Advances in Research

Volume 25, Issue 6, Page 41-53, 2024; Article no.AIR.124077
ISSN: 2348-0394, NLM ID: 101666096

Sediment Simulation Using Multi Layer Perceptron (MLP), Co-Active Neuro-Fuzzy Inference System (CANFIS) and Multiple Linear Regression Tecniques (MLR) for Hurdag Watershed

Vikram Singh [a++], Rananjay Singh [a#] and Sonali Kumari [a†*]

[a] SHUATS, Prayagraj, UP, India.

Authors' contributions

This work was carried out in collaboration among all authors. All authors read and approved the final manuscript.

Article Information

DOI: https://doi.org/10.9734/air/2024/v25i61177

Open Peer Review History:
This journal follows the Advanced Open Peer Review policy. Identity of the Reviewers, Editor(s) and additional Reviewers, peer review comments, different versions of the manuscript, comments of the editors, etc are available here: https://www.sdiarticle5.com/review-history/124077

Original Research Article

Received: 01/08/2024
Accepted: 03/10/2024
Published: 02/11/2024

ABSTRACT

Sediment modeling plays a crucial role in sustainable water resources planning, development, and management. Techniques like the Multilayer Perceptron (MLP), Co-active Neuro-Fuzzy Inference System (CANFIS), and Multiple Linear Regression (MLR) have proven effective for sediment modeling and forecasting. This study aimed to develop and assess the applicability of MLP, CANFIS, and MLR models by training and testing them during the monsoon season (June to September) for the Hurdag watershed in the Damodar-Barakar basin, located in Hazaribagh district,

++ Associate Professor;
 # Assistant Professor;
† Ph.D Scholar;
**Corresponding author: E-mail: aashusonali111@gmail.com;*

***Cite as:** Singh, Vikram, Rananjay Singh, and Sonali Kumari. 2024. "Sediment Simulation Using Multi Layer Perceptron (MLP), Co-Active Neuro-Fuzzy Inference System (CANFIS) and Multiple Linear Regression Tecniques (MLR) for Hurdag Watershed". Advances in Research 25 (6):41-53. https://doi.org/10.9734/air/2024/v25i61177.*

Jharkhand, India. Daily rainfall, runoff (or streamflow), and sediment concentration data from 1997 to 2006 were used, with the data split into two sets: training set (1997–2004) and a testing set (2005–2006). The analysis was conducted using NeuroSolution 5.0 software and Microsoft Excel for performance evaluation indices. The best input combinations for sediment yield simulation were identified, and 10 optimal models were selected from 31 different input combinations. These input combinations were applied to the network for training using the back propagation algorithm for MLP and Gaussian and generalized bell membership functions for CANFIS models. Multiple networks were trained individually, with the most accurate predictions during testing being chosen as the best models. The models' performance was evaluated using statistical indices such as root mean squared error (RMSE), coefficient of efficiency (CE), and correlation coefficient (r). The results showed that MLP and CANFIS models performed best in predicting sediment concentration for the Hurdag watershed, whereas MLR models showed poor performance for the given dataset. Specifically, sediment concentration for the current day could be modeled using current day rainfall and runoff data (MLP-7), while runoff could be simulated using the previous day's rainfall data (MLP-2).

Keywords: Soft computing; CANFIS; MLP; MLR; sedimentation prediction.

1. INTRODUCTION

Water is a vital and invaluable resource provided by nature, essential for life on Earth in forms such as rain, snow, rivers, and oceans. Rainfall, a key component of the hydrological cycle, is challenging to understand and model due to the complexity of atmospheric processes. Life on Earth is unimaginable without water. India receives about 4000 km³ of annual precipitation, with 75% occurring as monsoon rainfall, primarily from the south-west monsoon between June and September (Ministry of Water Resources, 1999). The north-east monsoon also contributes, especially in Tamil Nadu from October to November. In urban areas, rainfall significantly impacts sewer systems, waterlogging, traffic, and other activities. Advances in computer technology and GIS have enhanced the spatial interpolation of precipitation data, improving our ability to analyze and manage this critical resource.

Artificial neural networks (ANNs) have recently emerged as the most effective method for rainfall forecasting (Haykin, 1999, Rajurkar, et al., 2004). ANNs are non-linear mapping structures that mimic the human brain's learning process, making them powerful tools for modeling complex, noisy, and imprecise data, even when the underlying relationships are unknown. During the training process, ANNs identify and learn patterns between input data and target values, enabling them to predict outcomes for new data sets (Zhang, 2000). This approach yields reliable results without needing detailed information on catchment characteristics. ANNs combine linear and non-linear concepts in model building and can function in both dynamic and memory-less systems. In hydrology, ANNs are used for modeling daily rainfall-runoff, runoff-sediment yield, and assessing the ecological and hydrological impacts of climate change on streamflow, sediment transport, and groundwater quality (Maier and Dandy, 2000).

The Multi-layer Perceptron (MLP) is the most widely used ANN architecture today. An MLP consists of three layers: input, hidden, and output. Each neuron in the network computes an output by combining weighted inputs and applying a nonlinear activation function. In hydrological modeling, MLP has been extensively applied due to its ability to model complex relationships. The network processes data in one direction, from the input layer through the hidden layer to the output layer. Training involves adjusting the weights connecting neurons through error backpropagation, where the network learns from a series of training examples to model the relationship between predictors and predictands.

The conventional CANFIS model is an extension of the original ANFIS model, allowing for multiple input-output pairs. The core structure of CANFIS is similar to ANFIS, where a fuzzy neuron with a membership function (MF) is used to build the model. Several types of MFs can be employed, such as triangular, trapezoidal, sigmoidal, Gaussian, z-shape, pi, and general bell functions. The CANFIS model also normalizes output variables within the 0-1 range. Its architecture combines the output of MFs with the neural network target, aligning closely with the standard ANFIS process.

An approach to modelling the linear correlations between one dependent variable (Y) and two or more independent variables (X1, X2,.........Xn) is called multiple linear regressions (MLR).

Given the foregoing, this study was conducted to train, test, and validate the Sediment models for the watersheds of the Damodar-Barakar basin in the Hazaribagh district of Jharkhand state, India, using multiple linear regressions (MLR) and soft computing techniques (MLP,CANFIS based ANN).

2. MATERIALS AND METHODS

2.1 Study Area

As seen in Fig. 1, the study's chosen region is the Hurdag watershed of the Damodar-Barakar basin in the Hazaribagh district of Jharkhand state, India, which has an area of 23.04 km2 [6]. The Hurdag watershed is located between latitudes 230 47' 35'' and 230 52' 8'' N and longitudes 850 30' 30'' and 850 39' 45'' E. The Central Water Commission (CWC), the India Meteorological Department (IMD), and the Soil and Water Conservation Division of the Damodar Valley Corporation, located in Hazaribagh, Jharkhand, provided the hydrological data (rainfall, discharge, and sediment yield).

