INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY

A rainfall-runoff model using artificial neural networks for the district of Bankura in a time of climate change

Kush Kumar Dey[1]∗, **Vijay Kumar Dwivedi**[2], **Satanand Mishra**[3]

1 Research scholar, Department of Civil Engineering, National Institute of Technology, Durgapur, 713209, India. Tel.: +91 894 287 2286
2 Professor, Department of Civil Engineering, National Institute of Technology, Durgapur, 713209, India
3 Sr. Scientist, CSIR, Advanced Materials and Processes Research Institute (AMPRI), Bhopal, 462064, India

Abstract

Objectives: The main objective of this research is to determine the Hydrological (Rainfall-Runoff) model by using Artificial Neural Networks (ANNs) for the study area. **Methods**: The ANNs model was applied to the relative impact of different climatic variables such as Rainfall, Temperature, Cloud Cover, Potential Evapotranspiration, and Relative Humidity for the district of Bankura located on Lower Gangetic Plain (Zone no-III) in India. This study has also developed runoff hydrograph using various Slope, Rainfall Intensity, and Roughness over the catchment. The researcher has collected the Real-Time data series of 116 years (1901-2016) for the six meteorological stations of district Bankura from India Meteorological Department, Pune. For estimating runoff values, the study has been used Kothyari and Garde equation in which the most important factor i.e. the Vegetal Cover Factor (Fv) was considered. For developing the ANNs model, the available data were separated as 70% for training, 15% for testing, and 15% for validation. **Findings**: The Predicted values using ANNs model are more useful for better estimation of water resources management than previous researches. The model performance was with better efficiency (Nash-Sutcliffe Efficiency) and it was greater than 97%. **Novelty**: First time, this research established the Hydrological (Rainfall-Runoff) model by using Artificial Neural Networks (ANNs) for the study area.

Keywords: Artificial neural networks; climate change; meteorological data; water resources management; runoff; district Bankura

1 Introduction

Rainfall-Runoff plays a major role in Hydrology and Water Resources. ANNs are a physical-based and black box model which is a useful tool in hydrology[1,2]. Nowadays, Artificial Neural Networks (ANNs) have become one of the vital tools for modeling of complex hydrological processes. ANNs draw the relationship among the variables particularly inputs and outputs values[3–5].

In hydrology, the research work is to develop the model to find out the future basin discharge which is a safe and economical tool for hydrological engineering design[6–8]. The water resources management using ANNs which involves the estimation of Rainfall-Runoff event, river discharge forecasting, Climate Change, river inflow modeling and estimation of groundwater, etc[9–11]. The ANNs were developed to an observed daily runoff as a function of daily rainfall, temperature, cloud cover, potential evapotranspiration, and relative humidity. This model is also accurately predicting the basin response to rainfall. The physiological catchment character such as ground slope, vegetal cover factor (land use), and roughness is also considered in this model.

From a review of previous researches, several researchers have investigated and studied the different aspects of this area over the last century, but did not arrive with clear signals. Then, this study has started to research in the district of Bankura.

2 Study area and data

2.1 Site location

The study area is located on Lower Gangetic Plain (Zone no-III, as per Planning Commission of India) in between 22° 38'00" and 23°38'00" N latitude and 86°36'00" and 87°46'00" E longitude. The area of the district Bankura is 688200 hectares. The district consists of different groundwater potential and water table contour (shown in Figure 1). The river system of the district (shown in Figure 2) is divided into different sub-basin (As per Annual Flood Report, April 2017, Irrigation and Waterways Directorate, Government of West Bengal) such as i) Damodar sub-basin: The catchment area of this basin in West Bengal is 4325 km^2. Sali is one of the tributaries of Damodar River located in district Bankura. ii) Darakeswar sub-basin: The catchment area of this basin is 4292 km^2. The tributaries of Darakeswar River are Arkasha, Kansachor, Gangheswari, Berai, Khukra, and Shankari. iii) Shilabati sub-basin: The catchment area of the basin is 4088 km^2. The tributaries of Shilabati River are Jaiponda, Puratan, Champyan, and Ketia. iv) Kangsabati sub-basin: The catchment area of this basin in West Bengal is 6324 km^2. The tributaries of Kangsabati River are Jam, Jhinuk, and Kumari. Damodar River is originated from Palamau hill range in Jharkhand State. Darakeswar, Shilabati, and Kangsabati are originated from upland of Chhota Nagpur Plateau in district Purulia, West Bengal.

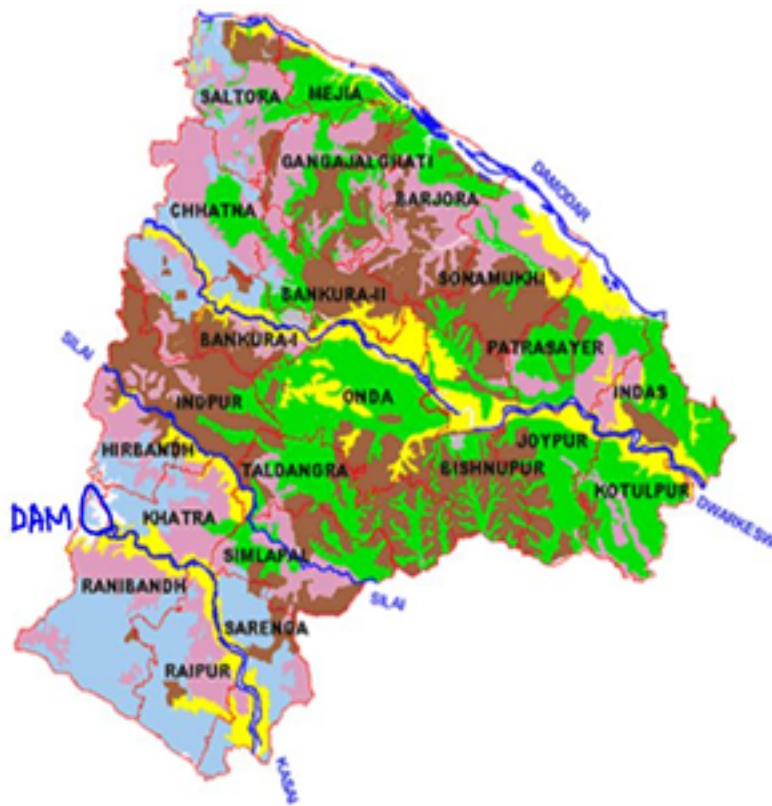**Fig 1.** Ground water potential and water table contour.

