

Daily Rainfall Prediction for Bihar Using Artificial Neural Networks

AMIT KUMAR¹, MAN MOHAN DEO^{2*}, PAWAN JEET³, ARTI KUMARI³ AND OM PRAKASH⁴

ABSTRACT

Accurate daily rainfall prediction is required for accurate stream flow prediction, flooding risk analysis and construction of reliable flood control and early warning system. However, because of its nonlinearity, the prediction of daily rainfall with high accuracy and long prediction lead time is difficult. In this study, the artificial neural network (ANN) model was applied to predict the daily rainfall using six meteorological parameters (minimum and maximum temperature, morning and evening relative humidity, pan evaporation and rainfall) for a period of 2015-19 were used. The three ANN models were trained and tested by using 1 (ANN-1), 3 (ANN-3) and 5 (ANN-5) days preceding meteorological parameters. For the development of the ANN models, 70% of data is chosen for the training process, and the remaining 30% of data is chosen for the testing phase. It was found that the ANN-3 and ANN-5 were a promising algorithm to predict daily rainfall. The model trained with one-day preceding information was a very poor performance to predict daily rainfall. The ANN-5 model had RMSE, MARE and MAE as 0.42, 0.04, 0.10 for training and 1.60, 0.07, and 0.24 for testing, respectively, which was the lowest as compared to ANN-1 and ANN-3 models. Nonetheless, the NSE and KGE were greater than 0.92 for both training and testing. The PBias were only positive for the ANN-5 model which was 1.35 and 3.90 for training and testing, respectively. The ANN model would be helpful in the quick and accurate prediction of daily rainfall.

Keywords: Rainfall, Streamflow, Prediction, ANN model, Flood control

ARTICLE INFO

Received on	:	18.09.2022
Accepted	:	11.12.2022
Published online	:	29.12.2022



INTRODUCTION

Water resource management involves the planning, development, and distribution of water resources. These activities are either directly or indirectly related to rainfall prediction (Altunkaynak and Nigussie 2015, Roushangar et al. 2018). Rainfall is an important basis for assessing water resources, agriculture, ecosystems and hydrology and also one of the key influencing components of the hydrological cycle, plays a significant role in runoff modeling. Thus, how to accurately predict rainfall has been a very crucial issue in the weather forecast community worldwide to identify the potential for heavy rainfall and flash flooding, as well as provide information for hydrologic interests (Hall et al. 1998, Ojo and Ogunjo 2022, Yen et al. 2019). Because of the dynamic changes in atmospheric processes, developing a rainfall prediction model is difficult. Rainfall is a stochastic procedure whose forthcoming event is contingent on some predecessors from other constraints such as the sea surface temperature for monthly to seasonal time scales, the surface pressure for weekly to daily time scales and other atmospheric constraints for daily to hourly time scales. The unpredictability of weather and climatic aspects, particularly those atmospheric constraints will be the major force for daily precipitation events. If unpredictability patterns could be documented and used for future paths, the feasibility of daily rainfall prediction is very much possible.

In recent years, the use of artificial intelligence algorithm proposed by artificial neural network (ANN) to solve

problems in hydrology has attracted considerable attention. The mechanism of hydrological parameters forecasting is a nonlinear system in terms of mathematics (Yen et al. 2019). ANNs is a powerful computing system for highly complex and nonlinear systems (Ghumman et al. 2011). ANN belongs to the black box time series models and offers a relatively flexible and quick means of modeling. These models can treat the nonlinearity of system to some extent due to their parallel architecture. A few studies reported poor performances of ANN models in comparison with the conventional ones. A few studies reported poor performances of ANN models in comparison with the conventional ones. For example, Gaume and Gosset (2003) compared feed forward ANNs with a linear model and a conceptual model (GR4J). They concluded that their conceptual model outclassed the linear and ANN models. Some other studies show the successful applications of ANN models in the simulation of future runoffs with high degree of accuracy (Agarwal and Singh 2004). Castellano-Méndez et al. (2004) compared ANN and Box and Jenkins techniques and found that ANN is an improvement on Box and Jenkins model. Sohail et al. (2008) compared ANN with MARMA (multivariate auto regressive moving average models) models in a small watershed of Tono area in Japan during wet and dry seasons. They concluded that ANN models show better results during wet seasons when the nonlinearity of rainfall-runoff process is high. Lee et al. (2018) ANN was used to create a late spring-early summer rainfall

