iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 Issue: VII Month of publication: July 2021

DOI: https://doi.org/10.22214/ijraset.2021.36470

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VII July 2021- Available at www.ijraset.com

Rainfall Simulation using ANN based Multilayer Perceptron (MLP) and Multiple Linear Regression (MLR) Technique for Bhopal, Madhya Pradesh

Prashant Singh Tomar[1], Dr. Vikram Singh[2], Tushar Rathod[3]
[1]*PG Scholar, SHUATS Prayagraj, UP*
[2]*Assistsnt Professor, SHUATS, Prayagraj, UP*
[3]*Ph.D Scholar, SHUATS, Prayagraj, UP*

Abstract: Rainfall forecasting represents a tremendously significant matter in field of hydrology. In this study, was undertaken to develop and evaluate the applicability of Multilayer Perceptron (MLP) and Multi Linear Regression (MLR) techniques. The performance of the developed models, on the basis of training and testing, was judged on the basis of four statistical measures such as Root Mean Squared Error (MSE), Coefficient of Efficiency (CE), Correlation Coefficient (r) and Coefficient of Determination(R^2) during monsoon period (June to September) for Bhopal, Madhya Pradesh, India. The daily data of minimum temperature, maximum temperature, wind speed and relative humidity were used for rainfall prediction. The appropriate parameter combination of input variables for MLP was used to predict rainfall. The Neuro Solution 5.0 software and Microsoft Excel were used in analysis and the performance evaluation of developed models, respectively. The input pairs in the training data set were applied to the network of a selected architecture and training was performed using back propagation algorithm for MLP models was designed with Gaussian membership function, Takagi- Sugeno- Kang fuzzy model, hyperbolic tangent activation function and Delta-Bar-Delta learning algorithm. Ten MLP models and MLR were selected based on the performance evaluation indices during testing period. MLP models were found to be much closer to the observed values of rainfall as compared to MLR.
Keywords: Soft computing, MLP, MLR, Rainfall Prediction.

I. INTRODUCTION

Rainfall is one of the most complex and difficult elements of the hydrology cycle to understand and model due to the complexity of atmospheric processes in its formation and occurrence. Rainfall in India is dependent on the south-west and north-east monsoons based on shallow cyclonic depressions and disturbances, and on the local storms. An accurate long-term rainfall prediction is necessary for water resources management, food production and maintaining flood risks. Rainfall models play a significant role in water resource management, planning and hydraulic design. The main purpose of this paper is to compare and analyze the performance of the MLP and MLR to see their applicability in rainfall forecasting.

ANN was first developed in the 1940s and the development has experienced a renaissance in iterative auto-associable neural networks (Hung et al, 2008). An ANN provides the user a model free tool, which can generate input output mapping for any set of data as complex pattern recognition can be attempted without making any initial assumptions. In addition, ANN could learn and generalize from examples to produce meaningful solution even when the input data contain errors or is incomplete. In recent years, ANNs have been used intensively for prediction and forecasting in a number of water-related areas, including water resource study(Najah et al., 2009;Ahmed et al.,2009;El-Shafie et al.,2007(a),2009(b);2010),hydrograph simulator (Deka and Chandramouli, 2005; rainfall estimating (Lin and Chen 2005; Luk et al.,2001).

The Multi-layer perceptron (MLP) is the most popular ANN architecture in use today (Dawson and Wilby, 1998). As a network, MLP was formed by simple neuron called perceptron, which computes a single output from multiple real valued inputs by forming a linear combination according to its input weights and then possibly expressing the output through a nonlinear transfer function. The MLP is a widely used ANN configuration that has been frequently applied in the field of hydrological modelling (Leahy *et al*. 2008; Tabari *et al*. 2010b; Zadeh and Tabari 2012).

A Multi Linear Regression (MLR) is the simplest, statistical and well-developed representation of the time-invariant relationship between an input function and corresponding output function. MLR models are considered as benchmark in reservoir inflow forecasting (Chavan and Ukrande, 2012).