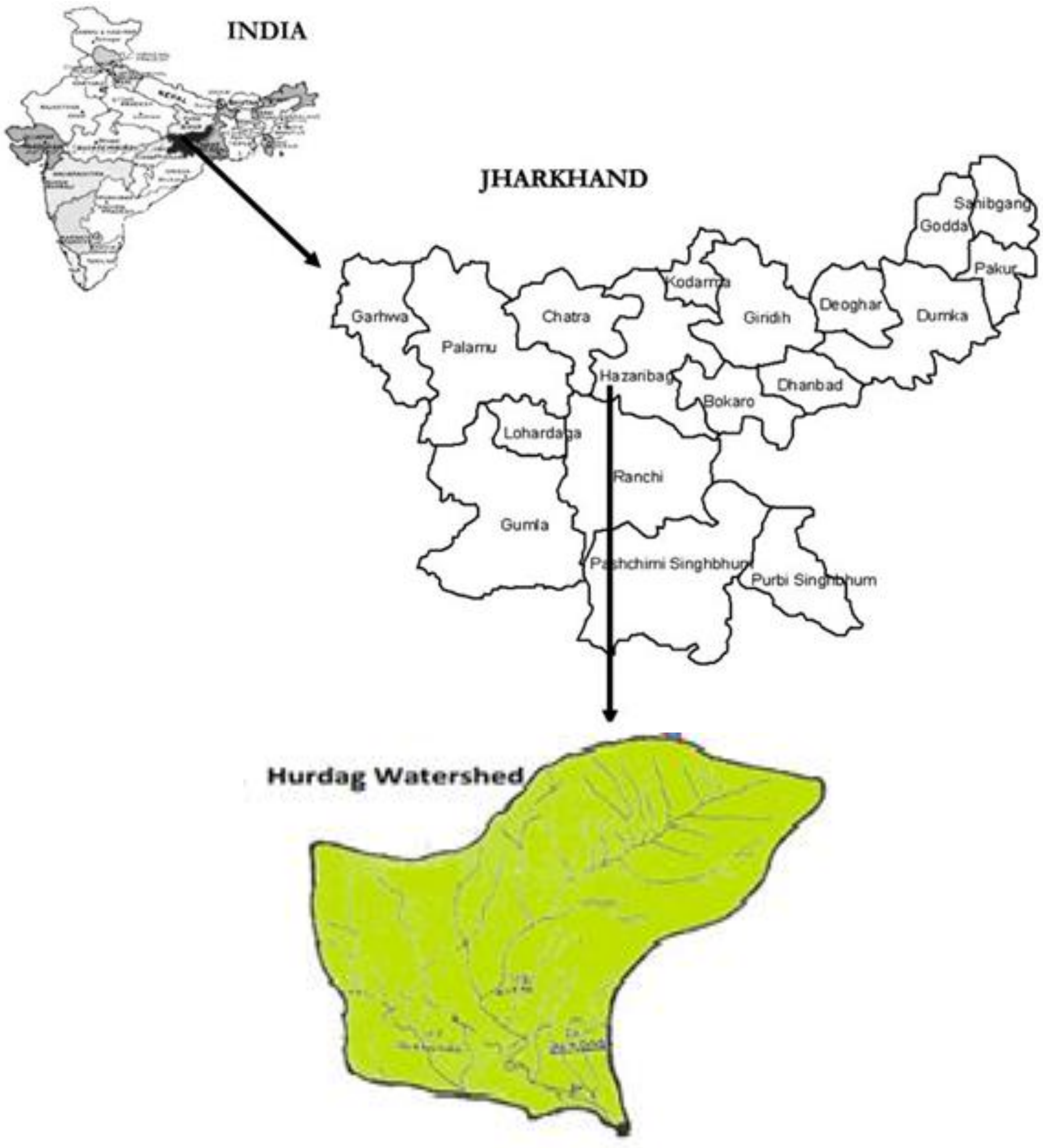

Fig. 1. Location map of the study area

Land use and land cover data are crucial for effective watershed planning and management. The area features diverse land cover types, including forests, vegetation-covered and bare lowlands, non-agricultural lands, settlements, uplands, wastelands, and water bodies. Due to the uneven terrain, cultivation occurs on terraced slopes and lowlands, with paddy grown in some areas and Rabi crops benefiting from high water availability. The construction of small check dams has converted much of the mono-cropped land into double-cropped areas. The watershed also has gully-eroded lands, and the primary tree species include Sal, Sisam, Mahua, Mango, and Eucalyptus. Settlements are dispersed throughout the area. The Soil Conservation Department, DVC, Hazaribagh, and Jharkhand created the watershed soil map. In the watershed, sandy clay loam makes up 56.9% of the soil (Singh and Kumar, 2016).

The study's watersheds are located in Jharkhand's Hazaribag district, within the Tilaiya catchment of the Damodar River valley, about 25 km from Hazaribag and 35 km from Tilaiya reservoir. The main river, Kothuwatari, joins the Mohaghat River and flows into the Barakar River. The area features diverse landscapes, from flatlands to steep hills, with elevations between 385 and 655 meters. The watershed is characterized by undulating uplands, dissected valleys, and significant erosion, including sheet, rill, and gully erosion.

The region receives an average annual rainfall of about 1240 mm, primarily between June and September, with occasional showers in December-January and heavy rains in May. Temperatures range from a maximum of 43°C in April-May to a minimum of 2.4°C in January. Winters are cold, while summers are hot and humid, with annual humidity between 66% and 77%. Daily rainfall, runoff, and sediment data were collected during the monsoon season (June-September) from 1997 to 2006 for the Hurdag watershed (Singh and Kumar, 2016).

2.2 ANN Based MLP Model

The concept of Artificial Neural Networks (ANNs) originated in 1943 when Warren McCulloch and Walter Pitts proposed a model based on the human brain, a natural neural network with billions of interconnected neurons (McCulloch and Pitts, 1943). Neurons receive signals through dendrites, process them, and transmit electric signals along axons (Zurada, 1992). Inspired by this, the ANN model was developed, where artificial neurons, or perceptrons, mimic biological neurons. The model learns from input data to reproduce outcomes. In the ANN model, there are three key components:

I. **Synapses:** The strength of the connection between an input and a neurone is represented by the weights that represent the synapses of the neurone.
II. **Adder:** This activity, also known as a linear combination, is what really happens inside the neurone cell. It consists of adding up all of the inputs that have been adjusted for their individual weights.
III. **Activation function:** This mechanism regulates the magnitude of a neuron's output.