¹ ICAR-Indian Agricultural Research Institute, New Delhi-110 012, India

² ICAR- Indian Institute of Pulses Research, Kanpur-208024, India

³ ICAR-Research Complex for Eastern Region, Patna-800 014, India

⁴ ICAR - Central Arid Zone Research Institute, Jodhpur-342003, India

*Corresponding Author E-mail: anandi888@gmail.com

forecasting model and found that the ANN model successfully predicts rainfall in Geum river basin, Korea. [Ejig and Nigatu \(2021\)](#) also proposed ANN and KNN (K-Nearest Neighbour) for rainfall prediction using three parameters *viz.* maximum temperature, minimum temperature, and average rainfall for crop recommendations in Ethiopia. However, both models did not forecast for all Ethiopian seasons.

Accurate daily rainfall prediction is required for accurate streamflow prediction, flooding risk analysis, and constructing a reliable flood control and early warning system. However, because of its nonlinearity, the prediction of daily rainfall with high accuracy and long prediction lead

time is difficult. The main purpose of this study is to develop a forecasting model of rainfall using meteorological data to investigate the possible factors in governing the rainfall forecast in the agro-climatic zone-1 of Bihar, India.

MATERIAL AND METHODS

Study area and data collection

Study area Bihar has a geographical area of 9.42 M ha. It is a landlocked state, located in the eastern part of India (Figure 1) and lies between the latitudes of 24° 20' 30" and 27° 31' 15" N and the longitudes of 83° 19' 50" and 88° 17' 40" E. The climate of Bihar is mostly subtropical, with three distinct seasons: winter, summer, and monsoon. Bihar is divided into three main agro-climatic zones. Zones I and II are in North Bihar, a flood-prone region, while agro-climatic zone III is in the drought-prone South Bihar ([Radda et al. 2021](#)). The daily meteorological for a period of 2015-19 of agro-climate zone-1 of Bihar was downloaded from <https://www.rpcau.ac.in/downloads/> (last accessed 20th July, 2022). The meteorological data included rainfall (mm), min. and max. temperature (°C), morning and evening relative humidity (%) and pan evaporation (mm).

Data Preprocessing

Initially, the data set contains some missing values. Thus, preprocessing data is very important for the accuracy of the model. The missing values are dropped and replaced with the mean value. Data scaling and normalization transform the data into a standardized form. Normalization helps to scale the data of an attribute so that it falls in a smaller range between 0 to 1 or -1 to 1. The min-max normalization method has been used for this study. Input data were normalized using the formula stated below in Equation 1.

$$Y_{normalized} = \frac{Y - Y_{min}}{Y_{max} - Y_{min}} \quad \dots(1)$$

where Y represents normalized data, Y is the actual value of rainfall data to be normalized, Y_{min} represents minimum value of rainfall data, represents Y_{max} value of rainfall data, respectively.

For the development of the ANN models, 70 % of processed data was chosen for the training process, and the remaining 30 % of data was chosen for the validation phase.

ANN model development steps and architecture

The most commonly used ANN in hydrological predictions is a feed-forward network with the back propagation (BP) training algorithm. The BP algorithm is a first-order gradient search method, which is capable of non-linear pattern recognition and memory association. Standard multi-layer feed-forward networks can approximate any measurable function to any desired degree of accuracy. The term 'bias' used in ANN models represents an adjustable parameter of the neurons that is the difference between the ANN-estimated and the observed output. For better model development, the sum of the biases for the entire data should be small. In the present study, the feed-forward neural network with BP training algorithm is used for the development of the ANN models.