II. LITERATURE REVIEW

Sorman *et al.* (2016) employed two different ANN models (MLP and RBF) and compared with each other using novel MODIS satellite snow covered area products as an alternative input into climatic data based models. Both models were performed to estimate the daily flows of Karasu River in the Upper Euphrates Basin, Turkey using 2002 – 2011 data. The main difference between the RBF network and MLP network is in the nature of the nonlinearities associated with hidden nodes. The nonlinearity in MLP is implemented by a fixed function such as a sigmoid. On the other hand, the RBF method bases its nonlinearities on the training set data. They investigated the determination of model architectures, optimization algorithms and methods to avoid overfitting are elaborated.

Chanu and Kumar (2017) compared the performance of two artificial neural network (ANN)- multilayer perceptron (MLP) and radial basis function (RBF) for modeling daily rainfall-runoff in a Himalayan watershed called Bino watershed situated at Almora and Pauri Garhwal districts of Uttarakhand, India. The time series monsoon data of rainfall and runoff between 2000 and 2009 were used to train and test the models. The best input combination was selected by gamma test (GT) technique. The performance of both the MLP and RBF neural network models were comprehensively evaluated in terms of indices viz. correlation coefficient (r), root mean square error (RMSE) and coefficient of efficiency (CE). The results of the study indicate that the choice of the network type has certainly an impact on the prediction accuracy of model. Both models performed satisfactorily for runoff predictions; however, the MLP model outperformed the RBF model. Results show that ANN models are useful tools for rainfall-runoff modelling the hydrologic response with good accuracy in the study watershed.

Anurag Malik et al. (2018) Radial Basis Neural Network(RBNN), Self-organizing map neural network(SOMNN), and Multiple Linear Regression(MLR) were used for the estimation of pan evaporation at Pantnagar located at the foothills of Himalayas in the Uttarakhand, India . Daily climatic data & pan evaporation data were used for model calibration & validation. Combination of significant input variables for RBNN, SOMNN & MLR models were deci.ded using Gamma test. Results obtained by models were compared with climate-based Emperical models such as Penman, Stephens-Stewart and Jensen-Burman-Allen models on the basis of Root mean squared Error(RMSE), Coefficient of efficiency (CE) & Correlation Coefficient(r)

Anurag Malik et al. (2019) The heuristic approaches including Co-Active Neuro fuzzy Inference system (CANFIS), multilayer Perceptron Neural Network (MLPNN) and Multiple Linear Regression (MLR) were utilized to predict the hydrologic draught based on Multi-Scalar Stream Flow Draught Index (SDI) at Naula and Kedar stations located in upper Rāmgangā River basin, Uttarakhand state, India. The predicted values of multi-scalar SDI using CANFIS, MLPNN and MLR models were compared with the calculated values, based on root mean squared error (RMSE), Nash-Sutcliffe efficiency (NSE), Coefficient of Correlation (COC) and Wilmot Index (WI). The visual interpretation was also made using line diagram, scatter diagram and Taylor Diagram (TD). The MLR model was found to be the best at 24-month time scale for Naula station only. The result is helpful in prediction of hydrological drought on multiple time scales and decision making for remedial schemes to cope with hydrological drought at Naula and Kedar stations.

III. MATERIALS AND METHOD

The area selected for the study is Bhopal which is located in Madhya Pradesh state of India. The coordinates of Bhopal are 23°15'N 77°25'E and 1,729 ft elevation. The daily data of runoff for monsoon season June 2004 to September 2013 were used. Total data of monsoon season for 10 years were divided into two sets: (i) training data set consisting of first 8 years data from 1 June 2004 to 31 September 2011; and (ii) testing data set consisting of remaining 2 years data from 1 June 2012 to 31September2013.