Fig 2. The river map of district. Bankura

2.2 Meteorological data

For this research, the data have been collected from India Meteorological Department, Pune, for six meteorological station located in the district Bankura such as i) Bankura, ii) Bankura (Central Water Commission), iii) Joypur, iv) Kangsabati dam, v) Ranibandh, vi) Indus. The researches have been considered for predicting runoff by using ANNs model. The following five input variables are Rainfall, Average Temperature, Cloud Cover, Potential Evapotranspiration and Relative Humidity. In the present investigation, the rainfall intensity pattern is used having three different types of intensity i.e. 30mm/hr, 60mm/hr and 90mm/hr.

2.3 Physiographical data

For physiographical data, viz. Land Use, Slope and Roughness have been collected from the chief soil survey officer, soil and land use survey of India, Pusa, New Delhi and Natural Resources Data Centre, Bankura, W.B. which is shown in Figure 3.

2.3.1 Land Use details
Land Use details of the district are 59.54 % under cultivation, 12.36% cultivable wasteland, 8.48% barren land, and 19.61% cover with forest. The land utilization pattern of the district reveals that Saltora, Mejia, Gangajalghati, Bankura, Bishnupur, and Patrasayer all have more than 57% of land under cultivation. The Joypur block has the highest cultivation (66%) whereas India has 83.01% highest cultivation. All these blocks are located in the central and eastern parts of the district. In the western and south-western part of the district have a less cultivated area. In Chhatna, Ranibandh and Raipur have more than 30% barren land and other 30% cover by forest. The percentage is covered by the forest of the following blocks – Barjora 23%, Ranibandh 32.87%, Taldangra 31.00%, and Bishnupur 35%. Researches have collected the all land use data from Soil and Land Use Survey of India, Pusa, New Delhi.
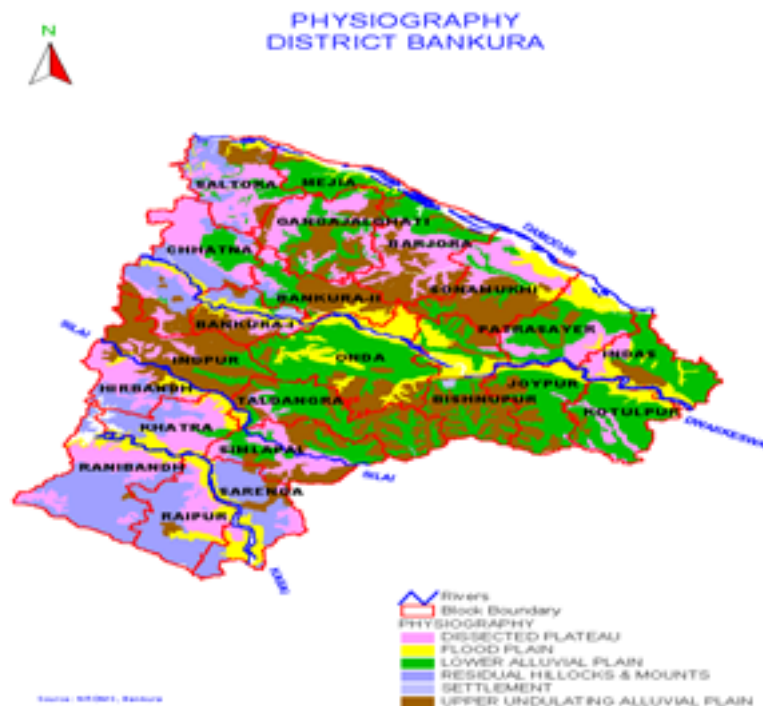
Fig 3. The physiographical map of district Bankura.

2.3.2 Slope details

The land of the district consists of nine slope categories. The slopes of the district according to their area are as per Table 1. In the present investigation, three different overland slopes of 1%, 2%, and 3% were used.

Table 1. Details classification of slope of the district Bankura

Sl no	Slopes classes	Area in hectare	Area in percentage (%)
1.	Level slope	23614.00	3.430
2.	Level – Very Gently slope	371351.00	53.960
3.	Very Gently slope	669.00	0.100
4.	Very Gently – Gently slope	238513.00	34.660
5.	Gently – Moderately slope	6301.00	0.920
6.	Strongly – Moderately steep slope	8088.00	1.180
7.	Moderately steep – Steep slope	1561.00	0.230
8.	Steep – Very steep slope	995.00	0.140
9.	Miscellaneous slope	37108.00	5.390
	Total	**688200.00**	**100.000**

*Sources: Soil and Land Use Survey of India, Pusa, New Delhi.

Table 2. Minimum and maximum Input / output parameters in model Hydrographs.

	Input parameters			Output parameters
Slope (%)		Rainfall intensity (mm/hr)	Rainfall duration (Sec)	Runoff (mm)
Minimum	1	30	0	0
Maximum	3	90	1140	958.90

2.4 Methodology

2.4.1 Estimation of observed runoff

The research has used the equation of Kothyari and Garde (1991) which is more suitable for Lower Gangetic Basin, district Bankura, West Bengal. This equation is recommended by the International Water Management Institute (IWMI) in working Paper-130 which is published in 2008. Simultaneously this is also recommended by the National Institute of Hydrology, Roorkee, India in 2008. The equation is the following-

$$R_m = \frac{F_V^{0.49}(P_m - 0.5T_m)}{26.5} \tag{i}$$

Where, R_m = Annual mean runoff in cm, P_m =the average annual rainfall in cm, T_m = the average annual temperature in ^{0}C, and F_V = vegetal cover factor.

$$F_V = \frac{(a_1F_F + a_2F_G + a_3F_A + a_4F_W)}{A} \tag{ii}$$

Where, a_1, a_2, a_3, a_4 are the weighting factors, F_F is the percentage area of forest, F_G is the percentage area of grass and scrub land, F_A is the percentage area of arable land, and F_W is the percentage area of waste land only.