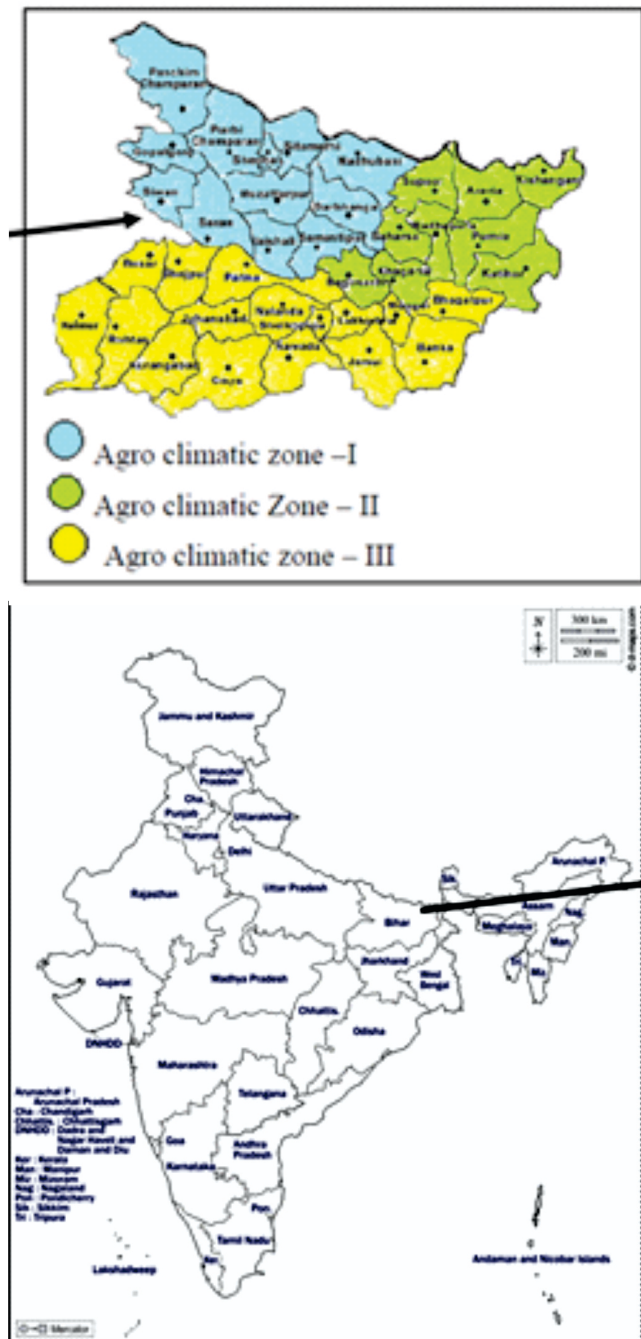


Fig. 1: Agro-climatic zones in Bihar, India ([Parthsarthy and Singh 2013](#))

In practice, the ANN architecture consists of an input layer, intermediate layers (hidden layer) and an output layer. The hidden layers may be one or more depending on the data type and the model error statistics. Also, the number of nodes in the hidden layer plays a significant role in ANN model performance (Sarangi and Bhattacharya 2005). An unresolved issue in applying ANNs to the modeling of the rainfall-runoff process is the architecture that should be used to map the process effectively. The input vectors to the selected ANN model, the number of hidden layers, the learning rule and the number of output vectors have an impact on the model's performance. There are no fixed rules for developing an ANN and a general framework is followed based on previous successful applications in engineering. Based on such successes, the number of neurons in the input layer and hidden layers are the same as with a single output layer as the rainfall approach was adopted to select the optimal ANN architecture in the present analysis. Three different combinations of input parameters and the three number of hidden layers with a single output (rainfall) were tried. The input parameters were 1, 3 and 5 days preceding meteorological data (Viz.6, 18 and 30 neurons in input and hidden layers) in the ANN models to be chosen as the criterion for the selection of optimal architecture. The general architecture for ANN is shown in Figure 2.

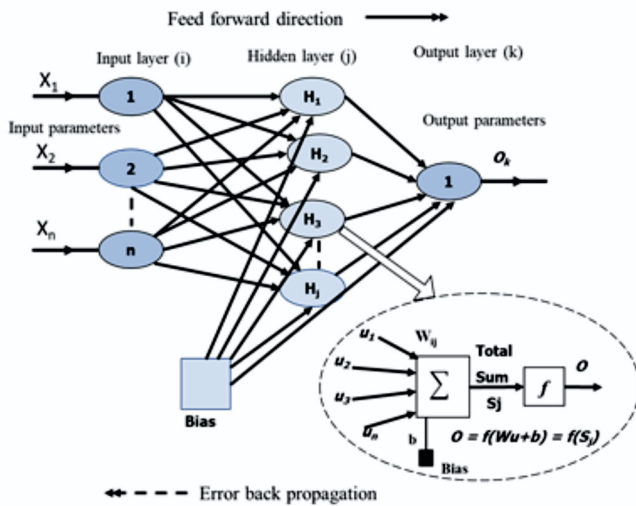


Fig. 2: Artificial neural network architecture for rainfall prediction

Each neuron has a number of input arcs connected (Figure 2), u_1 to u_n , and associated with each i , there is a weight W_{ij} which represents a factor by which a value passing to the neuron is multiplied. A neuron sums the values of all inputs as:

$$S_j = \sum W_{ij} + b \quad \dots(2)$$

In Fig. 2, Wu corresponds to the summation of weights W_{ij} . The term b is called bias. Finally, an activation function is applied to S_j to provide the final output from the neuron. When a BP training algorithm is used for training a network, the sigmoid activation function is most often used (Sivakumar *et al.* 2002). The sigmoid function is bounded above and below (0 and 1), is continuous and differentiable everywhere. The sigmoid function (w) is given by

$$\varphi(S_j) = \frac{1}{1 + e^{-S_j}} \quad \dots(3)$$

where S_j is the value of the neuron at j^{th} location.

