Journal of Applied Science, Engineering, Technology and Management
Vol. 02, Issue 01, June 2024 pp. 19 - 24

Integration of Lag Time Analysis and Antecedent Rainfall Scenarios in Artificial Neural Network Models for Streamflow Prediction

Nirmala Theresa
Water Resources and Hydroinformatics Department
Government Engineering College
Thrissur, India
theresa.nirmala@gmail.com

Smitha Mohan K
Water Resources and Hydroinformatics Department
Government Engineering College
Thrissur, India
smitha.mohanam@gmail.com

***Abstract*— Artificial Neural Networks (ANN) play a vital role in hydrological modeling, adept at capturing intricate, non-linear relationships in dynamic environmental systems. This study investigates the interplay between lag time in a basin and the predictive accuracy of ANN models for streamflow forecasting. Three distinct ANN models, employing varied antecedent rainfall scenarios as inputs, were evaluated. Results indicate a robust correlation between the lag time of the basin and the efficacy of the ANN model. Notably, the model incorporating a 3-day antecedent rainfall scenario demonstrated superior predictive performance, aligning precisely with the calculated lag time. This synchronicity underscores the relevance of tailoring input features to the inherent characteristics of the watershed. Beyond model optimization, the study's significance lies in its contribution to flood risk management, offering advanced tools for more accurate and efficient predictions in the face of increasing climate-induced extreme weather events.**

Keywords—*Artificial Neural Network, Lag Time, Streamflow Prediction, Antecedent Rainfall, Hydrological Modeling.*

I. Introduction

Water resource management demands accurate streamflow prediction models, crucial for effective decision-making in the face of evolving hydrological dynamics. Artificial Neural Networks (ANNs) have emerged as powerful tools in this domain, offering the capacity to capture complex nonlinear relationships within hydrological processes. However, the optimal configuration of input parameters remains a persistent challenge.

This study focuses on unraveling the intricate relationship between the lag time of a basin and the performance of ANN models, emphasizing the pivotal role of antecedent rainfall scenarios in enhancing predictive accuracy. Antecedent rainfall, representing the antecedent moisture condition of a watershed, is known to influence streamflow. Yet, the synergies between the lag time, which characterizes the temporal delay in basin response, and the choice of antecedent rainfall scenarios in shaping ANN model outcomes have not been comprehensively explored. Through meticulous experimentation, we present an empirical investigation into three distinct ANN models, each utilizing various combinations of antecedent rainfall scenarios. The outcomes reveal a compelling connection between the lag time of the basin and the optimal antecedent rainfall duration for improved predictive performance. Our findings not only contribute to advancing the understanding of hydrological processes but also provide a practical framework for refining ANN-based streamflow prediction models.

This paper unfolds a pathway towards the harmonization of model inputs with basin characteristics, contributing to the ongoing discourse on enhancing the precision of hydrological modeling techniques. The implications extend beyond the realm of academia, offering valuable insights for practitioners engaged in water resource management and environmental planning.

II. Study Area

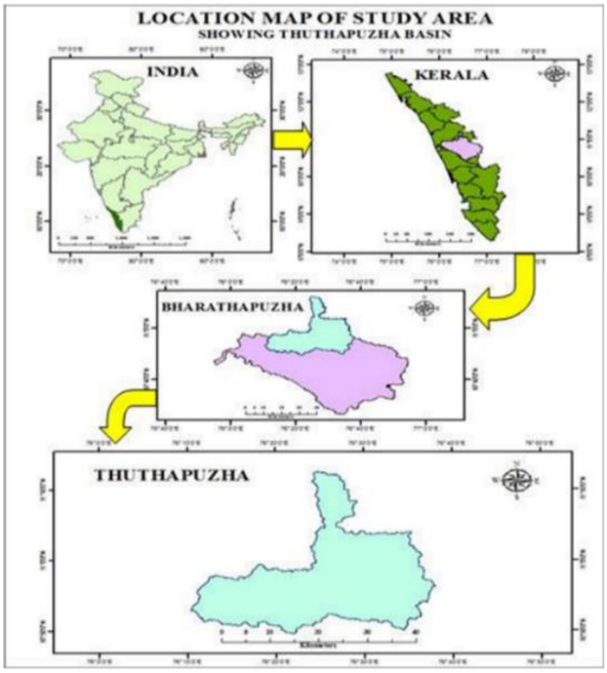

Fig.1. Thuthapuzha river basin.