2.4.2 Artificial Neural Network (ANN) modelling

In Hydrological research, ANN is a recognized tool to construct a structure between multiple input variables and specific output[12–14]. This technique is more suitable for forecasting and runoff analysis[15,16]. In this investigation, ANNs consist of five input layers, a single hidden layer, and an output layer. Hidden layers may be increased in case of a complex situation.

Firstly, ANNs are trained with a series of observed inputs data set viz. Rainfall, Average Temperature, Cloud Cover, Potential Evapotranspiration, and Relative Humidity, denoted as X_1, X_2, X_3, X_4, and X_5 respectively and output data, Runoff denoted as Y_i. In the training process, the coefficients (denoted as W_k and W_{kj}) are obtained. The first process carries out exploration for the optimum nonlinear correlation among input variables and output. In linear regression, the network involves input variables (X_j) linear functions operated by transfer function as shown in equation (iii). Where, the hidden unit receives from each and every input variable. We are writing mathematically by the following equations:

$$U_K = \varphi\left(\sum_{j=i}^m W_{kj}X_j + b_{kj}\right) \tag{iii}$$

$$Y_i = \varphi\left(\sum_{k=i}^l W_k u_k + b_k\right) \tag{iv}$$

Where, X_1, X_2, …… X_m are input variables; φ denotes the hyperbolic transfer function; W_{k1}, W_{k2},… W_{km} and W_1, W_2,… W_k are the coefficients (weights) of the network; u_1, u_2, ….u_k denotes hidden units; b_{kj} and b_k are the constants to linear regression, and Y_i is the output signal. Hidden units linear function along constant obtains final output of the network as shown in equation (iv). At first, normalization of the data series are performed as per equation (v) within ±1.

$$P_{no\ r} = 2\left(\frac{P_0 - P_{\min}}{P_{\max} - P_{\min}}\right) - 1 \tag{v}$$

In equation (iii), P_o, P_{nor}, P_{min}, and P_{max} denotes observed data, normalized data, minimum observed data, and maximum observed data respectively. During ANNs modeling, complete data series were randomly distributed into training, testing, and validation in 70% / 15% /15% format. Initially, the network was trained with training data series, and after that testing, data series was utilized to calibrate the behavior of trained models. After calibration and testing finally, the data series was used to validate and to complete the ANNs model.

During the training the network, Gradient Decent (GD) algorithm was applied to reduce the network error by a function minimization routine and to improve the network output.

$$Overall\ error\ E_D = \Sigma(t_j - y_i)^2 \tag{vi}$$

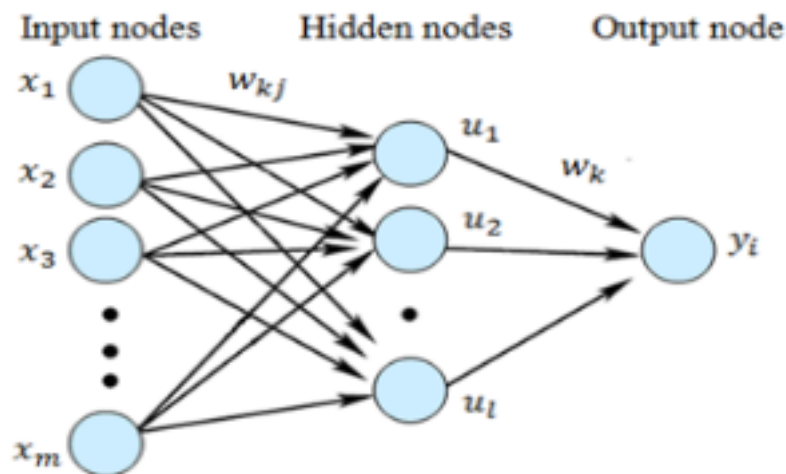Where t_j = desired error and y_i = calculated output.

Fig 4. Schematic illustrations of artificial neural networks with single hidden layer

*Sources: https://www.researchgate.net / Publication / 282903935

3 Testing of model performance

3.1 Using Nash – Sutcliffe criteria

For estimating the model performance, the Nash-Sutcliffe Efficiency (NSE) is an important measure that is expressed by the equation (vii).

Nash-Sutcliffe Efficiency (NSE) in percentage (%),

$$NSE = \left(1 - \frac{\sum_{i=1}^{n}(Y_o - Y_K)}{\sum_{i=1}^{n}(Y_o - Y_m)}\right) \qquad \text{(vii)}$$

Where, Y_O = Observed flow at time t, Y_K = Predicted flow (Kinematic flow) at time t, and Y_m = Mean observed flow.

Nash-Sutcliffe Efficiency (NSE) measures the variability of the model (Legates and McCabe, 1999). NSE ranges from -∞ at a worst-case to +1 for a perfect correlation. According to Shamseldin (1997), the value of NSE 0.9 and above is very satisfactory, 0.8 to 0.9 represents a fairly good model and 0.8 is an unsatisfactory result[6].

3.2 Root mean squared error (RMSE)

$$RMSE = \sqrt{\frac{\sum_{k=1}^{K}(Q_k - \bar{Q}_k)^2}{K}} \qquad \text{(viii)}$$

Where, Q_k = Observed flow, $\bar{Q}_k$ = Predicted flow, and K = Total number of year considered (Validation data).

The value of RMSE measures the accuracy of the model, and as minimum as possible. The value of RMSE and MAE more or less similar (near about same) and represents the good satisfactory model (C. W. Dawson and R. L. Wilby, 2015).