Model evaluation statistics

The statistical criteria, such as the root mean square error (RMSE), mean absolute relative error (MARE), mean absolute error (MAE), Nash-Sutcliffe Efficiency (NSE), Kling–Gupta efficiency (KGE), percent bias (PBias) and coefficient of determination (R^2) were used to ascertain the performance of the developed model.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_{est_i} - \bar{x}_{est_i})^2} \quad \dots(4)$$

$$MARE = \frac{1}{n} \sum_{i=1}^n \left| \frac{x_{est_i} - x_{obs_i}}{x_{est_i}} \right| \times 100 \quad \dots(5)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |x_{est_i} - x_{obs_i}| \quad \dots(6)$$

$$NSE = 1 - \frac{\sum_{i=1}^n (x_{est_i} - x_{obs_i})^2}{\sum_{i=1}^n (x_{obs_i} - \bar{x}_{obs})^2} \quad \dots(7)$$

$$KGE = 1 - \sqrt{(1 - r)^2 + (\beta - 1)^2 + (\gamma - 1)^2} \quad \dots(8)$$

$$r = \frac{\sum_{i=1}^n (x_{obs_i} - \bar{x}_{obs_i})(x_{est_i} - \bar{x}_{est})}{\sqrt{\sum_{i=1}^n (x_{obs_i} - \bar{x}_{obs_i})^2} \sqrt{\sum_{i=1}^n (x_{est_i} - \bar{x}_{est})^2}} \quad \dots(9)$$

$$\beta = \bar{x}_{est_i} / \bar{x}_{obs_i}; \quad \gamma = \frac{\sigma_{est} / \bar{x}_{est_i}}{\sigma_{obs} / \bar{x}_{obs_i}} \quad \dots(10)$$

where, est and obs are the standard deviation of the estimated and observed rainfall, respectively.

$$PBias = \left[\sum_{i=1}^n (x_{est_i} - x_{obs_i}) / \sum_{i=1}^n x_{obs_i} \right] \times 100 \quad \dots(11)$$

$$R^2 = \frac{\sum_{i=1}^n (x_{est_i} - \bar{x}_{est})(x_{obs_i} - \bar{x}_{obs_i})}{\sum_{i=1}^n (x_{est_i} - \bar{x}_{est})^2 \sum_{i=1}^n (x_{obs_i} - \bar{x}_{obs_i})^2} \quad \dots(12)$$

where n is the number of observations, X_{obs_i} and X_{est_i} are observed and predicted value, respectively, \bar{X}_{obs_i} and \bar{X}_{est_i} are the average values of observed and predicted data, respectively.

Open-source software

In present study open-source software were used. The programming language of choice is Python 3.7.14 (VanRossum and DrakeJ 1995). The libraries we use for preprocessing our data and for data management in general are Numpy (Harris *et al.* 2020), Pandas (McKinney 2010) and Scikit-Learn (Pedregosa *et al.* 2011). The Deep-Learning frameworks we use are TensorFlow (Abadi *et al.* 2016) and Keras (Chollet 2015). The figures are made using Matplotlib (Hunter2007).

RESULTS AND DISCUSSION

The daily rainfall of agro-climatic zone-1 of Bihar was predicted using 1, 3 and 5 days preceding meteorological information. The three ANN models were used with the same number of neurons in input and hidden layers with one output layer as rainfall. There were six input variables for ANN-1 model as meteorological data *i.e.*, one-day previous information used to predict the rainfall. Similarly, 15 and 30 input variables for 3- and 5-days preceding information were

used in ANN-3 and ANN-5 models, respectively. The output of the ANN-1 model for training and testing of as shown in Fig. 3. It could be observed that the ANN-1 model was underestimated for both the training and testing phases. Moreover, the coefficient of determinations was also very low for both training (0.48) and testing (0.14), respectively as shown in Fig. 4. Moreover, ANN-3 model predicted close to the observed values for both training and testing as shown in Fig. 5. Besides this, it could be also observed that the extreme events were underestimated. The coefficient of determinations was good 0.98 and 0.96 for training and testing, respectively as shown in Fig. 6. The ANN-5 model was trained and tested using 5 days of meteorological information predicted very close to the observed rainfall for training and testing (Fig. 7). The coefficient of determinations was good