The Thoothapuzha River basin (1), situated in the Indian state of Kerala, is a vital geographical and ecological region. It serves as a significant tributary to the Bharathapuzha River, the second-longest river in Kerala. Originating from the Western Ghats at an elevation of 1,375 meters above sea

eISSN 2584-0371

Received for review: 26-01-2024
Accepted for publication: 22-03-2024
Published: 26-03-2024

level, the Thoothapuzha River traverses approximately 130 kilometers, culminating in the Arabian Sea near Thootha in the Malappuram district. The river is an essential source of water for irrigation and domestic use in the area.

III. DATA SET

This study necessitates a robust dataset encompassing diverse parameters essential for a comprehensive hydrological analysis. The primary data requirements include a decade-long record of rainfall data, spanning ten years. This temporal depth is crucial for capturing long-term precipitation patterns and trends. Additionally, a decade-long dataset of discharge records spanning the same period is vital to construct a reliable training and validation set for the ANN model.

Beyond meteorological variables, geographical features play a pivotal role. Digital Elevation Model (DEM) data of the study area is indispensable for understanding the terrain's influence on hydrological processes. Moreover, a detailed Land Use Land Cover (LULC) map provides insights into the land surface characteristics, influencing runoff patterns. The inclusion of a Soil Map is equally vital, offering information on soil types and properties, which significantly impact water infiltration and retention.

A. Rainfall data

The rainfall and data is an important input variable in hydrological modelling and helps in understanding the water balance of the region. The rainfall data used in this study was collected from the Indian Meteorological Department (IMD) website. The data spans from 2005 to 2018 and was collected on a daily basis.

B. Discharge data

The discharge data is used as the output variable for the ANN model. The discharge data was collected on a daily basis for the time period of 2005 to 2015. This data was obtained from the Water Resources Information System (WRIS) website. The discharge data is an important parameter in hydrology and is used to understand the behaviour of rivers and streams. The availability of long-term discharge data is crucial in developing reliable hydrological models.

C. Digital Elevation Model (DEM)

A DEM is a digital representation of the topography of an area, and can be used to identify the boundaries of a watershed by tracing the flow of water from higher elevations to lower elevations.

Fig. 2. DEM of Thuthapuzha river basin.

By using the DEM to delineate the watershed, it is possible to identify the boundary of the Thuthapuzha watershed and to understand the physical characteristics of the basin, such as its size, shape, and elevation. Here, the DEM (Fig. 2) was downloaded from the USGS Earth Explorer site to delineate the Thuthapuzha watershed.

D. Land Use and Land Cover Map

To model the hydrological cycle, it is important to consider the land use and land cover (LULC) of the basin. The LULC map (Fig.3) can be used to calculate the curve number, which is a parameter used in hydrological models to estimate the rate of surface runoff. In this study, the LULC map was downloaded from the USGS Earth Explorer. By using the curve number, the lag time of the basin can be calculated, which is an important variable to be included in an artificial neural network (ANN) model. The accurate calculation of lag time can lead to better performance of the model in predicting the discharge.