3.3 Mean absolute error (MAE)

$$MAE = \frac{\sum_{i=1}^{K}|(Q_k - \bar{Q}_k)|}{K} \qquad \text{(ix)}$$

Where, Q_k = Observed flow, $\bar{Q}_k$ = Predicted flow, and K = Total number of year considered (Validation data).

The value of MAE measures the accuracy of the model and as minimum as possible.

3.4 Error in runoff computation

The error in Runoff for the present investigation was estimated by,

$$\text{Error in \% } e = \left(1 - \frac{Y_K}{Y_O}\right) * 100 \tag{x}$$

Where, Y_o = Observed Runoff, and Y_K = Predicted Runoff.

The value of Error in Runoff computation measures the accuracy of the model and as lower as possible.

3.5 Correlation coefficient for linear regression

$$\text{Correlation Coefficient value "}r\text{"} = \frac{\sum_{i=1}^{n} \left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)}{\sqrt{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 \sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2}} \tag{xi}$$

Where, X_i = Observed flow, $\overline{X}$ = Mean of observed flow, Y_i = Predicted flow, $\overline{Y}$ = Mean of Predicted flow. The values of "r" indicate the association with the two values, ranges from -1 (perfectly negative correlation), through 0 (no correlation) to +1 (strongly positive correlation).

4 Result and Discussion

The ANNs model (Rainfall-Runoff model) is a machine learning process for flow (discharge) evaluation using Microsoft Excel. The researcher has collected the real-time data series (1901 – 2016) from India Meteorological Department (IMD), Pune, for the district Bankura. The data were taken from IMD station located six places and equally distributed in the entire district viz. i) Bankura, ii) Bankura (Central Water Commission), iii) Joypur, iv) Kangsabati dam, v) Ranibandh, vi) Indus. The data were taken at various slopes and various intensities. The studies have been considered for predicting runoff for five major climatic variables. The five input variables are Rainfall, Average Temperature, Cloud Cover, Potential Evapotranspiration, and Relative Humidity. The database used for ANNs model development is mentioned in Table 2. For ANNs model, the observed data were considered as 70% for training, 15% for testing, and 15% for validation.

The behavior of the ANNs during training, testing, and validation are shown in Figure 9. The correlation coefficient value "r" above 97% indicates the strong positively correlated with each other (observed and predicted flow). The training process of the model network was stopped when the error of the data series was minimal. The objective of the training process is to reduce error based on model performance (NSE, RMSE, and MAE). The validation set is used to obtain the best result of the model network. The values of training "r" = 0.9961, testing "r" =0.9783, and validation "r" =0.9989, represents the good agreement with correlation coefficient values as shown in Figure 9. The data pairs are very closer to the diagonal line represent the excellent prediction. The hydrograph using ANNs model as shown in Figure 10 reveals that runoff data series is more efficient for the rising, equilibrium discharge, and recession limb. For events, model "C" and "D", the Nash-Sutcliffe Efficiency (NSE) greater than 95% (shown in Table 3) indicates the excellent results.

Table 3. Summarized the result and testing of model performance

Model	Observed Runoff in mm	Predicted Runoff in mm	Error in Runoff in %	Correlation Coefficient "r"	Testing of model performance			Remarks
					NSE	RMSE	MAE	
A	764.85	759.74	6.590	0.9736	81.38	43.495	38.788	Model "C" is the Best fitted model and very satisfactory.
B	789.80	792.38	1.092	0.9746	87.925	46.521	38.574	
C	787.46	769.31	0.062	0.9782	97.492	44.008	40.477	
D	783.53	773.32	0.290	0.9897	95.245	62.759	45.792	