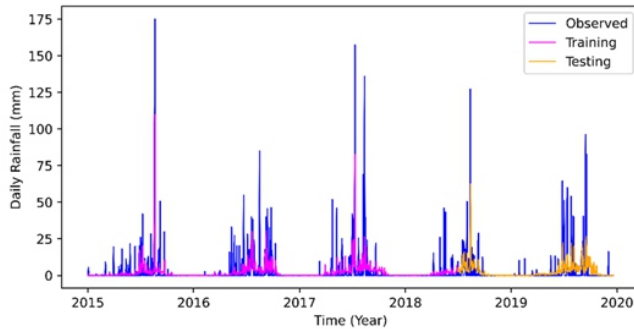


Fig. 3: Time series for observed and estimated rainfall (ANN-1) using one day preceding meteorological data

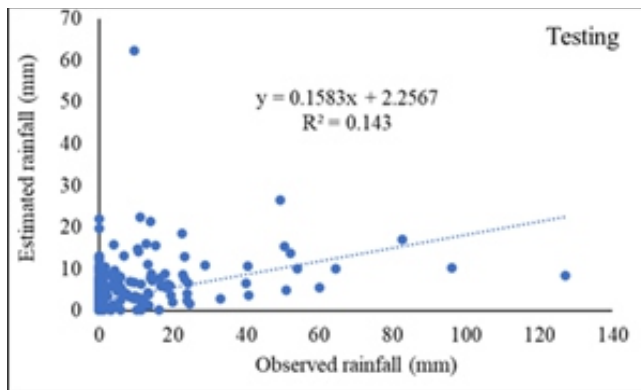
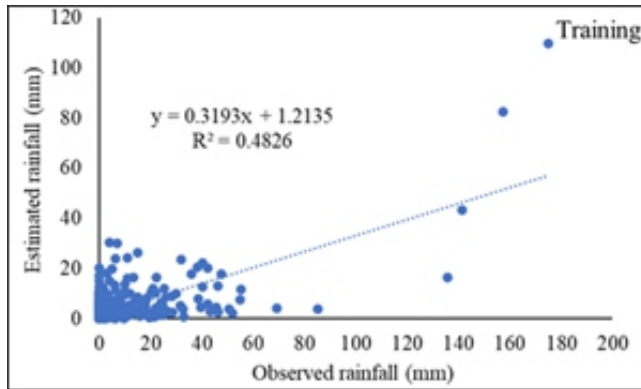


Fig. 4: Scatter plot of observed versus estimated rainfall values for the training and the testing (ANN-1) using one day preceding meteorological data

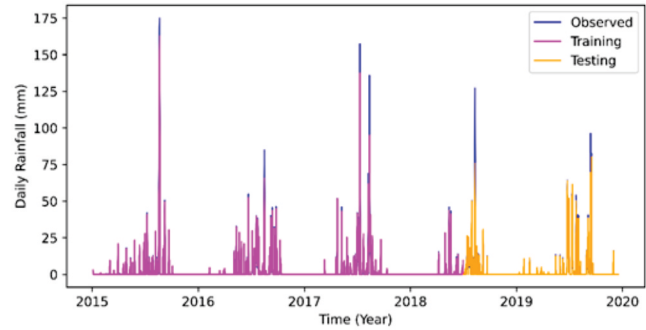


Fig. 5: Time series for observed and estimated rainfall (ANN-3) using three days preceding meteorological data