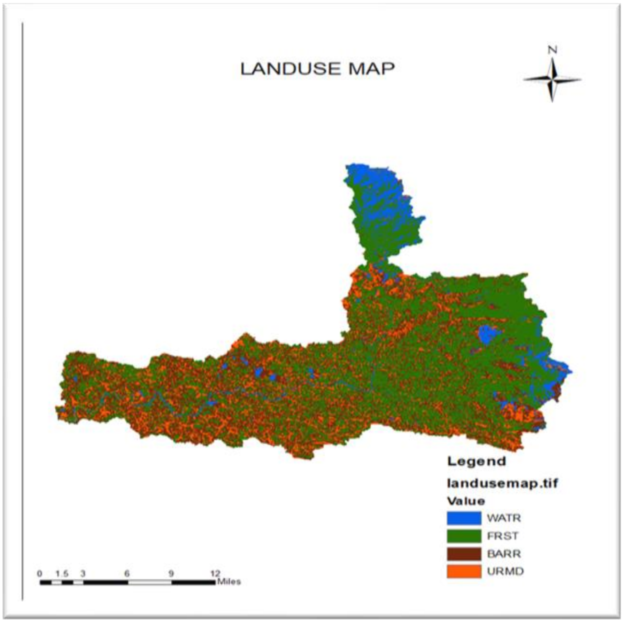

Fig.3 LULC Map of Thuthapuzha river basin.

E. Soil Map

A soil map is necessary to calculate the curve number and lag time of the Thuthapuzha basin. The world soil map was downloaded from the FAO website, and then the soil map of the Thuthapuzha basin (Fig. 4) was extracted. The soil map provides information about the soil type and characteristics of the basin, which are important factors in determining how quickly water infiltrates the soil and how much runoff occurs. By using the soil map, we are able to calculate both the curve number and the lag time of the Thuthapuzha basin, which are primary deciding factors of inputs for the hydrological modeling using ANN. The curve number is a measure of the runoff potential of a particular soil type, and it is used in hydrological modeling to estimate the amount of water that will enter a stream or river during a rainfall event.

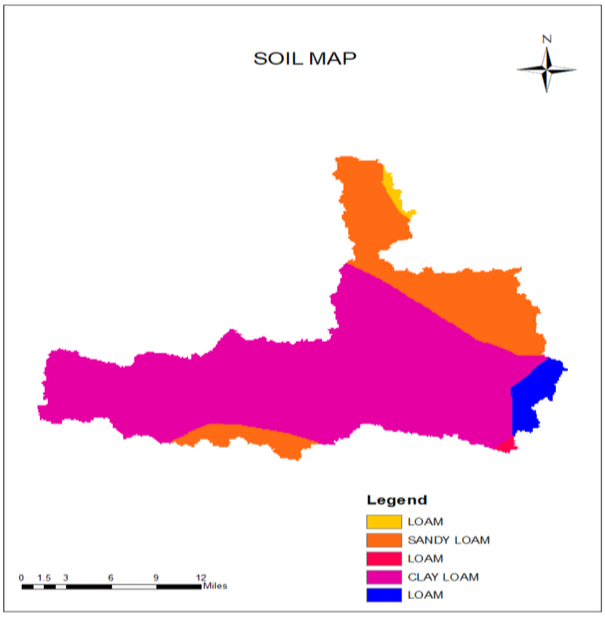

Fig.4 Soil Map of Thuthapuzha river basin.

IV. ANN MODELS IN STREAMFLOW PREDICTION

Artificial Neural Network (ANN) models have gained prominence in streamflow prediction due to their ability to capture complex, non-linear relationships inherent in hydrological processes [1]. These models, inspired by the human brain's neural networks, excel in recognizing patterns and learning from data. In the context of streamflow prediction, ANNs exhibit versatility by accommodating various input parameters, making them well-suited for the dynamic and interconnected variables influencing hydrological systems [2]. Researchers have explored diverse configurations of ANNs, incorporating antecedent rainfall scenarios, lag times, and other relevant factors to enhance predictive accuracy. The utilization of ANNs in streamflow prediction holds promise not only for its predictive prowess but also for its adaptability to evolving environmental conditions. This section delves into the application of ANN models, emphasizing their efficacy, challenges, and the ongoing quest for optimal model configurations in the realm of streamflow forecasting.

One of the significant advantages of ANN models is their ability to learn from historical data and make accurate predictions in real-time, which can be helpful in issuing flood alerts and taking necessary mitigation measures to minimize the damages caused by floods [2]. However, ANN models require a large amount of data for training and may suffer from overfitting and underfitting problems if not appropriately calibrated [3]. Therefore, careful selection of input variables, network architecture, and appropriate training algorithms is necessary for developing accurate ANN models for flood prediction.