The predicted result is useful in the field of water resources management and planning for decision-making[4]. Apart from that, the modeling can also be useful for rural as well as urban planners to take necessary measures.
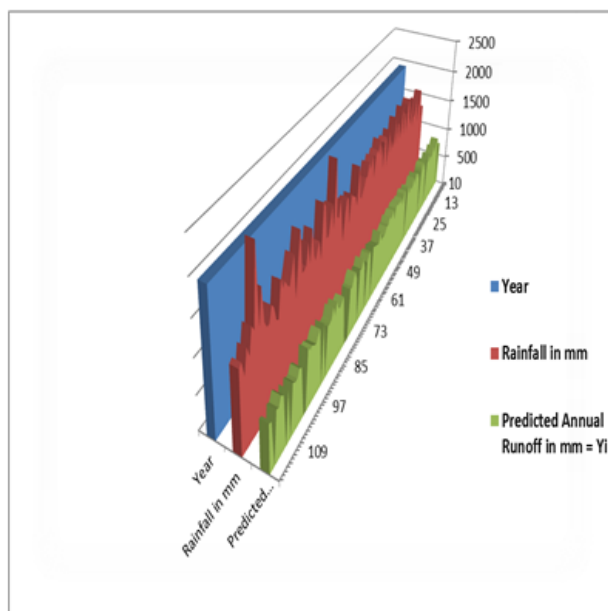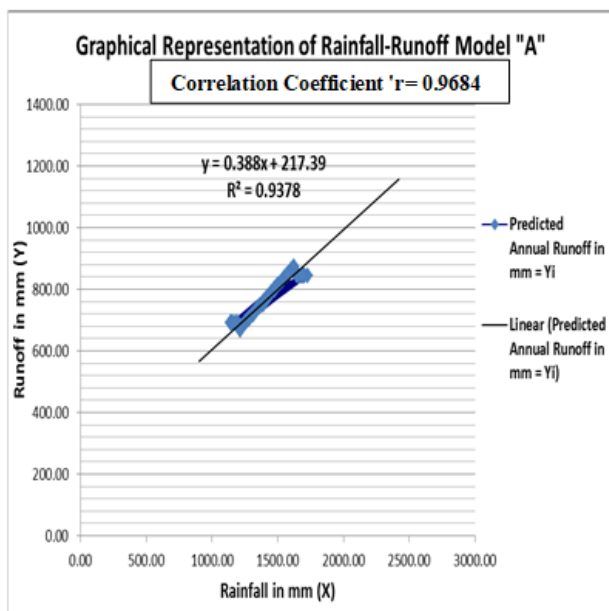
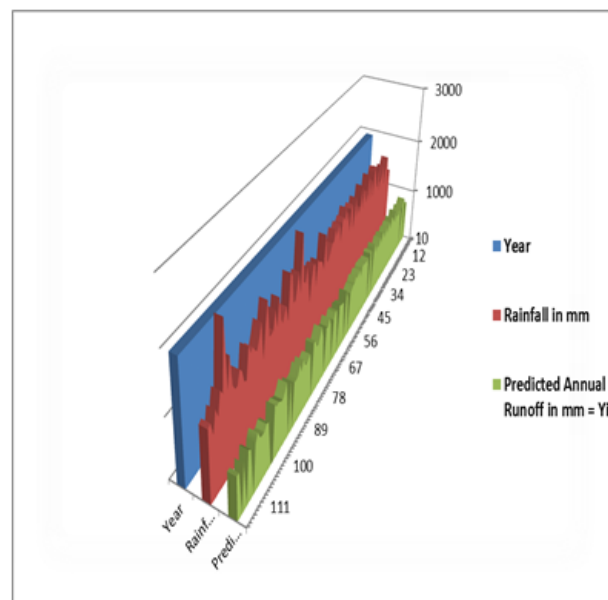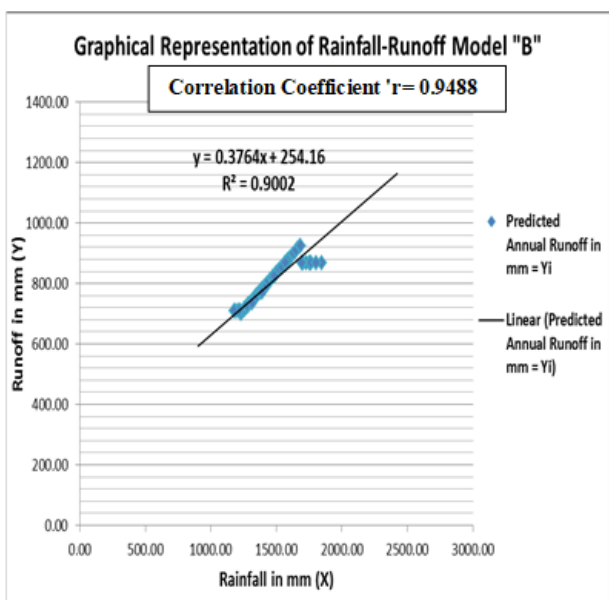**Fig 5.** The graphical representation of rainfall-runoff model "A"

Fig 6. The graphical representation of rainfall-runoff model "B"
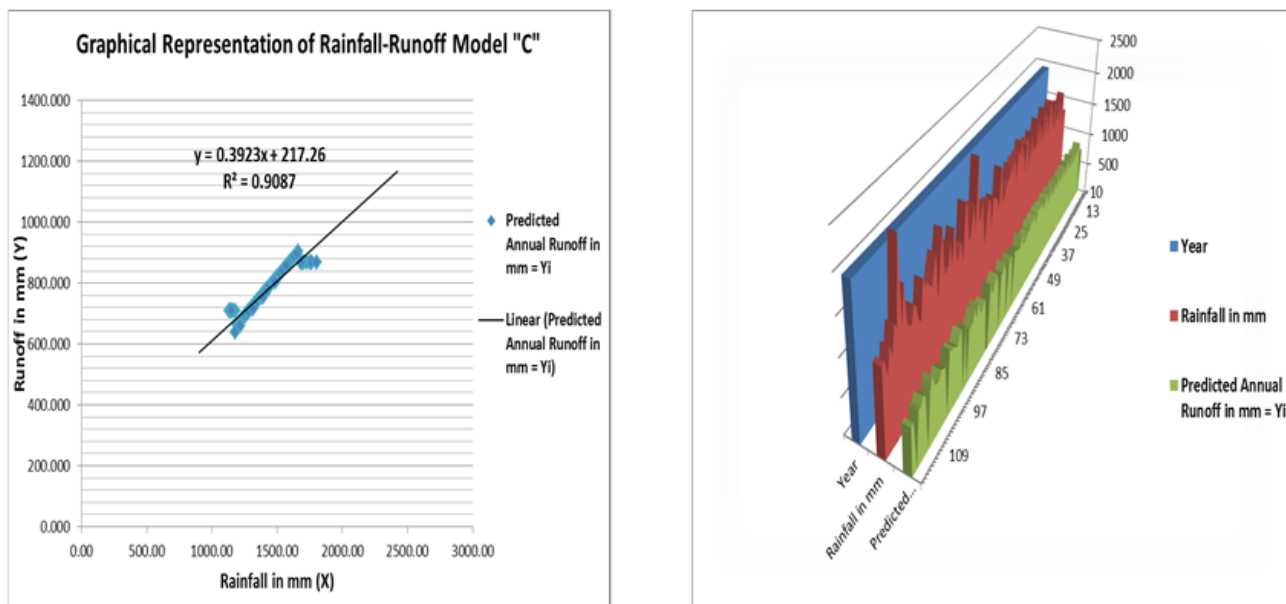
Fig 7. The graphical representation of Rainfall-Runoff Model "C"
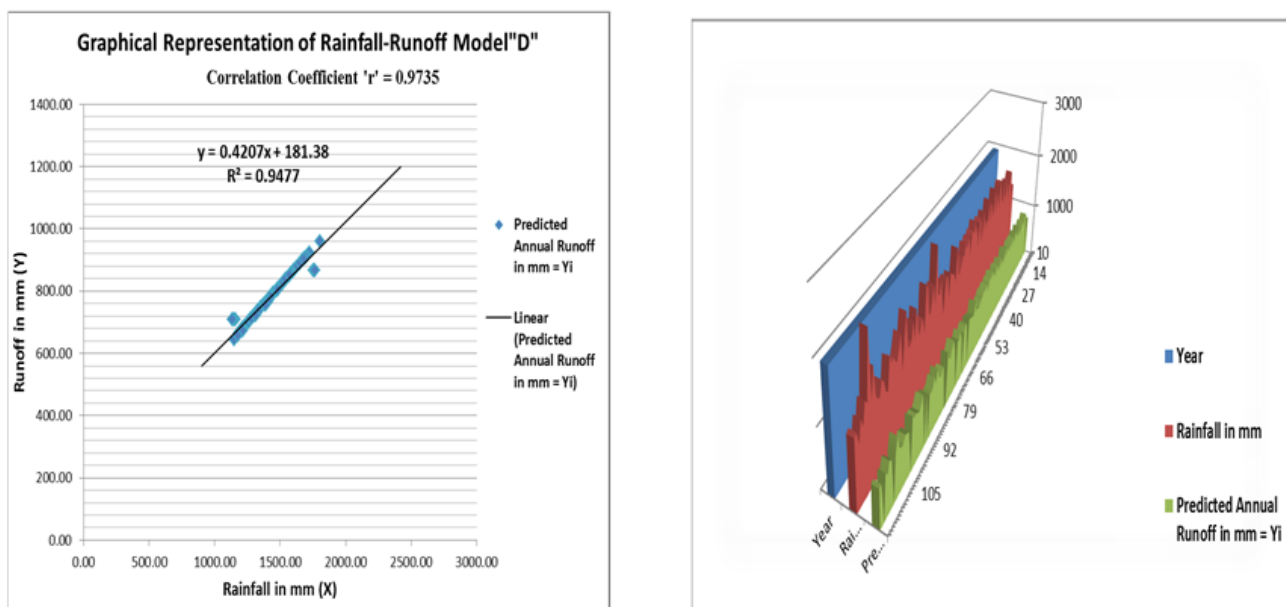