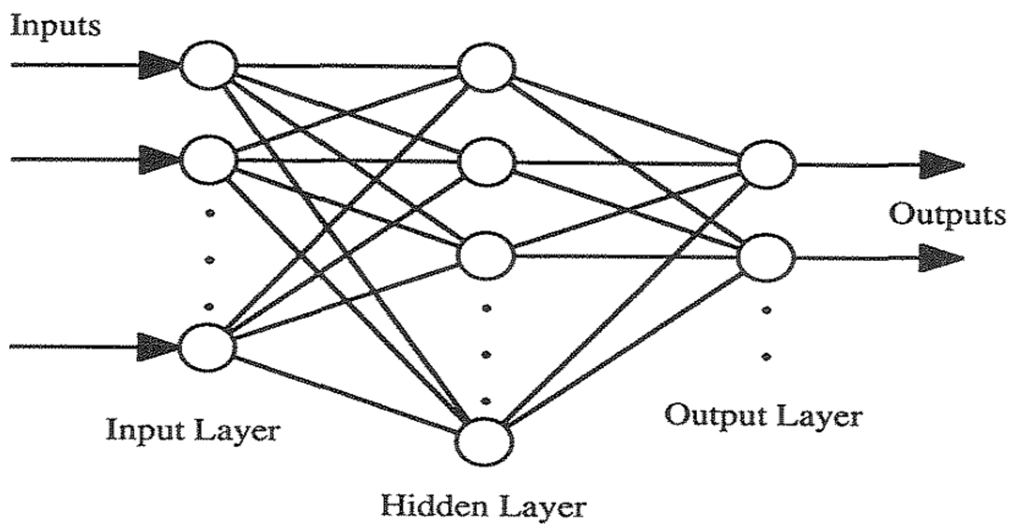

Fig. 2. A basic overview of MLP

Table 1. Input-output combinations for MLP model for sediment concentration simulation at Hurdag watershed

Model No.	Input-Output Variables	Model No.	Input-Output Variables
MLP,CANFIS-1	$S_t = f(R_t)$	MLP,CANFIS-17	$S_t = f(R_t, S_{t-1})$
MLP,CANFIS-2	$S_t = f(R_{t-1})$	MLP,CANFIS-18	$S_t = f(R_{t-1}, S_{t-1})$
MLP,CANFIS-3	$S_t = f(R_t, R_{t-1})$	MLP,CANFIS-19	$S_t = f(R_t, R_{t-1}, S_{t-1})$
MLP,CANFIS-4	$S_t = f(Q_t)$	MLP,CANFIS-20	$S_t = f(Q_t, S_{t-1})$
MLP,CANFIS-5	$S_t = f(Q_{t-1})$	MLP,CANFIS-21	$S_t = f(Q_{t-1}, S_{t-1})$
MLP,CANFIS-6	$S_t = f(Q_t, Q_{t-1})$	MLP,CANFIS-22	$S_t = f(Q_t, Q_{t-1}, S_{t-1})$
MLP,CANFIS-7	$S_t = f(R_t, Q_t)$	MLP,CANFIS-23	$S_t = f(R_t, Q_t, S_{t-1})$
MLP,CANFIS-8	$S_t = f(R_t, Q_{t-1})$	MLP,CANFIS-24	$S_t = f(R_t, Q_{t-1}, S_{t-1})$
MLP,CANFIS-9	$S_t = f(R_t, Q_t, Q_{t-1})$	MLP,CANFIS-25	$S_t = f(R_t, Q_t, Q_{t-1}, S_{t-1})$
MLP,CANFIS-10	$S_t = f(R_{t-1}, Q_t)$	MLP,CANFIS-26	$S_t = f(R_{t-1}, Q_t, S_{t-1})$
MLP,CANFIS-11	$S_t = f(R_{t-1}, Q_{t-1})$	MLP,CANFIS-27	$S_t = f(R_{t-1}, Q_{t-1}, S_{t-1})$
MLP,CANFIS-12	$S_t = f(R_{t-1}, Q_t, Q_{t-1})$	MLP,CANFIS-28	$S_t = f(R_{t-1}, Q_t, Q_{t-1}, S_{t-1})$
MLP,CANFIS-13	$S_t = f(R_t, R_{t-1}, Q_t)$	MLP,CANFIS-29	$S_t = f(R_t, R_{t-1}, Q_t, S_{t-1})$
MLP,CANFIS-14	$S_t = f(R_t, R_{t-1}, Q_{t-1})$	MLP,CANFIS-30	$S_t = f(R_t, R_{t-1}, Q_{t-1}, S_{t-1})$
MLP,CANFIS-15	$S_t = f(R_t, R_{t-1}, Q_t, Q_{t-1})$	MLP,CANFIS-31	$S_t = f(R_t, R_{t-1}, Q_t, Q_{t-1}, S_{t-1})$
MLP,CANFIS-16	$S_t = f(S_{t-1})$		

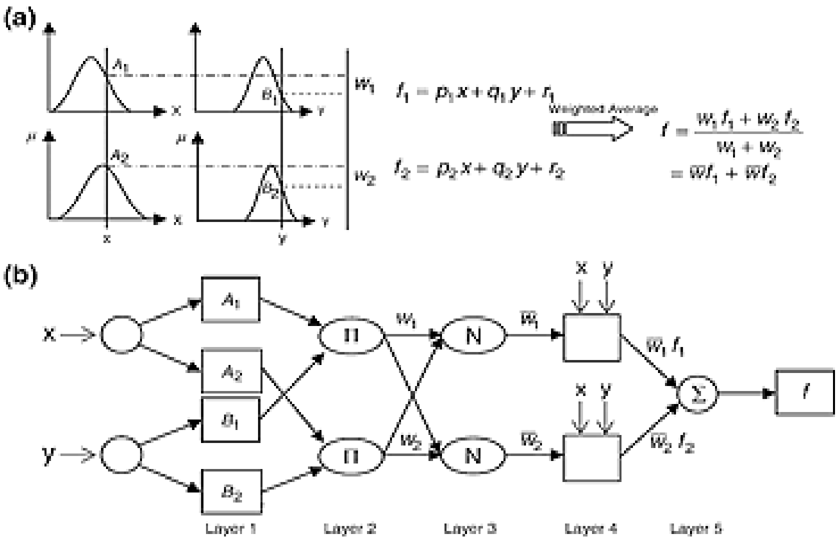

Fig. 3. A basic overview of CANFIS structure (Saemi and Ahmadi 2008)

2.3 Co-Active Neuro-Fuzzy Inference System

The co-active neuro-fuzzy inference system (CANFIS), a generalized form of ANFIS, integrates neural networks with fuzzy inference systems to approximate nonlinear functions. Its strength lies in pattern-dependent weights between the fuzzy association and consequent layers (Jang, et al., 1997). CANFIS uses fuzzy neurons that apply membership functions (e.g., triangular, Gaussian, sigmoidal) to inputs, with a normalization axon adjusting outputs within a set range. The system also includes a modular network that applies functional rules to inputs, with a combiner axon linking membership function outputs to modular network outputs.

2.4 Multiple Linear Regressions

The association between a dependent variable and two or more explanatory factors using a linear function is known as the multiple linear regression (MLR) model. A linear equation is chosen to represent the connection between the n independent variables (X1, X2,.... Xn) and the dependent variable Y. This regression equation may be expressed as follows (Malik and Kumar, 2015):

$$Y = a_0 + b_1X_1 + b_2X_2 + \cdots + b_nX_n \qquad (2.1)$$

where;
$a_0, b_1, b_2 \ldots b_n$, are regression coefficients.

