International Journal of Advanced Biochemistry Research 2025; 9(4): 32-36

International Journal of Advanced Biochemistry Research

ISSN Print: 2617-4693
ISSN Online: 2617-4707
NAAS Rating: 5.29
IJABR 2025; 9(4): 32-36
www.biochemjournal.com
Received: 03-02-2025
Accepted: 07-03-2025

Harshita Rani Ahirwar
Department of Farm Engineering, Institute of Agricultural Sciences, Banaras Hindu University Varanasi, Uttar Pradesh, India

Kuldeep Singh Kushwaha
National Institute of Hydrology, Jal Vigyan Bhavan, Roorkee, Haridwar, Uttarakhand, India

Vijay Shankar Yadav
Department of Farm Engineering, Institute of Agricultural Sciences, Banaras Hindu University Varanasi, Uttar Pradesh, India

AK Nema
Department of Farm Engineering, Institute of Agricultural Sciences, Banaras Hindu University Varanasi, Uttar Pradesh, India

MK Nema
National Institute of Hydrology, Jal Vigyan Bhavan, Roorkee, Haridwar, Uttarakhand, India

Corresponding Author:
Harshita Rani Ahirwar
Department of Farm Engineering, Institute of Agricultural Sciences, Banaras Hindu University Varanasi, Uttar Pradesh, India

Revolutionizing flood forecasting with machine learning techniques

Harshita Rani Ahirwar, Kuldeep Singh Kushwaha, Vijay Shankar Yadav, AK Nema and MK Nema

DOI: https://www.doi.org/10.33545/26174693.2025.v9.i4a.4055

Abstract
Flooding is a major natural disaster that disrupts lives, damages infrastructure, and devastates agriculture, emphasizing the importance of accurate forecasting for better water resource management and disaster preparedness. This research examines flood forecasting techniques at the Ramakona gauging station on the Wainganga River, leveraging 31 years (1987-2017) of peak flow data. The study evaluates two widely used methods: ARIMA and ANN, which are machine-learning approaches. Analysis of historical data revealed significant fluctuations in peak discharge, highlighting the complexity of flood behavior at the site. Results showed that ANN significantly outperformed ARIMA in predicting flood magnitudes, with lower error rates and greater accuracy in capturing nonlinear patterns. These findings demonstrate the potential of ANN as a powerful tool for short-term flood prediction, offering actionable insights for enhancing flood management strategies. The study underscores the role of advanced machine-learning techniques in reducing the impacts of floods and improving water management systems.

Keywords: Flood, flood forecasting, Wainganga River, Ramakona, machine learning, ARIMA, ANN

Introduction

India is traversed by an extensive network of rivers that span across the country. These rivers hold immense potential to serve as significant sources of wealth and energy if managed effectively [1]. Flood estimation is a crucial component of hydrological design and plays a key role in developing strategies for flood protection and management [2]. Assessing the maximum potential flood stage and its associated phase is vital for planning flood control measures, hydropower projects, and irrigation systems [3]. However, when a river exceeds its banks, the consequences can be devastating, leading to significant property damage, loss of lives, and destruction of agricultural land [4-7].

Global warming and climate change are anticipated to amplify the frequency and intensity of large-scale floods in the coming years. Additionally, changes in land use, population growth and deforestation in flood-prone regions are expected to further increase the damage caused by flood forecasting is a critical aspect of applied hydrology, given the frequency and destructive nature of floods and their potential to disrupt the socioeconomic progress of communities [8]. Flood frequency analysis (FFA) is widely used to estimate critical discharge levels required for designing hydraulic structures [9]. This involves analyzing historical peak discharge records to derive statistical variables and generate frequency distributions that represent the likelihood of different discharge levels occurring over time [10].

Various forecasting techniques have been developed to predict flood events. Machine learning algorithms offer a distinct advantage in their ability to predict hydrological patterns over extended periods, making them highly valuable for long-term water management planning [11]. Among these, autoregressive models like Artificial Neural Networks (ANN) and Autoregressive Integrated Moving Average (ARIMA) have proven effective [12]. These models utilize historical hydrological data, such as rainfall, temperature, and water levels, to identify patterns and forecast future trends [13]. ANN models are particularly adept at handling nonlinear time series and are applied to forecast at different time scales, ranging from daily to yearly [14]. Autoregressive models, including ARIMA and ANN, are effective for both short- and long-term flood forecasting [15]. Short-term models typically provide predictions for one to seven days, while long-term models extend forecasts to months or even years. Such extended forecasts are crucial for optimizing the operations of multipurpose reservoirs, which serve multiple functions like water supply, irrigation, hydropower, and flood control [16].