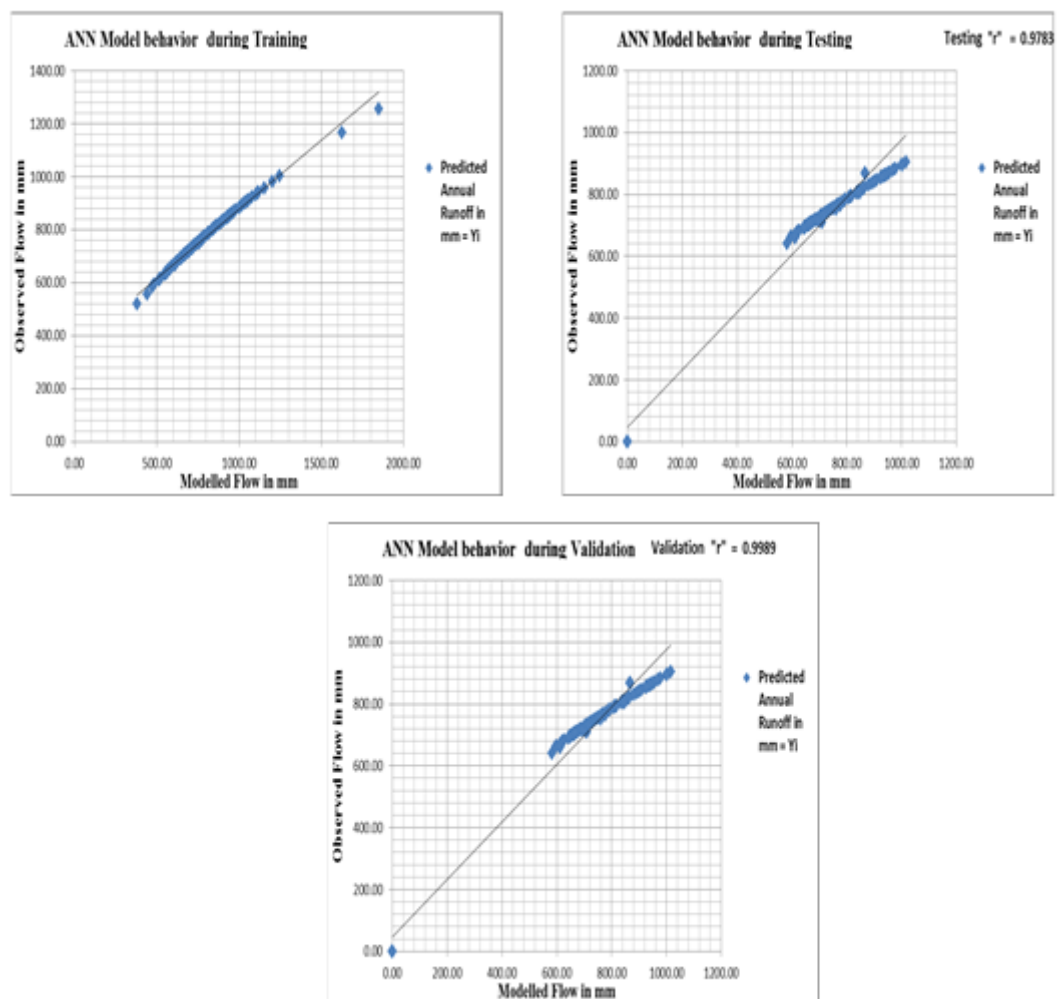**Fig 8.** The graphical representation of Rainfall-Runoff Model "D"