A. Training the network

The data sets required for the study were collected for a 10-year period (2005-2015) with daily time step. The collected data sets were pre-processed to remove any errors or inconsistencies and missing data were filled using appropriate techniques such as linear interpolation [8]. The feed-forward backpropagation architecture is selected for the ANN model based on its proven effectiveness in similar studies. The pre-processed data sets were split into three categories, i.e., training, validation, and testing, using a ratio of 70:15:15 respectively [5].

After selecting the feed-forward backpropagation architecture for the ANN model, the number of hidden layers is optimized to improve the accuracy of the model. To achieve this, a graph was plotted with the number of hidden layers on the X-axis and mean squared error (MSE) on the Y-axis. The MSE is computed for different numbers of hidden layers, and the hidden layer number corresponding to the lowest MSE was selected as the optimal number of hidden layers for training the ANN model. This number is then fixed, and the ANN model is trained using the Levenberg-Marquardt algorithm with the selected number of hidden layers [5]. The hyperparameters of the ANN model are also optimized, including the learning rate, momentum, and number of neurons in the hidden layers, to further improve the performance of the model [7].

Finally, the performance of the trained ANN model is evaluated using the testing data sets, and the ANN model with the best performance metrics was selected as the optimal model for predicting observed discharge based on present-day rainfall input. To evaluate the performance of the ANN model, several evaluation metrics are used, including mean squared error (MSE), root mean squared error (RMSE), and coefficient of determination (R^2) [10].

a) ANN with present day rainfall as input and discharge as output: The selection of present-day rainfall as the input parameter is grounded in its direct influence on the immediate hydrological response of the watershed. Rainfall, as a key driver of streamflow, captures the real-time precipitation dynamics that impact the river's discharge. By leveraging this instantaneous meteorological parameter, the ANN model aims to discern intricate patterns and relationships, crucial for precise streamflow predictions [12].

b) ANN with present day rainfall, previous day rainfall as input and discharge as output : The input to the ANN model is present-day rainfall and previous day rainfall which is a critical variable for predicting river discharge. The amount of rainfall and its intensity directly affect the water level in rivers and other water bodies, which in turn affects the likelihood of flooding. The inclusion of both present-day and previous-day rainfall as input parameters recognizes the temporal dynamics inherent in hydrological processes. Rainfall from the immediate past, coupled with the current day's precipitation, collectively influences the watershed's response and subsequent discharge. By incorporating this temporal aspect into the ANN model, we aim to capture the lagged effects of rainfall on streamflow, providing a more nuanced understanding of the hydrological system.The output of the ANN model is observed discharge, which is an essential indicator of hydrological processes.

c) ANN with present day rainfall rainfall, previous day rainfall, pre-previous day rainfall as input and discharge as output : The input to the ANN model is present-day rainfall, previous day rainfall and previous day rainfall which is a critical variable for predicting river discharge. These parameters capture the precipitation pattern in the watershed and its effect on the discharge. The inclusion of previous day and pre-previous day rainfall in addition to present day rainfall helps to capture the lag time of the basin, which is the time taken for water to flow from the upper reaches of the watershed to the outlet. The lag time is an important factor as it determines the time delay between rainfall and peak discharge. Hence, the inclusion of lagged rainfall data helps to capture the effect of rainfall on the discharge not only on the present day but also in the previous days. The inclusion of a multivariate rainfall approach recognizes the temporal progression of precipitation's influence on streamflow. By considering not only the current-day rainfall but also the rainfall from the two preceding days, the model captures the cumulative impact of precipitation, addressing potential lag effects. This temporal sensitivity enhances the model's ability to discern complex relationships within the hydrological system.Overall, the input parameters used in the ANN model provide a comprehensive representation of the rainfall-runoff relationship in the watershed.

B. *Lag Time of the basin*

In flood prediction and forecasting, the lag time of the basin is an important factor as it refers to the time taken for rainfall to reach the river and influence its water level and discharge. In general, the longer the lag time, the slower the response of the river to rainfall.
Maximum Potential Retention (S) is calculated (1) and thereby time of concentration of the basin can also be calculated (2)

$$S = \frac{25400}{CN} - 254 \qquad (1)$$

$$T_c = \frac{L^{0.8}(S+1)^{0.7}}{1140\, Y^{0.5}} \qquad (2)$$

Where L is the flow length in ft,Y is the average watershed land slope in percentage, and CN is the curve number.