This research focuses on one gauging station along the Wainganga River, a tributary of the Godavari River, chosen for its suitability for damming and safe water management. The study aims to evaluate peak discharge (Flood) magnitudes using advanced forecasting techniques, identify the most appropriate method for the station, and compare these methods to assess their predictive accuracy and reliability.

Study Area

The Wainganga River, a tributary of the Godavari River, originates in the Seoni district of Madhya Pradesh at an elevation of approximately 640 meters. The river initially flows eastward before curving south, traversing about 334 kilometers through Maharashtra and 274 kilometers through the Seoni and Balaghat districts of Madhya Pradesh. Before merging with the Wardha River, which contributes to forming the Pranhita River, the Wainganga served as a vital water source for the region. This study utilizes data from the Ramakona gauging station, located (Fig. 1) along the upstream of the Wainganga River. The geographical coordinates of Ramakona are 21.7200° North latitude and 78.8242° East longitude. Managed by the Central Water Commission (CWC). Ramakona employs manual methods for data collection and has a zero gauge level set at 335.95 meters, which serves as the reference point for measuring water levels. Data for this station were obtained from the INDIA WRIS platform (https://indiawris.gov.in/wris/#/RiverMonitoring), which provides detailed information on river monitoring across India. The insights from this station are crucial for understanding hydrological patterns in the Wainganga River and form the basis of this study's analysis.

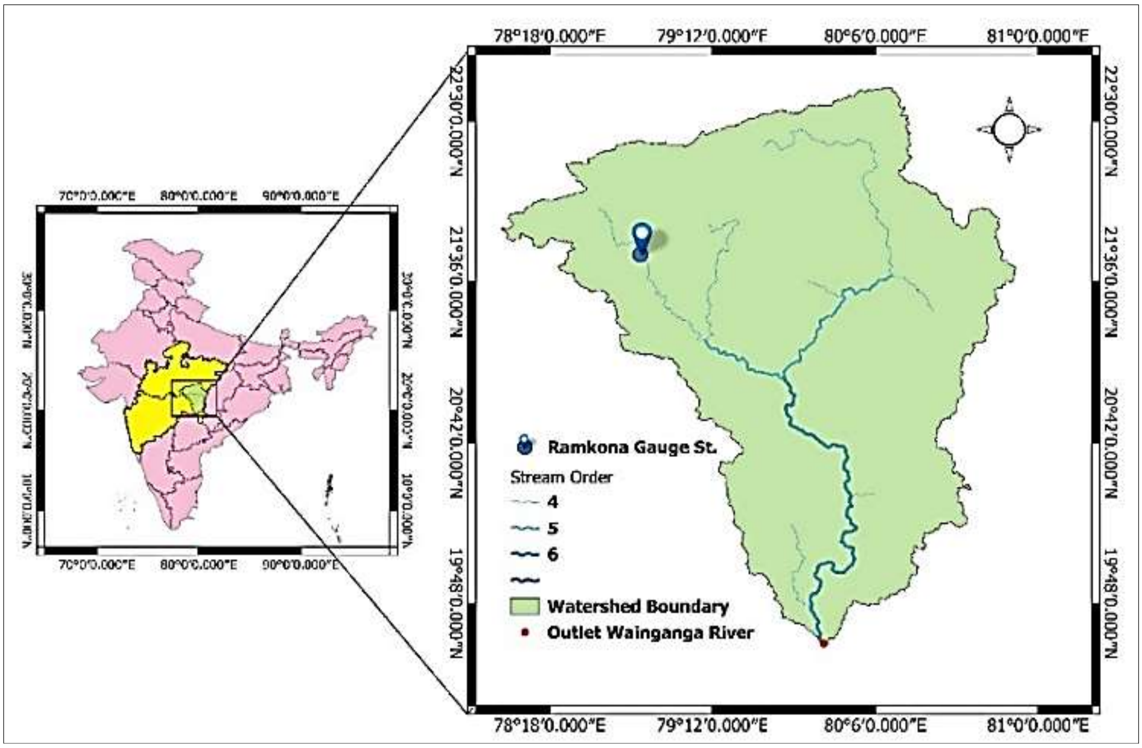

Fig 1: Location of Gauge Station in Wainganga River Basin

Methodology

The ARIMA approach for predicting future values in time series data

The ARIMA (Auto-Regressive Integrated Moving Average) method is a popular tool for time series forecasting. It works by combining three key components: the auto regression (AR) model, which looks at the relationship between a value and its previous values [17]. The indicates the count of lagged observations by p. For instance if p=1, the model has predicted the current value using the immediately preceding value [18]. Integration (I), which removes trends to make the data stable and easier to analyze. The total differences required to acquire stationarity in series is denoted by d. For example, if d=1 it is understood that the model has used differences between consecutive observations; and moving average (MA), which captures the influence of past errors on current values; The count of lagged forecast errors included in the model is denoted by parameter q [18].

