forecasting and hydrological analysis. Sehgal et al. (2014) explored the Wavelet–Bootstrap–ANN, (WBANN) model for daily discharge forecasting, demonstrating its superiority over traditional models such as Multiple Linear Regression, (MLR) and Artificial Neural Network,(ANN). Jena et al (2014) investigated trends in peak discharge and extreme rainfall in the Mahanadi basin, attributing recent high floods to increased extreme rainfall in the central reaches. Nanda et al (2016) developed the wavelet-based nonlinear autoregressive with exogenous involvements, (WNARX) model for flood-forecasting, highlighting its effectiveness when shared with satellite- based rainfall products. Nanda et al. (2019) evaluated the Variable Infiltration Capacity (VIC) model for real-time streamflow forecasting, concluding that the VIC WNARXu model is the most effective for short to medium-range flood forecasting. Samantaray et al. (2021) compared statistical distribution methodologies and neural network algorithms for flood forecasting, introducing a hybrid neural network approach, (ANFIS-FFA) that outperformed other methods. Khatun et al. (2023) proposed the Smooth-LSTM typical for flood forecasting in the Mahanadi River basin, demonstrating its superior performance linked to other models such as LSTM, ANN, and MIKE11 NAM-HD.

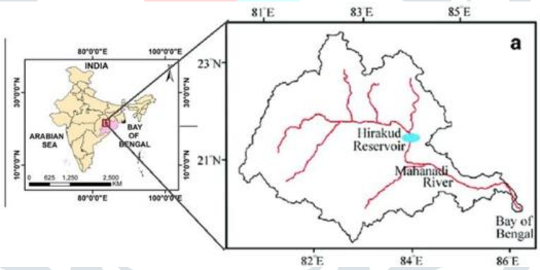


FIG: Index-map of the training area

II.2. Materials and methods

The research area encompasses the catchment of the Hirakud Reservoir (HR), spanning approximately 83,400 km2 and representing a substantial portion of the Mahanadi River sink in eastern India. Positioned between latitudes 19°90′–23°35′N and longitudes 80°30′–84°80′E, the catchment primarily depends on rainfall, receiving an average annual precipitation of about 1400 mm throughout the southwest monsoon season from June to September. With a tropical climate, temperatures range from 4–12°C in December and January to approximately 42–45°C in May and June. The topography varies from plains in the central area to hills in the northern regions, with usual advancements ranging from 200–400 m to 750–1000 m, respectively. The HR primarily serves for flood protection, boasting a live storage capacity of about 4823 × 106 m3. Given the recurrent extreme floods in the delta reaches of the Mahanadi River sink and the reservoir's pivotal role in flood mitigation, this study aims to develop data-driven models for forecasting influx into the HR with lead-time forecasting.

Data used,

This research relies on daily mean areal rainfall and mean areal temperature data, derived from spatially averaged gridded datasets of rainfall (0.25° × 0.25°), and temperature (1° × 1°) gained from the India Meteorological-Department (IMD), Pune. The daily observed inflows to the HR were sourced from the Hirakud Dam Circle, (HDC), Burla, Odisha. Additionally, daily discharge datasets from five gauging stations— Basantpur, Sundargarh, Kurubhata, Paramanpur, and Kelo— were collected from the India-Water Resources Information System, (India-WRIS). The analysis encompasses hydro- meteorological data aimed at the three southwest monsoon months of July, August, and September spanning from 2001 to 2011. Notably, data from July to September were utilized owing to the unavailability of continuous data intended for the month of June during the southwest monsoon season.

III. Methodology

In this section, a comprehensive explanation is provided regarding six distinct ML/DL methods employed for daily streamflow forecasting at the HR inflow location. Additionally, the methodology utilized for the expansion of these models is elucidated.

Artificial Neural Network (ANN)

Artificial Neural Network(ANN) had emerged as a widely adopted machine learning model for simulating streamflow within river basins. ANNs excel in capturing intricate relationships within hydro-meteorological time-series data due to their ability to process complex information through interconnected nodes with weighted links. The backpropagation algorithm (BPA) is commonly working for training multilayer ANNs, involving both a feed-forward and backward phase to adjust weights based on variance between computed and actual value. In this, a feed- forward ANN model is developed specifically for multi-step ahead daily streamflow forecasting, comprising an Input layer, a Hidden layer, and an Output layer. The architecture comprises an input layer with 50 neurons utilizing the ReLU activation function, followed by a hidden layer containing 25 neurons also using ReLU activation. Finally, there's an output layer with one neuron employing the sigmoid activation function. The model is enhanced using the Adam optimizer and trained to minimize Mean Squared Error (MSE) loss during training, with parameters determined through a trial-and-error approach.