Table 2. Input-output combinations for MLR models for sediment concentration simulation at Hurdag watersheds

Model No.	Input-Output Variables*
MLR-1	$S_t = a_1 + b_1R_t$
MLR-2	$S_t = a_2 + b_2R_{t-1}$
MLR-3	$S_t = a_3 + b_3R_t + b'_3R_{t-1}$
MLR-4	$S_t = a_4 + c_1Q_t$
MLR-5	$S_t = a_5 + c_2Q_{t-1}$
MLR-6	$S_t = a_6 + c_3Q_t + c'_3Q_{t-1}$
MLR-7	$S_t = a_7 + b_4R_t + c_4Q_t$
MLR-8	$S_t = a_8 + b_5R_t + c_5Q_{t-1}$
MLR-9	$S_t = a_9 + b_6R_t + c_6Q_t + c'_6Q_{t-1}$
MLR-10	$S_t = a_{10} + b_7R_{t-1} + c_7Q_t$
MLR-11	$S_t = a_{11} + b_8R_{t-1} + c_8Q_{t-1}$
MLR-12	$S_t = a_{12} +b_9R_{t-1} + c_9Q_t + c'_9Q_{t-1}$
MLR-13	$S_t = a_{13} +b_{10}R_t + b'_{10}R_{t-1} + c_{10}Q_t$
MLR-14	$S_t = a_{14} +b_{11}R_t + b'_{11}R_{t-1} + c_{11}Q_{t-1}$
MLR-15	$S_t = a_{15} + b_{12}R_t + b'_{12}R_{t-1} + c_{12}Q_t + c'_{12}Q_{t-1}$
MLR-16	$S_t = a_{16} + d_1S_{t-1}$
MLR-17	$S_t = a_{17} + b_{13}R_t + d_2S_{t-1}$
MLR-18	$S_t = a_{18} + b_{14}R_{t-1} + d_3S_{t-1}$
MLR-19	$S_t = a_{19} + b_{15}R_t + b'_{15}R_{t-1} + d_4S_{t-1})$
MLR-20	$S_t = a_{20} + c_{13}Q_t + d_5S_{t-1}$
MLR-21	$S_t = a_{21} + c_{14}Q_{t-1} + d_6S_{t-1}$
MLR-22	$S_t = a_{22} + c_{15}Q_t + c'_{15}Q_{t-1} + d_7S_{t-1}$
MLR-23	$S_t = a_{23} + b_{16}R_t + c_{16}Q_t + d_8S_{t-1}$
MLR-24	$S_t = a_{24} + b_{17}R_t + c_{17}Q_{t-1} + d_9S_{t-1}$
MLR-25	$S_t = a_{25} + b_{18}R_t + c_{18}Q_t + c'_{18}Q_{t-1} + d_{10} S_{t-1}$
MLR-26	$S_t = a_{26}+ b_{19}R_{t-1} + c_{19}Q_t + d_{11}S_{t-1}$
MLR-27	$S_t = a_{27}+ b_{20}R_{t-1} + c_{20}Q_{t-1} + d_{12}S_{t-1}$
MLR-28	$S_t = a_{28}+ b_{21}R_{t-1} + c_{21}Q_t + c'_{21}Q_{t-1} + d_{13}S_{t-1}$
MLR-29	$S_t = a_{29}+ b_{22}R_t + b'_{22}R_{t-1} + c_{22}Q_t + d_{14} S_{t-1}$
MLR-30	$S_t = a_{30}+ b_{23}R_t + b'_{23}R_{t-1} + c_{23}Q_{t-1} + d_{15}S_{t-1}$
MLR-31	$S_t = a_{31}+ b_{24}R_t + b'_{24}R_{t-1} + c_{24}Q_t + c'_{24}Q_{t-1} + d_{16}S_{t-1}$

*a_i, b_i, b'_i, c_i, c'_i and d_i are regression coefficients (i = 1, 2,…., 31)

Table 3. Training variables and their assigned values for CANFIS and MLP models

Training variables	Assigned values for CANFIS	Assigned values for MLP
Membership function	Gaussian, Bell	-
MFs per input	2 to 6	-
Fuzzy model	TSK	-
Activation function	Tanh	Tanh
Learning rule	Delta-Bar-Delta	Delta-Bar-Delta
Epoch	1000	1000
Training threshold	0.001	0.001

2.5 Training and Testing of MLP and MLR Models

The daily rainfall and sediment concentration (SC) data were divided into two sets: a training set spanning from 1997 to 2004, and a testing set spanning from 2005 to 2006 Hurdag watersheds. Utilising the back propagation technique for MLP, training was carried out on the network of a chosen architecture utilising the input pairs from the training data set (Table 3).

2.6 Performance Evaluation

Both qualitative and quantitative performance will be used to assess the models' performances that were constructed for this project. While the models' quantitative performance will be confirmed by estimating the values of statistical and hydrological indices like the correlation coefficient (CC), root mean square error (RMSE), coefficient of efficiency (CE), and coefficient of determination (R2), the models' qualitative performance will be assessed through visual observation (Aamir et al., 2019).

2.6.1 Correlation coefficient (r)

This is a number that represents the linear relationship between two variables; it ranges from -1.0 to +1.0. The correlation coefficient is equal to one if the two variables have a perfect linear connection with a positive slope (Singh et al., 2019). However, it takes a lot of work to calculate the measure from bigger observations. Karl Pearson's Correlation Coefficient between the observed and projected discharge is seen in equation 2.2.

$$r = \frac{\sum_{i=1}^{N}(X_{oi} - \overline{X_o})(Y_{pi} - \overline{Y_p})}{\sqrt{\sum_{i=1}^{N}(X_{oi} - \overline{X_o})^2 \sum_{i=1}^{N}(Y_{pi} - \overline{Y_p})^2}} \quad (2.2)$$

Where, $\overline{X_o}$ and $\overline{Y_p}$ are the mean of observed and predicted values, respectively.
A positive r indicates that the observed and predicted values tend to go up and down together (Ozer, 1985).

2.6.2 Root Mean Square Error (RMSE)

An overall level of agreement between the observed and simulated datasets is provided by this metric. With zero serving as the value for a perfect model, it has no upper bound. Although it can only be used to compare the prediction errors of many models for a given variable, RMSE is a useful metric for assessing accuracy (Wilby,1998). The RMSE between actual and expected values is displayed in Equation 2.3.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(X_{oi} - X_{pi})^2} \quad (2.3)$$

Where, X_{oi} and Y_{pi} are the observed and predicted values for i^{th} datasets and N is the total number of observations.