The ANN approach for predicting future values in time series data

Artificial Neural Networks (ANNs) are inspired by how the human brain processes information and are highly effective tools for time series forecasting. These networks consist of layers of interconnected nodes, or neurons, that work together to recognize patterns and relationships in data. Unlike traditional statistical methods, ANNs excel at capturing complex, non-linear patterns in historical data. By learning from past observations, they can predict future trends by adjusting their internal weights to reduce errors. This adaptability and ability to handle large datasets make ANNs a popular choice for forecasting in areas like finance, weather prediction and energy demand [19]. ANNs are structured with three key layers: the input layer, hidden layer(s), and output layer, each serving a unique purpose. The input layer is where data first enters the network [20]. The hidden layer(s) lies between the input and output layers and handle most of the heavy lifting in terms of computation. These layers are made up of neurons that process data by applying mathematical operations like weighted sums and activation functions. The output layer is the final stop, delivering the network's results. The number of nodes in this layer depends on what the network is trying to achieve. Together, these layers work in harmony to analyze data and make predictions.

Result and Discussion

Data Description and Pre-Processing

The data for the Ramakona gauging site covers 31 years of annual peak flow observations. During this period, the lowest recorded peak flow was 87.78 cubic meters per second (Cumecs), while the highest reached 4548.25 cumecs, reflecting significant variability in flood magnitudes. On average, the annual peak flow at the site was 1345.91 cumecs, with a standard deviation of 1125.75 cumecs, indicating considerable year-to-year fluctuations. The data distribution shows a moderate right skew, as evidenced by a skewness value of 0.93. This means that while most of the recorded values are below the mean, a few extreme high-flow events have a noticeable impact. The kurtosis value of 0.80 suggests a relatively flat distribution with fewer extreme outliers compared to a sharper, more peaked dataset. This statistical analysis highlights the variability and occasional extremity of flood events at the Ramakona site, providing valuable insights for flood risk assessment and water management planning.

Comparison and acceptability of ANN and ARIMA

To compare the performance of ARIMA and ANN models, evaluating errors is crucial for determining the best-fit model. In this study, the RMSE (Root Mean Square Error) was calculated by comparing the original peak discharge with the predicted peak discharge (Fig. 2) for the Ramakona gauging station during the testing years (2013-2017). The RMSE(s) were then compared to the series' mean to assess the models' acceptability. The acceptability of a model is established when the mean of the series exceeds the RMSE (original versus predicted discharge in the testing set). The lesser the RMSE value the better the result is considered. The RMSE from ANN is around 16% whereas from ARIMA it's about 26% of the maximum discharge value. The results for Ramakona, as shown in Fig. 3 the analysis, confirm that the ANN model meets this criterion, demonstrating superior performance. Conversely, the ARIMA model did not achieve acceptability at this station due to higher RMSE values than the mean. This highlights ANN's greater suitability for predicting peak discharge at the Ramakona gauging site. The predicted values for the proceeding 5 years (2018-2023) are shown in table 1.