The lag time of the basin is calculated and determined to be 2.88 days, which is approximated to 3 days. Thus, the input variables for the ANN model are selected as present day rainfall, previous day rainfall, and pre-previous day rainfall, which are considered significant for predicting discharge.
By including present day rainfall, previous day rainfall, and pre-previous day rainfall as input variables, the model can capture the short-term and long-term effects of rainfall on river discharge, as well as the cumulative effect of rainfall over multiple days. This is particularly relevant for basins with longer lag times, as the model can capture the delayed response of the river to rainfall.

V. PERFORMANCE EVALUATION OF THE ANN MODELS

a) ANN with present day rainfall as input and discharge as output: Based on the performance metrics obtained from the ANN model with present day rainfall as input, it can be concluded that the model is performing moderately well [13]. In Fig.5 , the training r value of 0.63008 indicates that the model has captured 63.01% of the variation in the data during training. The validation r value of 0.61499 indicates that the model has captured 61.50% of the variation in the data during validation. The testing r value of 0.60189 indicates that the model has captured 60.19% of the variation in the data during testing.

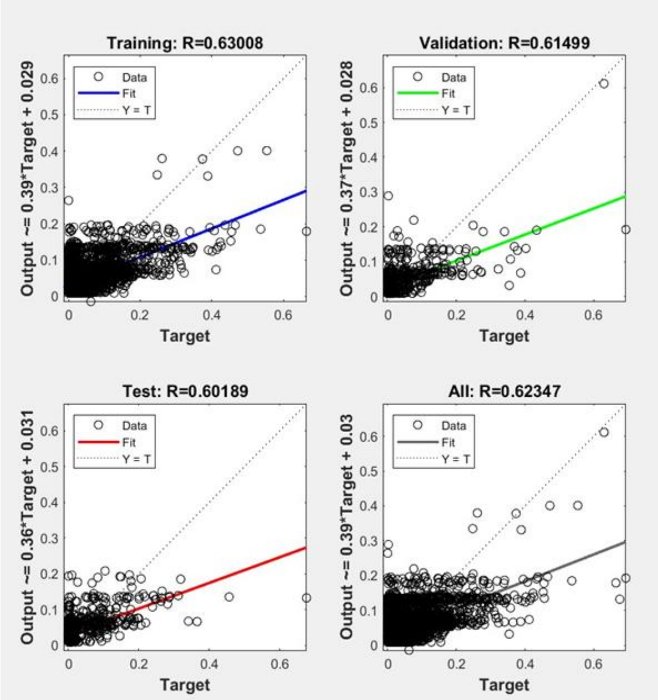

Fig.5. Performance of the ANN model with respect to training set, validation set, testing set and all the data

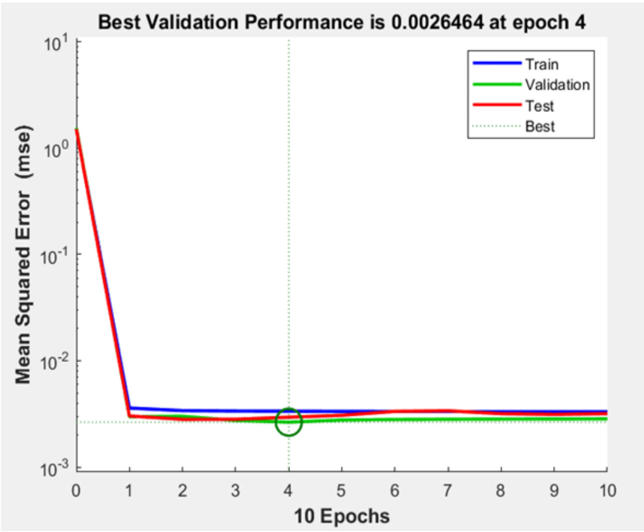

Fig.6. Best validation performance

The total r value of 0.62437 indicates that the model has captured 62.44% of the variation in the data overall. While this is not a very high percentage, it is still a reasonable performance for the model. Fig. 6 shows the best validation performance of 0.0026464 at epoch 4 which suggests that the model achieved its best performance during the fourth iteration of training. Overall, the performance analysis indicates that the ANN model with present day rainfall as input has moderate predictive power and could be used as a basis for further analysis and improvement.