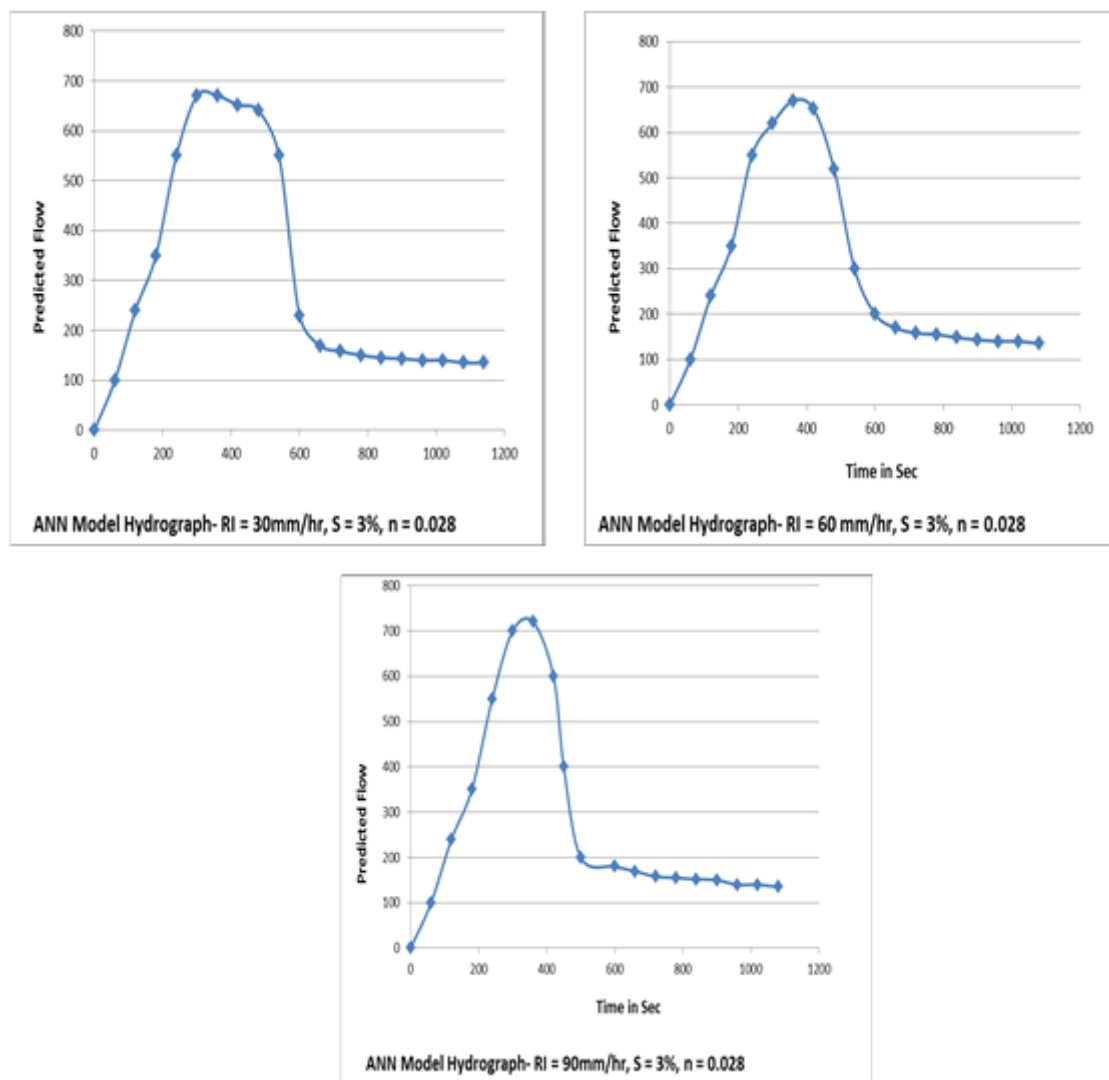**Fig 9.** The ANNs Model behaviour during training, testing, and validation.

Fig 10. The ANNs model hydrograph of different rainfall intensities (RI = 30 mm/hr, RI = 60mm/hr, and RI=90mm/hr)
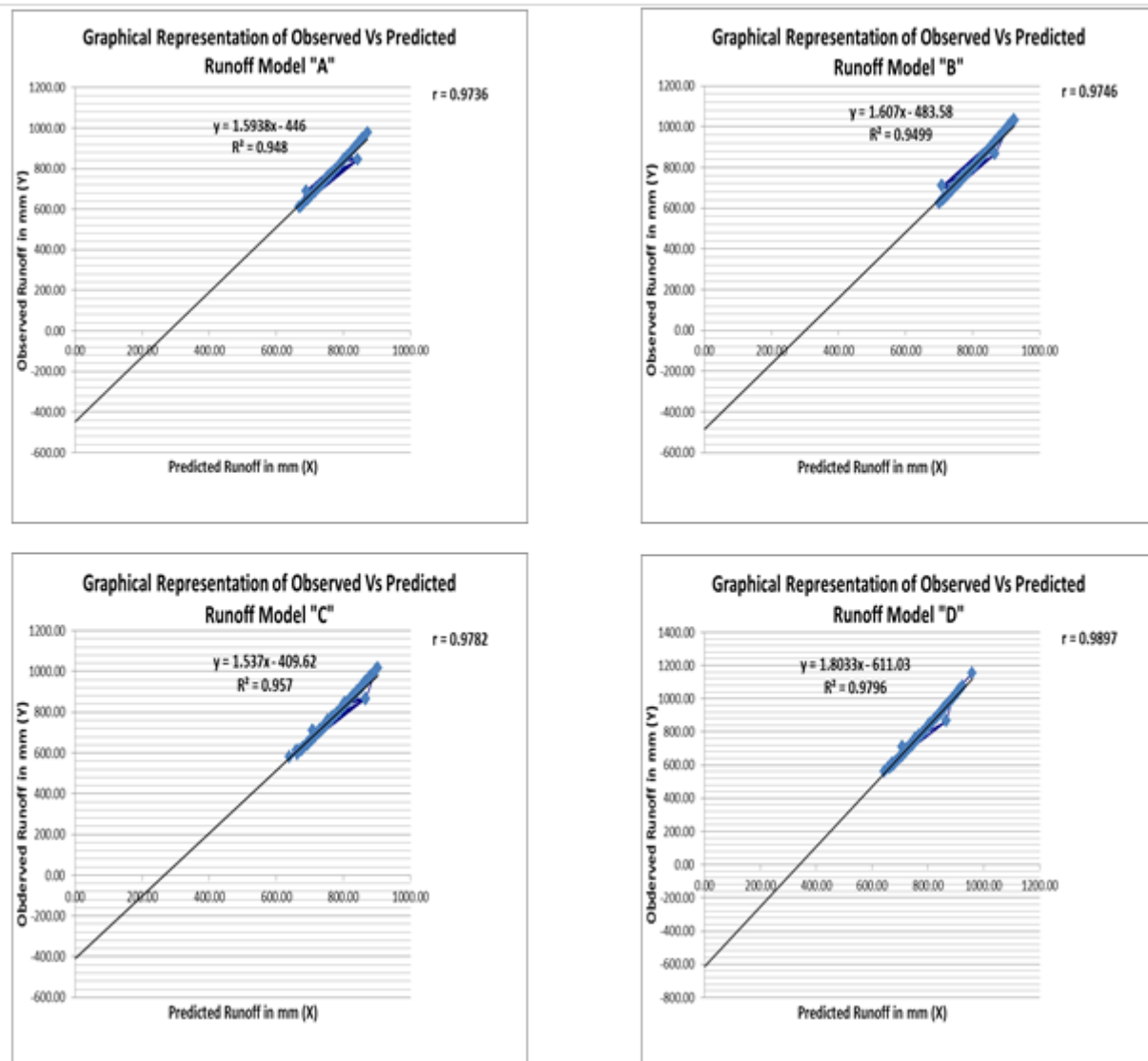
Fig 11. The graphical representation of observed vs predicted runoff for ANNs model (A, B, C and D)