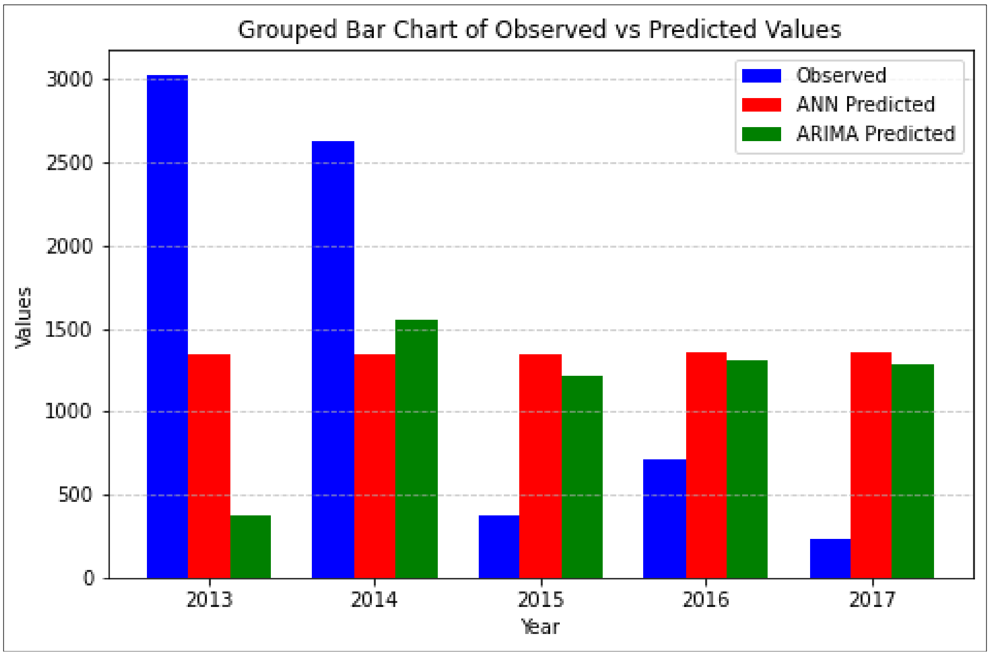

Fig 2: Observed Discharge Data with Predicted Data by ARIMA and ANN

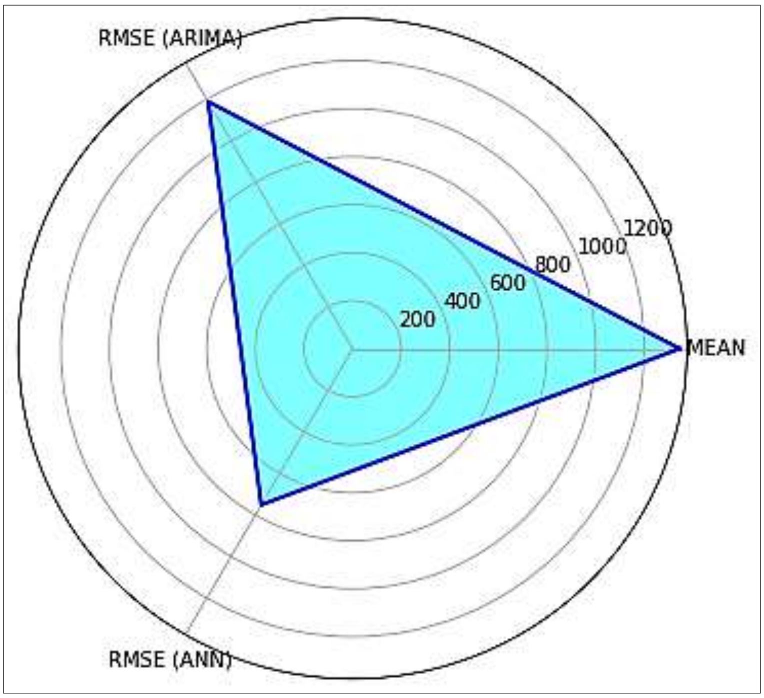

Fig 3: Mean (of series), RMSE of testing set with ARIMA and ANN.

Table 1: Predicted Discharge by ANN model

Year	2018	2019	2020	2021	2022
Discharge	1357.24	1357.91	1358.59	1359.26	1359.93

Conclusion

Flood forecasting is essential for managing water resources and reducing the negative impacts of floods. In this study, the focus was on the Ramakona gauging station, where machine-learning techniques were used to evaluate peak discharge. Among the methods tested, ARIMA and ANN models were applied for time series forecasting. The results showed that ANN performed better than ARIMA in predicting peak discharge at Ramakona, demonstrating its ability to handle the station's complex hydrological behavior. This research highlights the value of using machine-learning approaches, particularly ANN, for more accurate short-term flood predictions. By effectively capturing flood dynamics, ANN offers a promising tool for enhancing preparedness and mitigating flood risks. The insights gained from this study can help improve flood management strategies for Ramakona and similar regions.