This model was used to predict the streamflow of a flood event that occurred on August 16, 2018. However, when the model was used to predict the streamflow value, it produced a result of 470.6619, whereas the actual discharge value was 668.063. This indicates that there was a significant difference between the predicted and actual values. It is important to analyze the reasons behind this discrepancy and assess the performance of the model in such situations. The model did not capture all the variables or factors that influenced the flood event. Further analysis is required to improve the performance of the model.

b) ANN with present day rainfall, previous day rainfall as input and discharge as output : Based on the performance metrics obtained from the ANN model with present day rainfall and previous day rainfall as input, it can be concluded that the model is performing moderately well.

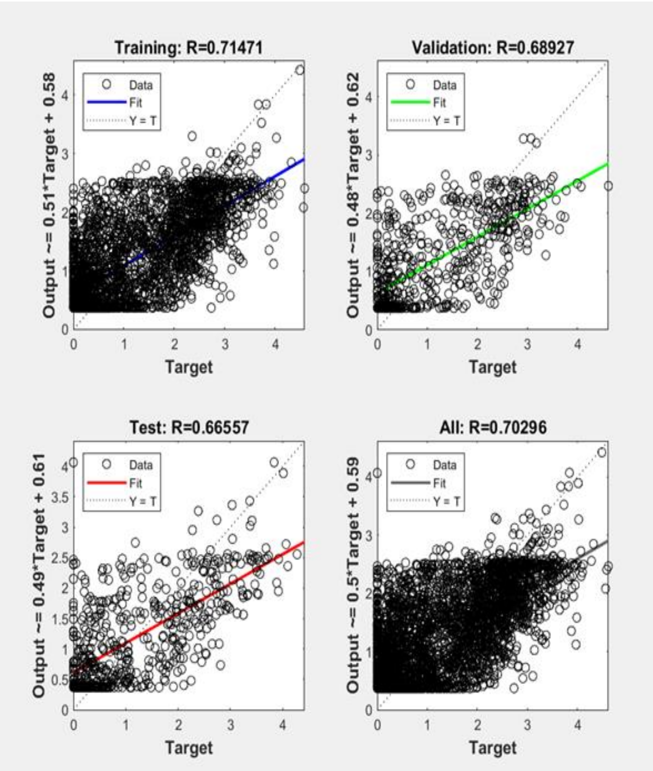

Fig.7. Performance of the ANN model with respect to training set, validation set, testing set and all the data

In Fig.7, the training r value of 0.71471 indicates that the model has captured 71.47% of the variation in the data during training. The validation r value of 0.68927 indicates that the model has captured 68.93% of the variation in the data during validation. The testing r value of 0.66557 indicates that the model has captured 66.56% of the variation in the data during testing.

The total r value of 0.70296 indicates that the model has captured 70.296% of the variation in the data overall. While this is not a very high percentage, it is still a reasonable performance for the model. Fig.8 shows the best validation performance of 0.61163 at epoch 23 which suggests that the model achieved its best performance during the twenty third iteration of training. Overall, the performance analysis indicates that the ANN model with present day rainfall and previous day rainfall as input has moderate predictive power and could be used as a basis for further analysis and improvement.