5 Conclusions

1. The predicted result should be more useful in the field of water resources management and planning in the future.
2. First time, the study has been used Kothyari and Garde equation (recommended by National Institute of hydrology, Roorkee, India) in which the most important factor i.e. the Vegetal Cover Factor (Fv) was considered over this area.
3. For all events, model "A", "B", "C", and "D", correlation coefficient "r" and the Nash-Sutcliffe Efficiency (NSE) (for model "C") greater than 97% (shown in Table 3) indicates an excellent result.

Acknowledgement

References

1) Kundu S, Khare D, Mondal A, Mishra PK. Future rainfall analytic thinking (1871-2011) for whole India. *Temperature Change and Bio-diversity*. 2018;20(2):45–60. Available from: https://doi.org/10.18231/2454-9150.2018.1186.

2) Kundu S, Mishra PK, Mondal A, Khare D. Study and trend analysis of rainfall and temperature change of M. *Environmental Earth Sciences*. 1901;73(12):8197–8216. Available from: https://doi.org/12.1821/2454-9150.2015.1008.

3) Meena PK, Khare D, Kundu S, Mishra PK. Future analytic thinking of mega cities in North India exploitation precipitation knowledge. *Indian Journal of science and technology*. 2015;8(3):247–253. Available from: https://doi.org/10.17485/ijst/2015/v8i3/59580.

4) Dawson CW, Wilby RL. Hydrological modeling using Artificial Neural Networks. *Progress in Physical Geography*. 2015;25(1):80–108. Available from: https://doi.org/10.1177/030913330102500104.

5) Jamaludin S, Abdul AJ. Fitting the statistical distributions to the daily precipitation amount in Peninsular Malaysia. *Journal Technology*. 2017;46(3):33–48. Available from: https://doi.org/10.3923/jas.2017.1880.1886.

6) Bhutiyani MR, Kale VS, Power NJ. Semi-permanent trends in most, minimum and mean annual surface air temperature across the North Western Himalayas throughout the 20th century. *Climatic Modification*. 2017;85(8):159–177. Available from: http://swat.tamu.edu/media/114860/usermanual_swatcup.

7) Duhan D, Pandey A. Mathematical regression analysis of future rainfall and Temperature trend M.P, India. *Atmospherically analysis*. 1901;122:136–149. Available from: https://doi.org/10.24941/ijcr.33157.11.2016.

8) Daniela S, Georgiana PE, Catalina N. Weather forecast by Statistical Methods. *Weather Prediction*. 2019;1:269–274. Available from: http://www.weather.gov.hk/wxinfo/pastwx/extract.

9) Chakravarti A, Joshi N, Panjiar H. Rainfall Runoff Analysis using Artificial Neural Network. *Indian Journal of Science and Technology*. 2015;8(14):229–235. Available from: https://dx.doi.org/10.17485/ijst/2015/v8i14/54370.

10) Kumar PS, Praveen TV, Prasad MA. Artificial Neural Network Model for Rainfall-Runoff -A Case Study. *International Journal of Hybrid Information Technology*. 2016;9(3):263–272. Available from: https://dx.doi.org/10.14257/ijhit.2016.9.3.24.

11) Dhar A, Vekariya LR, Bhadja P. n-Alkane isomerization by catalysis—a method of industrial importance: An overview. *Cogent Chemistry*. 2018;4:424–427. Available from: https://dx.doi.org/10.1080/23312009.2018.1514686.

12) Dhawal H, Mishra N. A Survey on Precipitation Prediction Techniques. *International Journal of Computer Application*. 2016;6(2):1797–2250. Available from: http://www.cs.waikato.ac.nz/ml/weka.

13) Mathur P, Mathur S. A Simple Weather Forecasting Model Using Regression analysis. *Current Science*. 2014;6(2):145–263.

14) Sethi N. Exploiting Data Mining Technique for rainfall prediction. *International Journal of Computer Science and Technologies*. 2014;5(3):3982–3984. Available from: https://www.researchgate.net/publication/275340859.

15) Ireland G, Volpi M, Petropoulos GP. Determining the capacity of machine learning classifiers in extracting flooded areas). *Remote Sensing*. 2015;7(3):3372–3399.

16) Jain SK, Kumar V. Study and analysis of rainfall and temperature. *Current Science*. 2016;102(1):37–39. Available from: https://www.researchgate.net/publication/2753408592.