2.6.3 Coefficient of Efficiency (CE)

To assess the goodness of fit between observed and predicted values of runoff simulation, the CE was suggested by Nash and Sutcliffe (1970). The coefficient of efficiency is computed by the following equation;

$$\mathbf{CE} = \left[1 - \frac{\sum_{i=1}^{N}(X_{oi} - Y_{pi})^2}{\sum_{i=1}^{N}(X_{oi} - \overline{X_o})^2}\right] \quad (2.4)$$

2.6.4 Coefficient of determination (R^2)

It is a measurement used to explain how much variability of one factor can be caused by its relationship to another related factor (Barrett, 1974). This correction, is known as “goodness of fit” is represented as a value between 0.0 and 1.0 (Ozer, 1985).

$$\mathbf{R}^2 = \frac{SSR}{SST} \quad (2.5)$$

Where,
SSR = Sum of squared regression
SST = Total variation in data

3. RESULTS AND DISCUSSION

3.1 Sediment Modeling Using MLP and CANFIS

MLP and CANFIS models were used to simulate sediment concentration (SC) based on combinations of current and previous days’ rainfall, runoff, and prior day's SC.Ten models were selected for further analysis, and their performance was evaluated using RMSE, CE, and correlation coefficient (r). Results showed poor performance in MLP-25 and CANFIS-25 during testing due to hysteresis between SC and streamflow, where SC is higher during the rising stage of the hydrograph than the falling stage. Improvement is expected by including previous day’s SC in the input. An improvement in the simulation performance is expected by adding the previous day’s SC values into the input combinations, since the measurements of streamflow and SC are taken together at the same cross-section of the river (Alp and Cigizoglu, 2005).

As shown in Table 4, the RMSE for the ten selected MLP models during training ranged from 0.397 to 0.414 g/l, while for testing it ranged from 0.139 to 0.159 g/l. The coefficient of efficiency (CE) varied between 0.435 and 0.483 for training, and between 0.782 and 0.886 for testing. The correlation coefficient (r) ranged from 0.663 to 0.692 during training and from 0.927 to 0.990 during testing. Similarly, Table 5 shows that for the CANFIS models, the RMSE ranged from 0.305 to 0.413 g/l during training and from 0.118 to 0.175 g/l during testing. CE values ranged from 0.436 to 0.528 for training and from 0.737 to 0.881 for testing. The correlation coefficient (r) for CANFIS models was between 0.660 and 0.727 during training and

0.921 to 0.982 during testing. The higher CE and r values during testing indicate good generalization by both MLP and CANFIS models. Based on lower RMSE (0.115) and higher CE (0.886) and r (0.990) during testing, MLP-7 and CANFIS-7 were the best performing models, suggesting the current day's sediment concentration (SC) can be predicted using the current day's rainfall and runoff data. MLP-10 and CANFIS-10 also performed well, indicating SC depends on the previous day's rainfall and current day's runoff.

Table 4. Statistical indices for selected MLP sediment models during training and testing phase for Hurdag watershed

ModelNo.	Structure	Training			Testing		
		RMSE	CE	r	RMSE	CE	r
MLP-7	2-10-1	0.414	0.435	0.663	0.115	0.886	0.990
MLP-10	2-8-1	0.411	0.443	0.666	0.117	0.882	0.977
MLP-17	2-2-1	0.402	0.474	0.682	0.139	0.832	0.933
MLP-20	2-2-1	0.402	0.466	0.683	0.139	0.832	0.945
MLP-21	2-6-1	0.398	0.476	0.690	0.145	0.818	0.935
MLP-22	3-2-1	0.398	0.477	0.691	0.141	0.829	0.939
MLP-25	4-6-1	0.405	0.458	0.685	0.159	0.782	0.927
MLP-26	3-2-1	0.397	0.485	0.692	0.144	0.824	0.942
MLP-27	3-2-1	0.398	0.478	0.691	0.144	0.820	0.940
MLP-28	4-2-1	0.402	0.466	0.683	0.144	0.822	0.936

Table 5. Statistical indices for selected CANFIS sediment models during training and testing phase for Hurdag watershed

Model No.	MF per input	Training			Testing		
		RMSE	CE	r	RMSE	CE	r
CANFIS-7	Gauss-6	0.402	0.466	0.684	0.118	0.881	0.982
CANFIS-10	Gauss-2	0.413	0.436	0.660	0.131	0.852	0.979
CANFIS-17	Gauss-4	0.378	0.528	0.727	0.142	0.825	0.921
CANFIS-20	Gauss-3	0.390	0.497	0.707	0.159	0.783	0.933
CANFIS-21	Gauss-3	0.391	0.495	0.704	0.133	0.847	0.939
CANFIS-22	Gauss-2	0.398	0.477	0.690	0.147	0.814	0.934
CANFIS-25	Gauss-3	0.305	0.485	0.698	0.175	0.737	0.916
CANFIS-26	Gauss-3	0.397	0.479	0.692	0.149	0.807	0.945
CANFIS-27	Gauss-2	0.395	0.484	0.696	0.153	0.799	0.932
CANFIS-28	Gauss-3	0.391	0.494	0.704	0.155	0.794	0.932

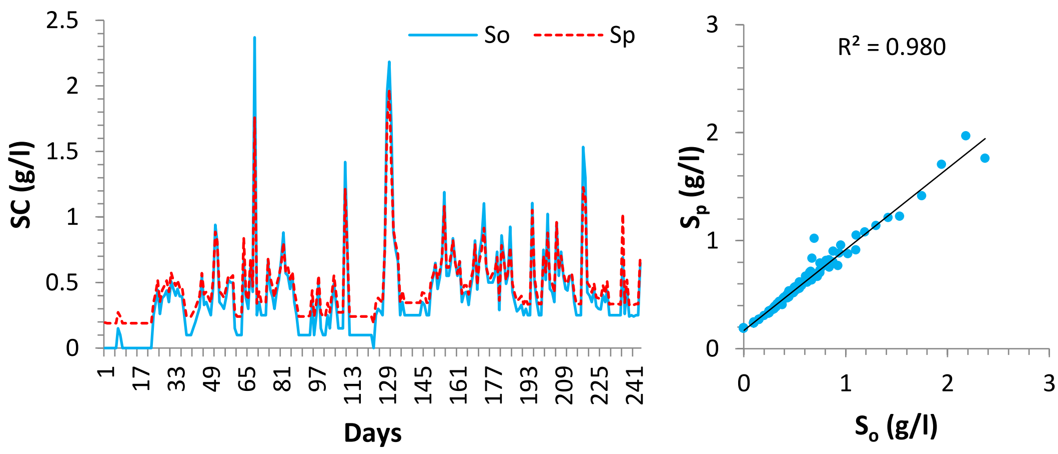

Fig. 4. Comparison of observed (S_o) and predicted (S_p) SC values and corresponding scatter plot in testing period by MLP-7 model for Hurdag watershed