References

1. Islam MR, Islam MR, Beg MRA. Renewable energy resources and technologies practice in Bangladesh. Renew Sustain Energy Rev. 2008 Feb;12(2):299-343. DOI: 10.1016/j.rser.2006.07.003.
2. Oyebode OJ, Oyerinde AO, Oyebode FA. Engineering Hydrology for Flood Control, Adequate Water Resources and Sustainable Environment in Nigeria. J Water Resour Eng Manag. 2023;10(2):1-10.
3. Nguyen-Tien V, Elliott RJR, Strobl EA. Hydropower generation, flood control and dam cascades: A national assessment for Vietnam. J Hydrol. 2018 May;560:109-26. DOI: 10.1016/j.jhydrol.2018.02.063.
4. Rahman A-U, Khan AN. Analysis of flood causes and associated socio-economic damages in the Hindukush region. Nat Hazards. 2011 Dec;59(3):1239-1260. DOI: 10.1007/s11069-011-9830-8.
5. Alderman K, Turner LR, Tong S. Floods and human health: A systematic review. Environ Int. 2012 Oct;47:37-47. DOI: 10.1016/j.envint.2012.06.003.
6. Abegaz R, Wang F, Xu J. History, causes, and trend of floods in the U.S.: A review. Nat Hazards. 2024 Dec;120(15):13715-13755. DOI: 10.1007/s11069-024-06791-y.
7. Doocy S, Daniels A, Murray S, Kirsch TD. The Human Impact of Floods: a Historical Review of Events 1980-2009 and Systematic Literature Review. PLoS Curr. 2013 Apr;5:ecurrents.dis.f4deb457904936b07c09daa98ee8171a. DOI: 10.1371/currents.dis.f4deb457904936b07c09daa98ee8171a.
8. Roy P, Chandra Pal S, Chakrabortty R, Chowdhuri I, Malik S, Das B. Threats of climate and land use change on future flood susceptibility. J Clean Prod. 2020 Nov;272:122757. DOI: 10.1016/j.jclepro.2020.122757.
9. England JF. Flood Frequency and Design Flood Estimation Procedures in the United States: Progress and Challenges. Australas J Water Resour. 2011 Jan;15(1):33-46. DOI: 10.1080/13241583.2011.11465388.
10. Reed S, Schaake J, Zhang Z. A distributed hydrologic model and threshold frequency-based method for flash flood forecasting at ungauged locations. J Hydrol. 2007 Apr;337(3):402-420. DOI: 10.1016/j.jhydrol.2007.02.015.
11. Shen C. A Transdisciplinary Review of Deep Learning Research and Its Relevance for Water Resources Scientists. Water Resour Res. 2018;54(11):8558-8593. DOI: 10.1029/2018WR022643.
12. Jain A, Kumar AM. Hybrid neural network models for hydrologic time series forecasting. Appl Soft Comput. 2007 Mar;7(2):585-592. DOI: 10.1016/j.asoc.2006.03.002.
13. Azad AS, *et al*. Water Level Prediction through Hybrid SARIMA and ANN Models Based on Time Series

Analysis: Red Hills Reservoir Case Study. Sustainability. 2022 Jan;14(3):1843. DOI: 10.3390/su14031843.

14. Ustaoglu B, Cigizoglu HK, Karaca M. Forecast of daily mean, maximum and minimum temperature time series by three artificial neural network methods. Meteorol Appl. 2008;15(4):431-445. DOI: 10.1002/met.83.
15. Banihabib ME, Bandari R, Peralta RC. Auto-Regressive Neural-Network Models for Long Lead-Time Forecasting of Daily Flow. Water Resour Manag. 2019 Jan;33(1):159-172. DOI: 10.1007/s11269-018-2094-2.
16. Jain SK, Shilpa LS, Rani D, Sudheer KP. State-of-the-art review: Operation of multi-purpose reservoirs during flood season. J Hydrol. 2023 Mar;618:129165. DOI: 10.1016/j.jhydrol.2023.129165.
17. Kaur J, Parmar KS, Singh S. Autoregressive models in environmental forecasting time series: a theoretical and application review. Environ Sci Pollut Res. 2023 Feb;30(8):19617-19641. DOI: 10.1007/s11356-023-25148-9.
18. Shumway RH, Stoffer DS. ARIMA Models. In: Shumway RH, Stoffer DS, editors. Time Series Analysis and Its Applications: With R Examples. Cham: Springer International Publishing; c2017. p. 75-163. DOI: 10.1007/978-3-319-52452-8_3.
19. Shanmuganathan S. Artificial Neural Network Modelling: An Introduction. In: Shanmuganathan S, Samarasinghe S, editors. Artificial Neural Network Modelling. Cham: Springer International Publishing; c2016. p. 1-14. DOI: 10.1007/978-3-319-28495-8_1.
20. Chen M, Challita U, Saad W, Yin C, Debbah M. Artificial Neural Networks-Based Machine Learning for Wireless Networks: A Tutorial. IEEE Commun Surv Tutor. 2019;21(4):3039-3071. DOI: 10.1109/COMST.2019.2926625.