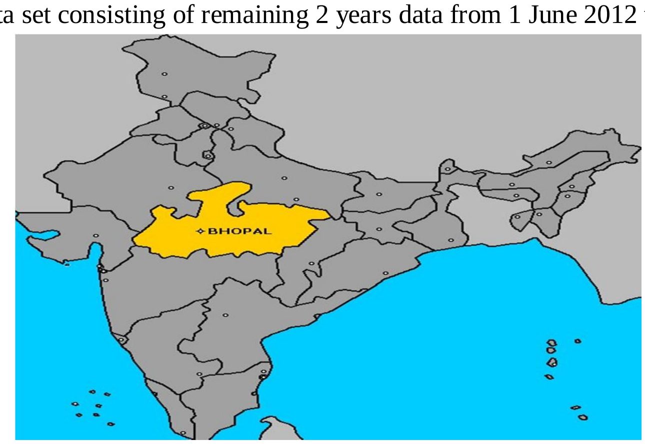


Figure 1. Map of Bhopal in India.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VII July 2021- Available at www.ijraset.com

In this study, the soft computing technique based on ANN, Multilayer Perceptron (MLP) and Multi Linear Regression (MLR) has been developed for predicting the total rainfall in Bhopal, Madhya Pradesh. The methodology of developing the MLP models along with training and testing of developed models, the Neuro Solution 5.0 software and Microsoft Excel were used in analysis and the performance evaluation indices for developed models.

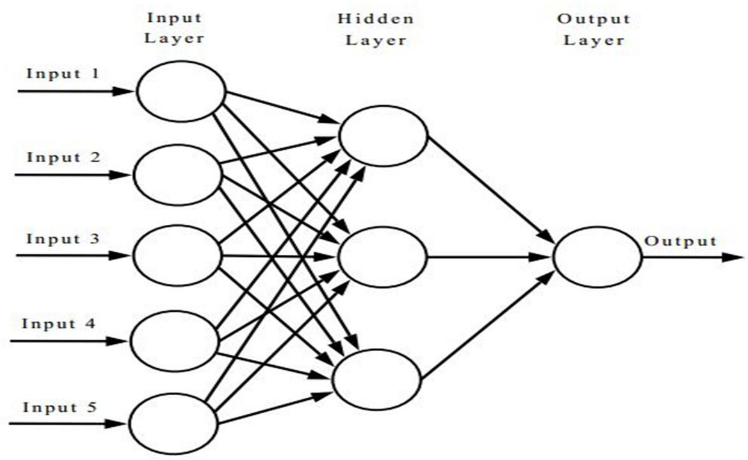


Figure 2. A basic overview of MLP

Identification of input and output variables is the first step for developing the MLP and MLR models. The output from the models is the Rainfall R_t at a daily time step t based on respective input variables. The various output-input combinations of MLP and MLR models for the study areas are listed in Tables 1 & 2.

Model No.	Output-Input Variables*
MLP-1	$R_t = f(T_{max})$
MLP-2	$R_t = f(T_{min})$
MLP-3	$R_t = f(WS)$
MLP-4	$R_t = f(RH)$
MLP-5	$R_t = f(T_{max}, T_{min})$
MLP-6	$R_t = f(T_{max}, WS)$
MLP-7	$R_t = f(T_{max}, RH)$
MLP-8	$R_t = f(T_{min}, WS)$
MLP-9	$R_t = f(T_{min}, RH)$
MLP-10	$R_t = f(WS, RH)$
MLP-11	$R_t = f(T_{max}, T_{min}, WS)$
MLP-12	$R_t = f(T_{max}, WS, RH)$
MLP-13	$R_t = f(T_{min}, WS, RH)$
MLP-14	$R_t = f(T_{min}, T_{max}, RH)$
MLP-15	$R_t = f(T_{max}, T_{min}, WS, RH)$

Table 1. Input-output combination for MLP models for rainfall prediction at Bhopal city of Madhya Pradesh.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VII July 2021- Available at www.ijraset.com

The daily data of rainfall and meteorological data (maximum temperature, minimum temperature, wind speed, relative humidity) on daily basis were split into two sets: a training data set from 2004 to 2011 and a testing data set from 2012 to 2013 for Bhopal.
The model attempts to reproduce the outcome based on the learning or training on data input information. In ANN model of the biological neurons, there are three basic components such as:

1) *Synapses:* The synapses of the neuron are modelled as weights which represent the strength of the connectivity between an input and a neuron.
2) *Adder:* This activity is referred to as linear combination which sums up all the inputs modified by their respective weights and is the actual activity within the neuron cell.
3) *Activation Function:* The activation function controls the amplitude of the output of the neuron