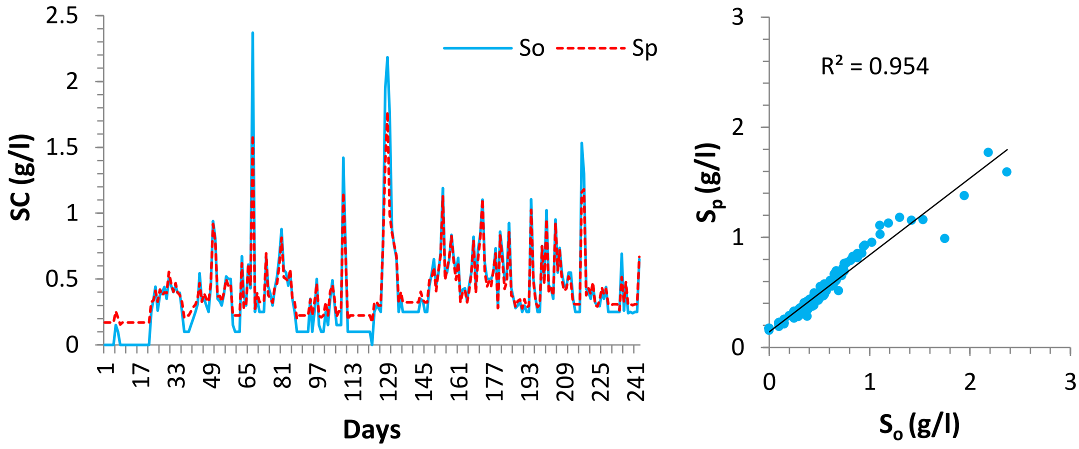

Fig. 5. Comparison of observed (S_o) and predicted (S_p) SC values and corresponding scatter plot in testing period by MLP-10 model for Hurdag watershed

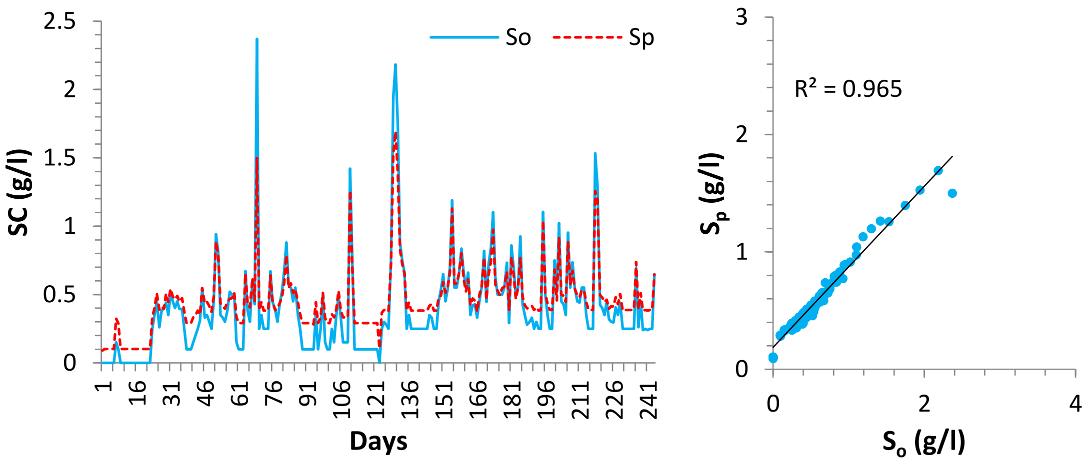

Fig. 6. Comparison of observed (S_o) and predicted (S_p) SC values and corresponding scatter plot in testing period by CANFIS-7 model for Hurdag watershed

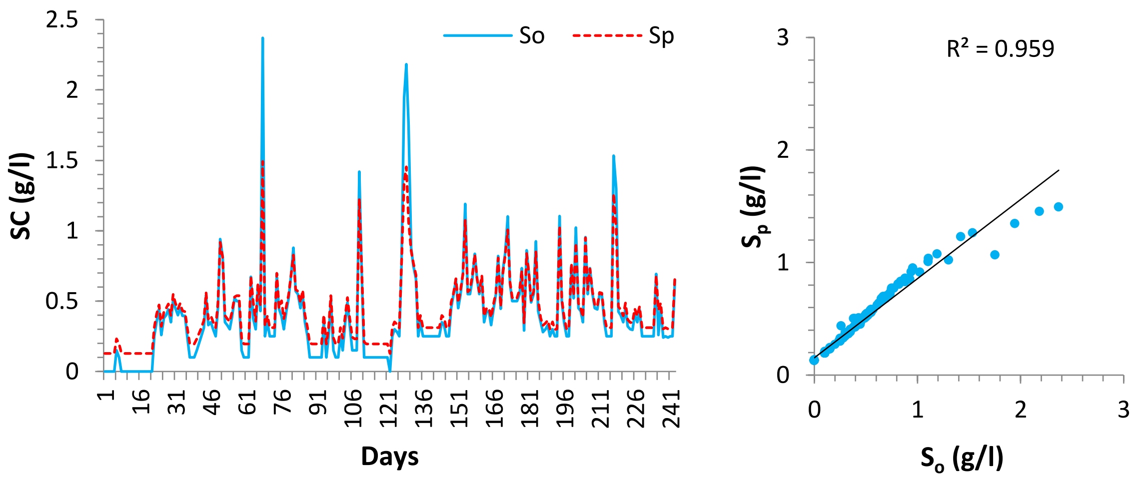

Fig. 7. Comparison of observed (S_o) and predicted (S_p) SC values and corresponding scatter plot in testing period by CANFIS-10 model for Hurdag watershed

Table 6. Statistical indices for selected MLR sediment models during testing phase for Hurdag watershed

Model No.	Regression equation	Statistical index		
		RMSE	CE	r
MLR-7	$S_t = 0.354 + 0.096\ Q_t + 0.011R_t$	0.281	0.327	0.632
MLR-10	$S_t = 0.364 + 0.116Q_t + 0.007R_{t-1}$	0.296	0.250	0.597
MLR-17	$S_t = 0.150 + 0.007R_t + 0.602S_{t-1}$	0.251	0.457	0.688
MLR-20	$S_t = 0.140 + 0.077Q_t + 0.608S_{t-1}$	0.269	0.376	0.619
MLR-21	$S_t = 0.184 - 0.008Q_{t-1} + 0.639S_{t-1}$	0.286	0.299	0.561
MLR-22	$S_t = 0.151 + 0.098Q_t - 0.051Q_{t-1} + 0.628S_{t-1}$	0.270	0.373	0.617
MLR-25	$S_t = 0.134 + 0.086Q_t - 0.052Q_{t-1} + 0.005R + 0.608S_{t-1}$	0.250	0.464	0.689
MLR-26	$S_t = 0.141 + 0.077Q_t - 0.0003R_{t-1} + 0.610S_{t-1}$	0.275	0.353	0.594
MLR-27	$S_t = 0.183 - 0.009Q_{t-1} + 0.0006R_{t-1} + 0.636S_{t-1}$	0.285	0.301	0.564
MLR-28	$S_t = 0.149 + 0.098Q - 0.052Q_{t-1} + 0.0006R_{t-1} + 0.626S_{t-1}$	0.269	0.376	0.620