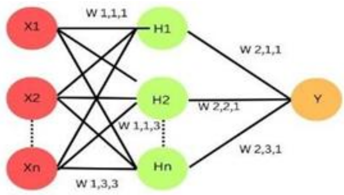


Fig: ANN Architecture

IV. DEVELOPMENT OF PROPOSED MODELS

(a) Types of input variables

In the study, proposed ML/DL models consider various hydro-meteorological factors influencing the rainfall-runoff process. The meteorological inputs for the Hirakud reservoir (HR) catchment include Mean Areal Rainfall (R) and Mean Areal Temperature (T). Additionally, the discharge data from five gauging stations—Basantpur, Sundargarh, Kurubhata, Paramanpur, and Kelo (collectively represented as Q)— within the HR catchment are incorporated. Furthermore, the inflow data (I) to the HR are also included as input variable quantity in the models.

(b)Time step selection

Since the presentation of data-driven models heavily relies on the selection of input variables and their time steps (Tiwari and Chatterjee, 2010b), this study initially adopts the conventional method of choosing input time steps. This approach involves conducting statistical analyses such as cross-correlation function, (CCF), autocorrelation function, (ACF), and partial autocorrelation function, (PACF) between the variables (Sudheer et al., 2002; Tiwari and Chatterjee, 2011; Sehgal et al. 2014). Consistent with previous literature (Nanda et al. 2016; Khatun et al., 2023), Additionally, to evaluate the significance of input time step selection, an investigation involving constant lag times is carried out, as discussed later.

V. MODEL DEVELOPMENT PROCEDURE

In the advance technique for the ANN model, the first step includes breaking down the time-series data into multiple input/output patterns, also known as samples. To expedite the learning process, data normalization is achieved using the scikit-learn 'MinMax' Scaler, which normalizes the input time series within the range of [0,1]. The Keras layers, a high-level TensorFlow API, are utilized to construct neural network (NN) models. An 80:20 ratio of input time-series data is allocated for model training (2001- 2007) and testing (2008-2011), respectively. Additionally, 20% of training data (2006-2007) is reserved for model validation to avoid overfitting or underfitting.

Initially, a base model is established, defining the network architecture, hyperparameters, and loss function. This base model considers all four input variables, with model limits determined through a trial-and-error approach. The time steps (lags) for to each input variable are determined based on the ACF, PACF, and CCF values. The model architecture specifies the input shape argument in the first hidden layer, defining the input shape for each sample.

To progress a well-trained model, various parameters such as the number of layers, nodes per layer, time-steps, dropout ratio, number of epochs, and batch size need to be tuned. Subsequently, two distinct exercises are conducted to discern the result of different input variables and time lags:

i. Case I, mentioned to as the 'correlation' approach, 15 combinations of input variables are considered, with time lags determined using the correlation-based approach.

ii. Case II, known as the 'constant' approach, the same 15 combinations of input variables are examined, with constant time lags.

These steps ensure a comprehensive analysis of input variables and time lags, facilitating the advance of an optimized ANN model for daily streamflow forecasting.

ANN

The time-series plot depicted in Fig. 4.3a illustrates that the ANN model's calculations closely line up with observed values, particularly at a 1-day lead time, regardless of whether constant or correlation approaches are employed. Evaluation metrics provided in Tables 4.1 and 4.2 reinforce this observation, indicating superior performance of ANN model with NSE values of 0.93 and 0.92, RSR values of 0.27 and 0.29, and MAE value of 688.47 m3/s and 639.77 m3/s for the constant and correlation scenarios, respectively.

However, as the Lead Time upsurges to 2 and 3 days, The ANN model's predictive accuracy diminishes notably. The scatter plot demonstrates that while the model adeptly captures low-flow events, it struggles with medium and high-flow events. Specifically, for a 2-day lead time, NSE values of 0.76 and 0.76, RSR values of 0.49 and 0.49, and MAE values of 955.27 m3/s and 983.27 m3/s are recorded for the constant and correlation cases, respectively. At a 3- day lead time, the model's performance deteriorates further, evidenced by NSE<0.5, RSR>0.5, and MAE values of 1426.46 m3/s and 1504.41 m3/s.