A. *Multiple Linear Regression (MLR)*

Multiple Linear Regression (MLR) is simply extended form of Simple regression in which two or more variables are independent variables are used and can be expressed as **(Kumar and Malik, 2015**)

$Y = \alpha + \beta 1\ X1 + \beta 2\ X2 + \dots \dots \dots \dots \beta p\ Xp$ (3.1)Where,

Y= Dependent variable;
α = Constant or intercept;
β1 = Slope (Beta coefficient) for X1;
X1 = First independent variable that is explaining the variance in Y;
β2 = Slope (Beta coefficient) for X2;
X2 = Second independent variable that is explaining the variance in Y;
P = Number of independent variables;

Model No.	Input-Output Variables
MLR-1	$Q_t = e_1 + g_1T_{min}$
MLR-2	$Q_t = e_2 + g_2Tmax$
MLR-3	$Q_t = e_3 + h1Ws$
MLR-4	$Q_t = e_4 + h2Rt$
MLR-5	$Q_t = e_5 + g_3T_{min} + g_{3'}T\ max$
MLR-6	$Q_t = e_6 + g_4T_{min}, + h3Ws$
MLR-7	$Q_t = e_7 + g_5T_{min} + h4Rt$
MLR-8	$Q_t = e_8 + g_6T_{max} + h5Ws$
MLR-9	$Q_t = e_9 + g_7T_{max} + h6Rt$
MLR-10	$Q_t = e_{10} + h7Ws + h8\ Rt$
MLR-11	$Q_t = e_{11} + g_8T_{min} + g_9T_{max} + h9Ws$
MLR-12	$Q_t = e_{12} + g_{10}T_{min} + h10Ws + h11Rt$
MLR-13	$Q_t = e_{13} + h6T_{max} + k7Ws + I6Rt$
MLR-14	$Q_t = e_{14} + g_7T_{min} + h7\ T_{max} + I7Rt$
MLR-15	$Q_t = e_8 + g_8T_{min} + h8\ T_{max} + k8Ws + I8Rt$

Table 2. Input-output combination for MLR models for rainfall prediction at Bhopal

B. Performance Evaluation

Performance of the models developed in this study will be evaluated by using qualitative performance and quantitative performance. Qualitative performance of the models will be checked by the visual observation, whereas, quantitative performance will be verified by estimating the values of statistical and hydrological indices such as Correlation Coefficient (r), Mean Square Error (MSE), Coefficient of Efficiency (CE) and Coefficient of Determination(R^2).

C. Correlation Coefficient (r)

This is a number between -1.0 and +1.0, which means the degree to which two variables are linearly related. If there is a perfect linear relationship with a positive slope between the two variables, the correlation coefficient is equal to 1. The measure, however, is very intensive to derive from larger observations. Equation 3.2 denotes Karl Pearson's Correlation Coefficient between observed and predicted discharge.

$$r = \frac{\sum_{i=1}^{N}(X_{oi}-\bar{X}_o)(Y_{pi}-\bar{Y}_p)}{\sqrt{\sum_{i=1}^{N}(X_{oi}-\bar{X}_o)^2\sum_{i=1}^{N}(Y_{pi}-\bar{Y}_p)^2}} \qquad (3.2)$$

Where, $\overline{Xo}$ 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.

D. Mean Square Error (RMSE)

The mean square error (MSE) is determined to measure the prediction accuracy. It always produces positive values by squaring the errors. The MSE is zero for perfect fit and increased values indicate higher discrepancies between predicted and observed values (Wilks, 1995). The Mean Square Error between observed and predicted values is determined by the following equation:

$$MSE = \frac{\sum_{j=1}^{n}(Y_j-Y_{ej})^2}{n} \qquad (3.3)$$

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

E. 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;

$$\textbf{CE} = \left[1 - \frac{\sum_{i=1}^{N}(X_{oi} - Y_{pi})^2}{\sum_{i=1}^{N}(X_{oi} - \overline{X}_o)^2}\right] \qquad (3.4)$$

F. 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. This correction, is known as "goodness of fit" is represented as a value between 0.0 and 1.0.