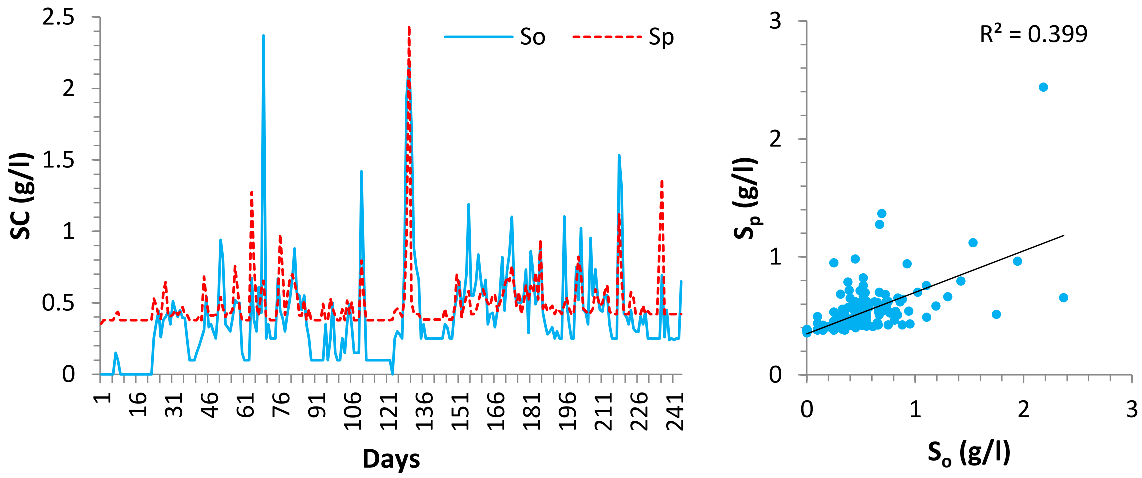

Fig. 8. Comparison of observed (S_o) and predicted (S_p) SC values and corresponding scatter plot in testing period by MLR-7 model for Hurdag watershed

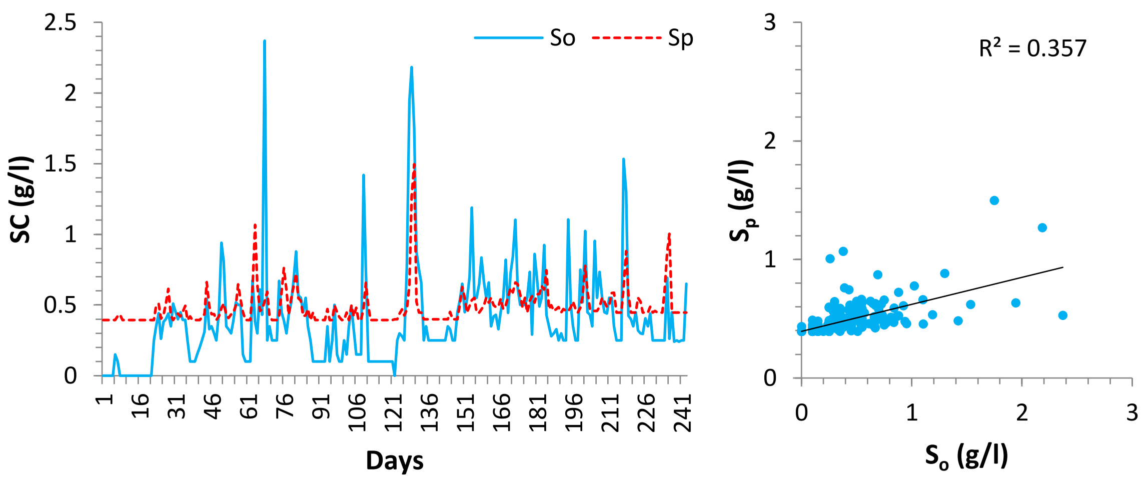

Fig. 9. Comparison of observed (S_o) and predicted (S_p) SC values and corresponding scatter plot in testing period by MLR-10 model for Hurdag watershed

The temporal variation of observed (S_o) and predicted (S_p) SC values simulated by the MLP-7 and CANFIS-7 models for testing period are compared using SC graph and corresponding scatter plot as shown in Figs. 4 and 5 for MLP models, and Figs. 6 and 7 for CANFIS models. These SC graphs indicate that all the models, in general, under-predict the peak values and over predict the lowest values of SC. The scatter plot also indicates that the simulated and observed sediment concentration during testing or validation period match very closely for MLP-7 model with R^2 value of 0.980 and CANFIS-7 model with R^2 value of 0.965.

3.2 Sediment Modelling Using MLR

Equations simulating the silt concentration of today were developed using the conventional statistical method known as MLR (Malik and Kumar, 2015). The input-output relationships for MLR models were developed using the input-output combinations of several MLR models (Table 2) for the research region. The performance evaluation index values for the various MLR models evaluated based on the RMSE, CE, and r values throughout the testing period are shown in Table 6. RMSE, CE, and r ranged from 0.250 to 0.296 g/l, 0.561 to 0.689, and MLR-7, MLR-10, MLR-17, MLR-20, MLR-21, MLR-22, MLR-25, MLR-26, MLR-27, and MLR-28 were the ten models that were chosen.It provides the generated input-output equations for these models. Based on the statistical criteria of the lowest RMSE and the highest (but not statistically significant) values of r and CE, the MLR-25 model was determined to be the best-performing model, closely followed by the MLR-17 model. As a result, the rainfall and runoff of the present day as well as the runoff and SC of the previous day determine the SC for the current day, according to the best-performing MLR-25 model.As the value of correlation coefficient of MLR-25 gives the best result.

The MLR models' simulated observed (So) and predicted (Sp) SC values for the testing period are compared using a SC graph and matching scatter-plot, as illustrated in Figs 8 and 9. These models generally underpredict the peak SC, a finding also supported by the scatter plots. The very low values of CE and r throughout the testing period plainly suggested that the MLR models were not appropriate for the prediction of SC for the studied watersheds.

3.3 Discussion

Sedimentation simulation is a crucial aspect of watershed management, especially in regions like the Hurdag watershed in the Damodar-Barakar basin, where sediment transport influences water quality, reservoir storage capacity, and overall hydrological dynamics. The objective of this study was to formulate Artificial Neural Network (ANN)-based Multi-Layer Perceptron (MLP), Co-Active Neuro-Fuzzy Inference System (CANFIS), and Multiple Linear Regression (MLR) models to simulate sediment yield and to evaluate the performance of these models using various statistical indices.

The models were trained using rainfall, runoff, and sediment data from 1995-2001 and tested on data from 2002-2003. Performance was evaluated using Root Mean Square Error (RMSE), Coefficient of Efficiency (CE), and Correlation Coefficient (r). Both MLP and CANFIS showed superior predictive accuracy, with lower RMSE and higher CE and r values compared to MLR. While MLR performed adequately during training, it failed to generalize well to unseen data, highlighting its limitations in simulating sediment yield (Agarwal, 2002, Agarwal et al, 2006, Agarwal, et al., 2003).