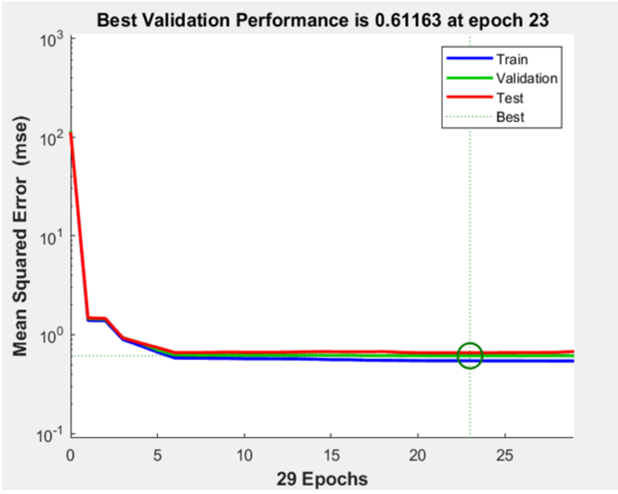

Fig.8. Best validation performance

This model was used to predict the streamflow of a flood event that occurred on August 16, 2018. However, when the model was used to predict the streamflow value, it produced a result of 479, whereas the actual discharge value was 668.063. This indicates that there was a significant difference between the predicted and actual values. The model did not capture all the variables or factors that influenced the flood event. Further analysis is required to improve the performance of the model.

c) ANN with present day rainfall rainfall, previous day rainfall, pre-previous day rainfall as input and discharge as output : The results of the ANN model trained with present day rainfall, previous day rainfall and pre previous day rainfall as input, showed a training r value of 0.90936 indicates a strong correlation between the predicted and actual values during the training phase (Fig.9). The validation r value of 0.91218 and test r value of 0.89658 also indicate a good fit of the model with the validation and testing data. The overall r value of 0.90764 shows that the model is performing well across all phases of the analysis.

In Fig. 10, the best validation performance of 0.18293 at epoch 26 indicates that the model is reaching its optimum performance level at an early stage of the training phase. Overall, the performance evaluation suggests that the ANN model is capable of accurately predicting the output based on the input parameters used in this study. The model showed promising results in predicting the discharge of the flood event, based on the input rainfall data from three consecutive days which is also lag time of the basin.

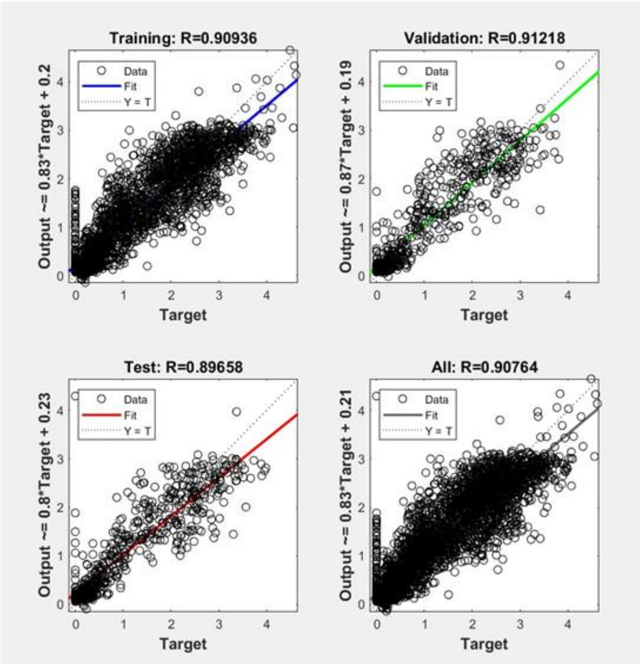

Fig.9. Performance of the ANN model with respect to training set, validation set, testing set and all the data

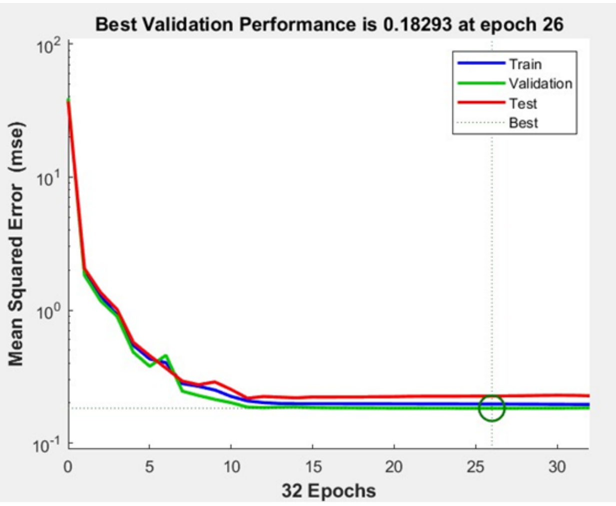

Fig.10. Best validation performance

VI. Conclusion

This study culminates in a significant advancement in streamflow prediction, spotlighting the intricate relationship between the lag time of the basin and the input variables of an Artificial Neural Network (ANN) model. The optimized configuration, integrating present-day rainfall, previous day rainfall, and pre-previous day rainfall as inputs, emerges as the key driver behind the model's exceptional predictive power.