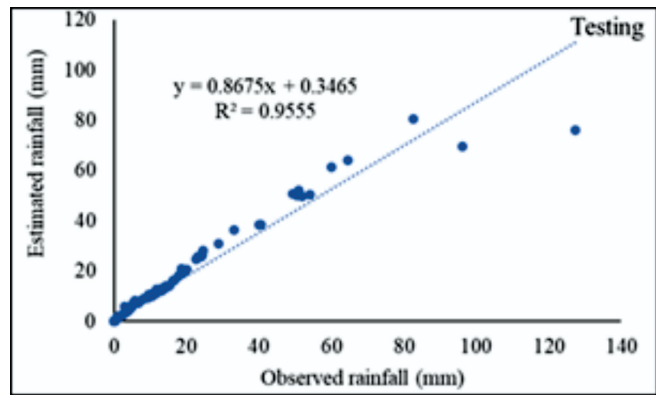
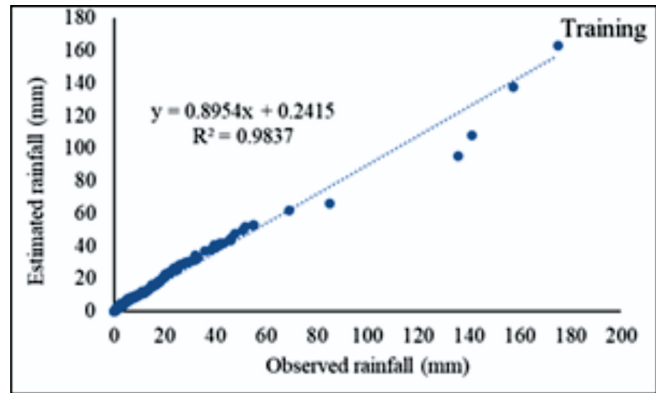


Fig. 6: Scatter plot of observed versus estimated rainfall values for the training and the testing (ANN-3) using one day preceding meteorological data

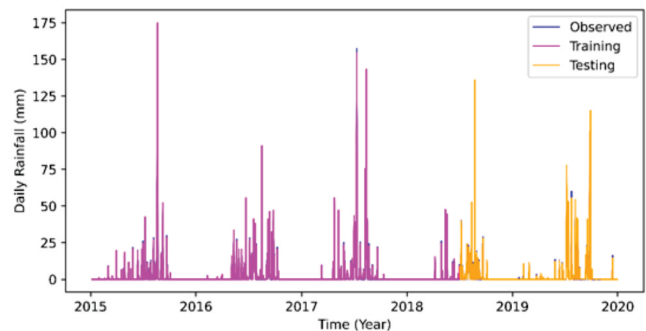


Fig. 7: Time series for observed and estimated rainfall (ANN-5) using five days preceding meteorological data

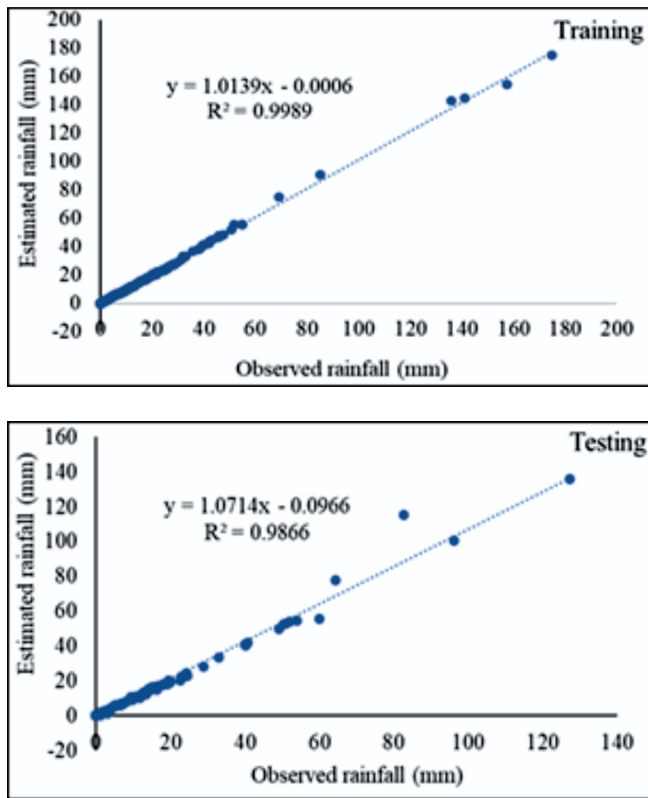


Fig. 8: Scatter plot of observed versus estimated rainfall values for the training and the testing (ANN-5) using five days preceding meteorological data

0.99 and 0.99 for training and testing, respectively as shown in Fig. 8. Nonetheless, ANN-5 also precisely simulated the extreme events for both training and testing. Overall, it was observed that the ANN has the potential to simulate the daily rainfall.