Model	Combination	Criteria	1	2	3
ANN	Q, R	NSE	0.92	0.76	0.47
		RSR (-)	0.29	0.49	0.73
		MAE (m^3/s)	639.77	983.27	1504.4
		E_{vol} (%)	7.18	5.34	11.67

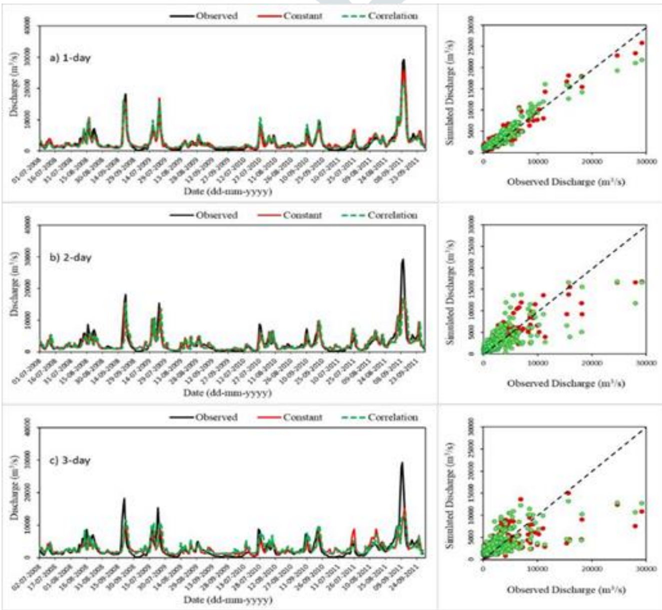
Fig: presents the time sequence and scatter-plots illustrating the observed and predicted discharges using the ANN model.

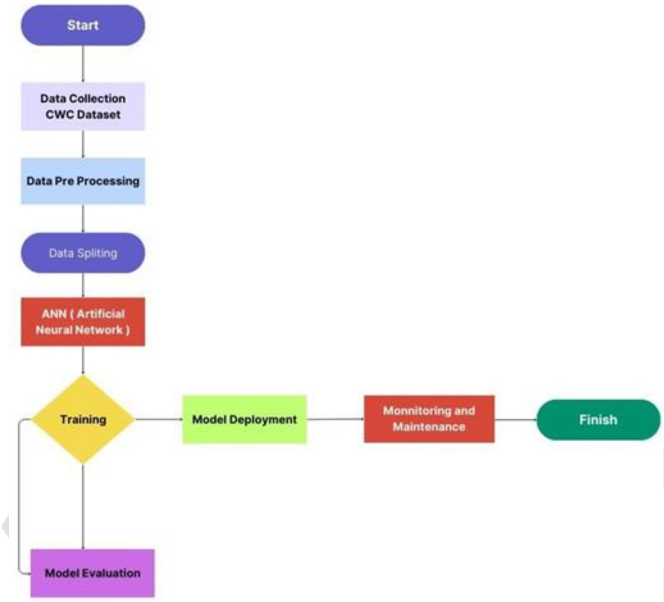
Model	Combination	Criteria	1	2	3
ANN	Q, R, T, I	NSE	0.93	0.76	0.44
		RSR (-)	0.27	0.49	0.75
		MAE (m^3/s)	688.47	955.27	1426.46
		E_{vol} (%)	4.08	-3.63	2.29

Fig: Performance valuation metrics during testing with lag inputs based on cross-correlation function (CCF) and autocorrelation function (ACF).

These conclusions suggest possible issues with the ANN model, such as slow learning rates, susceptibility to converging towards local minima, and overfitting, mostly in the situation of higher lead time flood forecasting. Notably, the ANN model demonstrates optimal performance when utilizing Q, R, T, I, and Q, R as input variables.

Fig: Flow chart of Flood forecasting using ANN





VI. RESULTS AND DISCUSSION

Selection of model inputs

In the setting of this study, selecting the appropriate input features is crucial for optimizing ML/DL model performance by enhancing training efficiency, overall accuracy, and mitigating overfitting. Initially, a base model is constructed incorporating all input variables and correlation-based time lags. Subsequently, various model runs are conducted to measure the influence of input variables and associated time lags.

The correlation coefficient function (CCF) between discharge data from five gauging stations (Basantpur, Sundargarh, Kurubhata, Paramanpur, Kelo) and inflows to the HR is analyzed, as depicted. Additionally, the CCF between Rainfall-discharge and Temperature- discharge pairs for the HR catchment. Generally, the correlation magnitude diminishes with increasing lag time. Notably, significant correlations (CCF>0.3) between inflows and discharge (across the five stations) are observed within 1 to 4 days lag. Based on these findings, time lags of 3, 3, 1, 0, and 2 days are designated for Basantpur, Sundargarh, Kurubhata, Paramanpur, and Kelo, respectively. Regarding the CCF values between rainfall and inflows, a 1-day lag is chosen. Despite the negative correlation observed between temperature and discharge time series at the HR inflow location, the temperature dataset is retained in the analysis. This decision aligns with preceding studies by Khatun et al. (2023), which also described negative associations between temperature-inflow data pairs. Reliable with the approach adopted by Nanda et al. (2016), where temperature data was included despite negative correlations, this study incorporates temperature data in subsequent analyses.