The revelation of a tangible link between these input variables and the lag time of the basin is pivotal. The identified lag time of 3 days mirrors the temporal responsiveness of the watershed to precipitation events. This finding not only underscores the accuracy of the ANN model in capturing the nuanced dynamics of the hydrological system but also establishes a direct and meaningful correlation between the chosen inputs and a fundamental watershed characteristic. The implications of this correlation extend beyond model performance. It provides a deeper understanding of how the temporal sequence of rainfall events influences the basin's response, shaping the lag time. This newfound knowledge holds practical significance for water resource management, where a precise grasp of the temporal dynamics enhances the efficacy of decision-making processes.

In conclusion, the optimized ANN model not only excels in streamflow prediction but also unravels a crucial connection between input variables and the lag time of the basin. This holistic insight enhances our ability to comprehend and manage hydrological systems, paving the way for more informed and effective strategies in water resource planning and environmental conservation.

References

[1] Abhijit Paul & Prodipto Das, "Flood Prediction Model using Artificial Neural Network," International Journal of Computer Applications Technology and Research, pp. 473 – 478, 2022

[2] Ahmed M. Tawfik, "River flood routing using artificial neural networks,"Journal of hydrology, 14(3), pp 1000-1018, 2022

[3] Biragani, Y.T, "Flood forecasting using artificial neural networks : an application of multi-model data fusion technique," Journal of Hydraulic Structures, pp 62–73, 2016

[4] Demb´el´e, M. & Zwart, S.J., "Evaluation and comparison of satellite-based rainfall products in Burkina Faso, West Africa," International Journal of Remote Sensing, pp 3995–4014, 2016

[5] Habtamu Tamiru & Megersa O. Dinka, "Application of ANN and HEC-RAS model for flood inundation mapping in lower Baro Akobo River Basin, Ethiopia," Journal of hydrology, pp 191-209, 2021

[6] Masoud Bakhtyari Kia, Saied Pirasteh, Biswajeet Pradhan & Ahmad Rodzi Mahmud, "An artificial neural network model for flood simulation using GIS: Johor River Basin, Malaysia," Journal of Environmental Earth Science, pp 251–264, 2012

[7] Riad, S., Mania, J., Bouchaou, L., & Najjar, Y., "Rainfall-runoff model using an artificial neural network approach," Journal of hydrology, pp 839–846, 2004

[8] R. Peters, G. Schmitz, J. Cullmann, ''Flood routing modelling with Artificial Neural Networks," Journal of Hydrology and Meteorology, pp 3-11, 2006

[9] Sulafa Hag Elsaf, "Artificial Neural Networks (ANNs) for flood forecasting at Dongola Station in the River Nile, Sudan," Alexandria Engineering Journal 53, pp 655–662, 2014

[10] Tegegne, G., Park & D.K., Kim, Y, "Comparison of hydrological models for the assessment of water resources in a data scarce region, the Upper Blue Nile River Basin," Journal of Hydrology : Regional Studies, pp 49–66, 2017

[11] Tim Hill, Leorey Marquezb, "Artificial neural network models for forecasting and decision making", International Journal of Forecasting, pp 5-12, 1994

[12] Toth, E., Brath, A. & Montanari, A., "Comparison of short-term rainfall prediction models for real-time flood forecasting," Journal of Hydrology, pp 132–147, 2000

[13] 13. Van S.P., Le, H.M., Thanh D.V., & Dang T.D, "Deep learning convolutional neural network in rainfall-runoff modelling," Journal of Hydroinformatics, pp 541–561, 2020