The performance error statistics of ANN models are shown in Table 1. The ANN-1 model trained with one day preceding meteorological data had poor performance to estimate the daily rainfall of agro-climatic zone-1 of Bihar, as compared to models trained using three and five days preceding meteorological data (ANN-1 & ANN-5). The RMSE, MARE, and MAE were 8.55, 1.43, 2.33 for training and 10.32, 1.41, 2.90 for testing, respectively for the ANN-1 model. Besides this, the NSE, KGE and PBias were also negative for training (-1.72, -1.02, -26.34) and testing (-3.90, -0.52, -12.15), respectively. The

Table 1: Model performance error statistics

Preceding data used	ANN models	RMSE	MARE	MAE	NSE	KGE	PBias	R ²
ANN-1	Training	8.55	1.43	2.90	-1.72	-0.24	-26.34	0.48
	Testing	10.32	1.41	3.87	-3.90	-0.52	-12.15	0.14
ANN-3	Training	1.76	0.09	0.23	0.97	0.89	-1.07	0.98
	Testing	2.56	0.11	0.32	0.93	0.867	-2.00	0.96
ANN-5	Training	0.42	0.04	0.10	0.99	0.98	1.35	0.99
	Testing	1.60	0.07	0.24	0.98	0.92	3.90	0.99

ANN-3 and ANN-5 model trained with 3- and 5-day meteorological data has better performance of error statistics. Moreover, the ANN-5 model had RMSE, MARE and MAE *viz.* 0.42, 0.04, 0.10 for training and 1.60, 0.07, and 0.24 for testing, respectively, which was the lowest as compared to ANN-1 and ANN-3 models. Nonetheless, ANN-54 had NSE and KGE were greater than 0.92 for both training and testing. The PBias were only positive for the ANN-5 model which were 1.35 and 3.90 for training and testing. [CommeH *et al.* \(2022\)](#) classified the model for the group based on NSE and PBias *viz.* very good (0.75<NSE1; PBias<10), good (0.65<NSE0.75; PBias<15), satisfactory (0.5<NSE0.65; PBias<25) and unsatisfactory (NSE0.5; PBias 25). [Suryaningtyas *et al.* \(2020\)](#) also described based on correlation coefficient (R²) as very low (0-0.19), low (0.20-0.39), moderate (0.40-0.59), strong (0.60-0.79) and very strong (0.80-1). Thus, the ANN-3 and ANN-5 models had NSE close to 1 and PBias<10 for both training and testing along with the correlation coefficient was also greater than 0.96. But ANN-3 did not predict the extreme event of rainfall for both training and testing (Figure 5), while the ANN-5 model predicted very well (Fig. 7). Globally increasing intensity and frequency of extreme rainfall events demand reliable early warning systems. Despite significant improvements in the skills of weather models, the state-of-art extreme rainfall forecasts, at a sufficient lead time, still suffer from high biases, high uncertainties, low hit rates, and high false alarms ([Tripathy *et al.* 2021](#)). Extreme rainfall events are of great concern as they often lead to excessive river flows and severe flooding in the Kosi River basin, North Bihar is often exposed to extreme rainfall events and associated flooding, with significant impacts on human lives, properties, infrastructure, agriculture, and livestock. Therefore, the ANN-5 model developed using 5 days preceding meteorological data predicted as closed to the observed along with extreme rainfall events, so the artificial neural network has great potential to predict the rainfall by using meteorological data.

CONCLUSION

The meteorological data consists six weather parameters such as min. and max. temperature, morning and evening relative humidity, pan evaporation and rainfall for a period of 2015-19 were used. For the development of the ANN models 70 % and 30% were chosen for the training testing process, respectively. The three ANN models were trained and tested by using 1 (ANN-1), 3 (ANN-3) and 5 (ANN-5) days preceding meteorological parameters. The training and testing results of

ANN models indicated that the ANN-3 and ANN-5 had a promising algorithm to predict daily rainfall of agroclimatic zone-1 (Bihar). The performance of ANN-1 was a very poor to predict daily rainfall.

The ANN-5 model had better performance for lower as well extreme event of rainfall with RMSE, MARE and MAE as 0.42, 0.04, 0.10 for training and 1.60, 0.07, 0.24 for testing,