VII. SUMMARY

This study focuses on developing a standalone ANN model to forecast daily short-to-medium range streamflow at 1-3 days lead time within the HR catchment, situated in the upper region of the Mahanadi-River basin in Eastern-India. Evaluating the ANN model involves assessing its performance in predicting streamflow forecasts using various input variables and time steps (lags), employing both correlation-based methods (autocorrelation and cross-correlation functions) and constant lags.

Compared to preceding research, this learning proposals innovation by investigating the influence of four input parameters on ANN model performance concerning the number of lag days considered. Through this analysis, the study contributes valued visions into the utilization of ANN models for runoff-simulation and the inspiration of input parameters on model effectiveness.

VIII. CONCLUSION

a) The ANN model exhibits promising performance for daily streamflow forecasting up to a 2-day lead time, outperforming other variants considered. However, it struggles to deliver precise predictions for a 3-day lead time.

b) The optimal of input variables significantly impacts the ANN model's predictive capability. Specifically, using discharge (Q), rainfall (R), temperature (T), and inflow (I) data, or a grouping of discharge, rainfall, and inflow data, yields the best results for runoff forecasting up to two days lead time.

c)Input time step selection is critical designed for enhancing the ANN model's predictive accuracy. While employing constant lag times yields reasonably accurate forecasts, incorporating correlation-based lag times generally improves performance. Given the data-dependent nature of constant lag time selection, leveraging the traditional approach of correlation-based lag times is advisable for developing ANN-based daily streamflow forecasting representations.

IX. REFERENCES

1. Tiwari, M. K., & Chatterjee, C. (2010). “Development of an accurate and reliable hourly flood forecasting model using wavelet–bootstrap– ANN (WBANN) hybrid approach”. Journal of Hydrology, 394(3-4), 458-470

2. Rahman, M. M., Goel, N. K., & Arya, D. S. (2012). “Development of the Jamuneswari flood forecasting system: Case study in Bangladesh”. Journal of Hydrologic Engineering, 17(10), 1123-1140.

3. Rahman, M. M., Goel, N. K., & Arya, D. S. (2012). “Development of the Jamuneswari flood forecasting system: Case study in Bangladesh”. Journal of Hydrologic Engineering, 17(10), 1123-1140.

4. Adnan RM, Petroselli A, Heddam S, Santos CAG, Kisi O (2021), “Comparison of different methodologies for rainfall–runof modeling: machine learning vs conceptual approach”. Nat Hazard 105(3):2987–3011.

5. Akhtar, M. K., Corzo, G. A., van Andel, S. J. & Jonoski, A. “River flow forecasting with artificial neural networks using satellite observed precipitation pre-processed with flow length and travel time information: Case study of the Ganges River basin”. Hydrol. Earth Syst. Sci. 13, 1607–1618 (2009).

6. Azadeh, A., & Alem, S. M. (2010). “A flexible deterministic, stochastic and fuzzy Data Envelopment Analysis approach for supply chain risk and vendor selection problem: Simulation analysis”. Expert Systems with Applications, 37(12), 7438-7448.

7. Beven, K., & Binley, A. (1992). “The future of distributed models: model calibration and uncertainty prediction”. Hydrological processes, 6(3), 279- 298.Chen, C., Jiang, J., Liao, Z., Zhou, Y., Wang, H., & Pei, Q. (2022). A short-term flood prediction based on spatial deep learning network: A case study for Xi County, China. Journal of Hydrology, 607, 127535.

8. Chen, Y., Li, J., & Xu, H. (2016). “Improving flood forecasting capability of physically based distributed hydrological models by parameter optimization”. Hydrology and Earth System Sciences, 20(1), 375- 392.

9. Lecun Y et al (1989), “Backpropagation applied to handwritten zip code recognition”. Neural Comput 1(4):541–551.

10. Li, X., Xu, W., Ren, M., Jiang, Y., & Fu, G. (2022). “Hybrid CNN-LSTM models for river flow prediction”. Water Supply, 22(5), 4902-4919.

11. Mosavi, A., Ozturk, P., & Chau, K. W. (2018). “Flood prediction using machine learning models: Literature review”. Water, 10(11), 1536.

12. Nanda, T., Sahoo, B., Beria, H., & Chatterjee, C. (2016). “A wavelet-based non-linear autoregressive with exogenous inputs (WNARX) dynamic neural network model for real-time flood forecasting using satellite-based rainfall products”. Journal of Hydrology, 539, 57-73.

13. Nourani, V. (2017). “An emotional ANN (EANN) approach to modelling rainfall-runoff process”. Journal of Hydrology, 544, 267-277.