The observed (So) and predicted (Sp) sediment concentration (SC) values for the testing period, simulated by MLP-7 and CANFIS-7 models, are compared through SC graphs and scatter plots (Figs. 4 – 7). The SC graphs show that both models tend to under-predict peak values and over-predict the lowest values. The scatter plots reveal a strong match between observed and predicted SC values, with MLP-7 achieving an R^2 of 0.980 and CANFIS-7 an R^2 of 0.965 during the validation period.

The observed (So) and predicted (Sp) sediment concentration (SC) values from MLR models for the testing period, presented in SC graphs and scatter plots (Figs. 8 – 9), show that the models generally under-predict peak SC. The low CE and r values during testing confirm that MLR models are not suitable for accurate SC prediction in the study watersheds (Aqil, et al., 2007, Rumelhart and McClelland, 1986, Werbos, 1974).

The results confirm that MLP and CANFIS are better suited for sediment yield simulation in the Hurdag watershed, offering more accurate and reliable predictions than MLR. These advanced techniques can significantly improve sediment management and hydrological forecasting in similar watersheds.

4. CONCLUSION

In this study, we attempted to forecast the daily Sedimentation on the basis of Multi Layer Perceptron (MLP), Co-Active Neuro-Fuzzy Inference System (CANFIS) and Multi Linear Regression (MLR) techniques for the Hurdag watershed in the Damodar-Barakar basin, located in Hazaribagh district, Jharkhand, India.

The performance of models was evaluated qualitatively by visual observation and various statistical and hydrological indices viz.root mean square error (RMSE), correlation coefficient (r) and coefficient of efficiency (CE). The model having higher values of correlation coefficient and coefficient of efficiency and low value of mean square error is consider as the best fit model.

The following conclusions were drawn from the results in this study:

- The MLP model used to simulate silt concentration by providing input values for the runoff and rainfall of the current day; similarly, runoff can be simulated by providing input parameters for the rainfall of the day before.
- The results obtained on the basis of statistical indices (RMSE, CE, r and R^2) indicates that the MLP and CANFIS model in general gave consistently better performance than the MLR models
- It was abundantly obvious that the MLR model suited the dataset under investigation extremely poorly.

DISCLAIMER (ARTIFICIAL INTELLIGENCE)

Author(s) hereby declare that NO generative AI technologies such as Large Language Models (ChatGPT, COPILOT, etc) and text-to-image generators have been used during writing or editing of this manuscript.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

Aamir, M., Srivastava, S. K., Singh, V., & Denis, D. M. (2019). Suspended sediment simulation using ANN based generalized feed forward neural network (GFF) and multi-linear regression method (MLR). *International Research Journal of Engineering and Technology (IRJET)*, 6. e-ISSN: 2395-0056, p-ISSN: 2395-0072.

Agarwal, A. (2002). *Artificial neural networks and their application for simulation and prediction of runoff and sediment yield* (Ph.D. Thesis, Department of Soil and Water Conservation Engineering, G. B. Pant University of Agriculture and Technology, Pantnagar).

Agarwal, A., Mishra, S. K., Sobha Ram, & Singh, J. K. (2006). Simulation of runoff and sediment yield using artificial neural networks. *Biosystems Engineering, 94*(4), 597-613.

Agarwal, A., Singh, J. K., & Ray, S. (2003). Artificial neural network rainfall-runoff modeling in varying domain. *Institution of Engineers (I) Journal, 83*, 166-172.

Alp, M., & Cigizoglu, H. K. (2005). Suspended sediment load simulation by two artificial neural network methods using hydrometeorological data. *Environmental Modelling & Software, 22*(1), 2-13.

Aqil, M., Kita, I., Yano, A., & Nishiyama, S. (2007). A comparative study of artificial neural networks and neuro-fuzzy in continuous modeling of the daily and hourly behaviour of runoff. *Journal of Hydrology, 337*, 22-34.

Barrett, J. P. (1974). The coefficient of determination—some limitations. *The American Statistician, 28*(1), 19-20.

Haykin, S. (1999). *Neural networks: A comprehensive foundation*. Prentice Hall.

Jang, J. S. R., Sun, C. T., & Mizutani, E. (1997). Neuro-fuzzy and soft computing: A computational approach to learning and machine intelligence [Book Review]. *IEEE Transactions on Automatic Control, 42*(10), 1482-1484.

Maier, H. R., & Dandy, G. C. (2000). Neural networks for the prediction and forecasting of water resources variables: A review of modelling issues and applications. *Environmental Modelling & Software, 15*(1), 101-124.

McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics, 5*, 115-133.

Ministry of Water Resources, Government of India. (1999). *Integrated water resources development: A plan for action. Report of the National Commission for Integrated Water Resources Development*.

Ozer, D. J. (1985). Correlation and the coefficient of determination. *Psychological Bulletin, 97*(2), 307.

Rajurkar, M. P., Kothyari, U. C., & Chaube, U. C. (2004). Modeling of the daily rainfall-runoff relationship with artificial neural network. *Journal of Hydrology, 285*(1-4), 96-113.

Rumelhart, D. E., & McClelland, J. L. (1986). *Parallel distributed processing: Explorations in the microstructures of cognition*. Cambridge, MA: MIT Press.

Saemi, M., & Ahmadi, M. (2008). Integration of genetic algorithm and a coactive neuro-fuzzy inference system for permeability prediction from well logs data. *Transport in Porous Media, 71*, 273-288.

Singh, K., Singh, V., Srivastava, S. K., & Wesley, C. J. (2019). Rainfall simulation using ANN based generalized feed forward and MLR technique. *International Research Journal of Engineering and Technology (IRJET), 6*(7). ISSN: 2395-0056, p-ISSN: 2395-0072.

Singh, R., & Panda, R. K. (2012). Daily sediment yield modeling with artificial neural network and support vector machines. *Journal of Hydrology, 424-425*, 367-383.

Singh, V., & Kumar, A. (2016). Rainfall-runoff prediction using co-active neuro fuzzy inference system (CANFIS) and multi-linear regression (MLR) technique for Karso watershed. *International Journal of New Agriculturist, 27*(1).

Werbos, P. J. (1974). *Beyond regression: New tools for prediction and analysis in behavior sciences* (Ph.D. Thesis, Harvard University, Cambridge, Mass).

Wilby, R. L., Wigley, T. M. L., Conway, D., Jones, P. D., Hewitson, B. C., Main, J., & Wilks, D. S. (1998). Statistical downscaling of general circulation model output: A comparison of methods. *Water Resources Research, 34*(11), 2995-3008.

Zhang, G. P. (2000). Neural networks for classification: A survey. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 30*(4), 451-462.

Zurada, J. M. (1992). *Introduction to artificial neural systems*. West Publishing Co.

Peer-review history:
The peer review history for this paper can be accessed here:
https://www.sdiarticle5.com/review-history/124077