REFERENCES

- Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, and Zheng X. 2016. Tensorflow: Large-scale machine learning on heterogeneous distributed systems. *arXiv preprint arXiv:1603.04467*.
- Agarwal A and Singh R D. 2004. Runoff modelling through back propagation artificial neural network with variable rainfall-runoff data. *Water resources management* **18**(3): 285-300.
- Altunkaynak A and Nigussie T A. 2015. Prediction of daily rainfall by a hybrid wavelet-season-neuro technique. *Journal of Hydrology* **529**:287-301.
- Castellano-Méndez M, González-Manteiga W, Febrero-Bande M, Prada-Sánchez J M and Lozano-Calderón R. 2004. Modelling of the monthly and daily behaviour of the runoff of the Xallas river using Box-Jenkins and neural networks methods. *Journal of hydrology* **296**(1-4): 38-58.
- Chollet F and *et al.* 2015. Keras. GitHub. Retrieved from <https://github.com/fchollet/keras>.
- Comme M K, Agyei-Agyemang A, Tawiah P O and Asaaga, B A. 2022. CFD analysis of a flat bottom institutional cookstove. *Scientific African* **16**: e01117.
- Ejigu Y B and Nigatu H M. 2021. Rainfall Prediction and Cropping Pattern Recommendation Using Artificial Neural Network. In *International Conference on Advances of Science and Technology: Springer*, p:500-516.
- Gaume E and Gosset R. 2003. Over-parameterisation, a major obstacle to the use of artificial neural networks in hydrology? *Hydrology and Earth System Sciences* **7**(5): 693-706.
- Ghumman A R, Ghazaw Y M, Sohail A R and Watanabe K. 2011. Runoff forecasting by artificial neural network and conventional model. *Alexandria Engineering Journal* **50**(4): 345-350.
- Hall T, Brooks H E and Doswell C A. 1999. Precipitation forecasting using a neural network. *Weather and Forecasting* **14**(3):338-345.
- Harris C R, Millman K J, vanderWalt S J *et al.* 2020. Array programming with NumPy. *Nature* **585**:357-362.
- Hunter J D. 2007. Matplotlib: A 2D Graphics Environment, *Computing in Science & Engineering* **9**(3):90-95.
- Lee J, Kim C G, Lee J E, Kim N W and Kim H. 2018. Application of artificial neural networks to rainfall forecasting in the Geum River basin, Korea. *Water* **10**(10):1448.
- McKinney W and *et al.* 2010. Data structures for statistical computing in python. In *Proceedings of the 9th Python in Science Conference*, 445:51-56.
- Ojo O S and Ogunjo S T. 2022. Machine learning models for prediction of rainfall over Nigeria. *Scientific African* **16**: e01246.
- Parthsarathi P and Singh A K. 2013. A simple approach about the characteristics of available surface water in Bihar state of India. *Geosciences* **3**(2):68-76.
- Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M and Duchesnay E. 2011. Scikit-learn: Machine Learning in Python, *J. Mach. Learn. Res.*, **12**:2825-2830.
- Radda I A, Kumar B M and Pathak P. 2021. Land degradation in Bihar, India: an assessment using rain-use efficiency and residual trend analysis. *Agricultural Research* **10**(3):434-447.
- Roushangar K, Alizadeh F and Adamowski J. 2018. Exploring the effects of climatic variables on monthly precipitation variation using a continuous wavelet-based multiscale entropy approach. *Environmental Research* **165**:176-192.
- Sarangi A and Bhattacharya A K. 2005. Comparison of artificial neural network and regression models for sediment loss prediction from Banha watershed in India. *Agricultural Water Management* **78**(3):195-208.
- Sivakumar B, Jayawardena A W and Fernando T M K G. 2002. River flow forecasting: use of phase-space reconstruction and artificial neural networks approaches. *Journal of Hydrology* **265**: 225-245.
- Sohail A, Watanabe K and Takeuchi S. 2008. Runoff analysis for a small watershed of Tono area Japan by back propagation artificial neural network with seasonal data. *Water Resources Management* **22**(1): 1-22.
- Suryaningtyas L S, Ery S and Rispiningtati R. 2020. Hydrological analysis of TRMM (Tropical rainfall measuring mission) data in Lesti sub watershed. *Civil and Env. Science Journal* **3**:18-30.
- Tripathy S S, Karmakar S and Ghosh S. 2021. Hazard at weather scale for extreme rainfall forecast reduces uncertainty. *Water Security* **14**:100106.
- VanRossum G and Drake F L. 1995. Python tutorial. *Centrum voor Wiskunde en Informatica Amsterdam*, The Netherlands.
- Yen M H, Liu D W, Hsin Y C, Lin C E and Chen C C. 2019. Application of the deep learning for the prediction of rainfall in Southern Taiwan. *Scientific Reports* **9**(1): 1-9.

Citation:

Kumar A, Deo MM, Jeet P, Kumari A and Prakash O.2022. Daily rainfall prediction for Bihar using artificial neural networks. *Journal of AgriSearch* **9**(4):320-325