$$\mathbf{R^2} = \frac{SSR}{SST} \qquad (3.5)$$

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

IV. RESULTS AND DISCUSSION

The performances of models were evaluated qualitatively and quantitatively by visual observation and various statistical and hydrological indices viz. Mean Square Error (MSE), Correlation Coefficient (r), Coefficient of Efficiency (CE) and Coefficient of Determination (R^2). 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.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VII July 2021- Available at www.ijraset.com

A. Rainfall Modelling using MLP

MLP models (Table 3) were used to predict daily rainfall as output based on various input combinations of minimum and maximum temperature, wind speed and relative humidity. MLP3(1-10-1), MLP12(3-10-1), MLP8(2-8-1), MLP1(1-10-1), MLP4(1-10-1), MLP3(1-8-1), MLP6(2-10-1), MLP1(1-8-1), MLP13(3-10-1) and MLP4(1-6-1) were selected for further analysis and comparison based on the statistical indices, such as mean squared error (MSE), coefficient of efficiency (CE), correlation coefficient (r) and coefficient of determination(R^2). The values of statistical indices for the selected MLP models during testing are presented in Tables 3 respectively.

Table 3. Statistical Indices for selected MLP models during testing phase for Bhopal

Sr. No.	Model	Combination	Training				Testing			
			MSE	R	CE	R^2	MSE	r	CE	R^2
1	N3	(1-10-1)	0.0245	0.3135	0.091	0.0965	0.0246	0.9928	0.4404	0.9856
2	N12	(3-10-1)	0.0226	0.5123	0.1612	0.1657	0.0271	0.8661	0.4104	0.7501
3	N8	(2-8-1)	0.0225	0.4325	0.165	0.1852	0.0293	0.8633	0.3337	0.7453
4	N1	(1-10-1)	0.0225	0.4193	0.165	0.1766	0.0271	0.8506	0.3834	0.7235
5	N4	(1-10-1)	0.0223	0.4841	0.1727	0.1758	0.0261	0.8231	0.4056	0.6775
6	N3	(1-8-1)	0.0225	0.4222	0.1646	0.1772	0.0288	0.8090	0.3446	0.6545
7	N6	(2-10-1)	0.0217	0.4511	0.1955	0.2038	0.0281	0.7801	0.361	0.6086
8	N1	(1-8-1)	0.0212	0.4643	0.214	0.2162	0.0256	0.7784	0.4167	0.6059
9	N13	(3-10-1)	0.0220	0.4316	0.1845	0.1857	0.0271	0.7500	0.3827	0.5625
10	N4	(1-6-1)	0.0211	0.4680	0.2193	0.2197	0.0265	0.7411	0.3973	0.5492

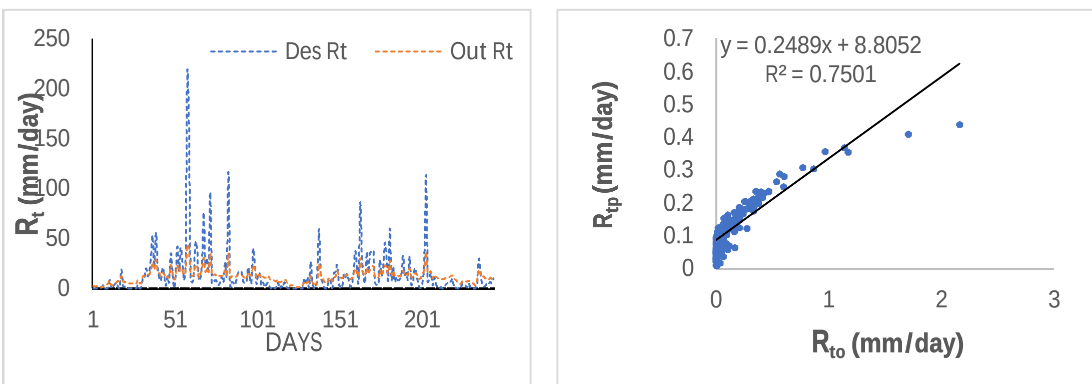


Figure 3. Comparison of observed (Rto) and predicted (Rtp) line and corresponding scatter plot by MLP -3(1-10-1)during testing period

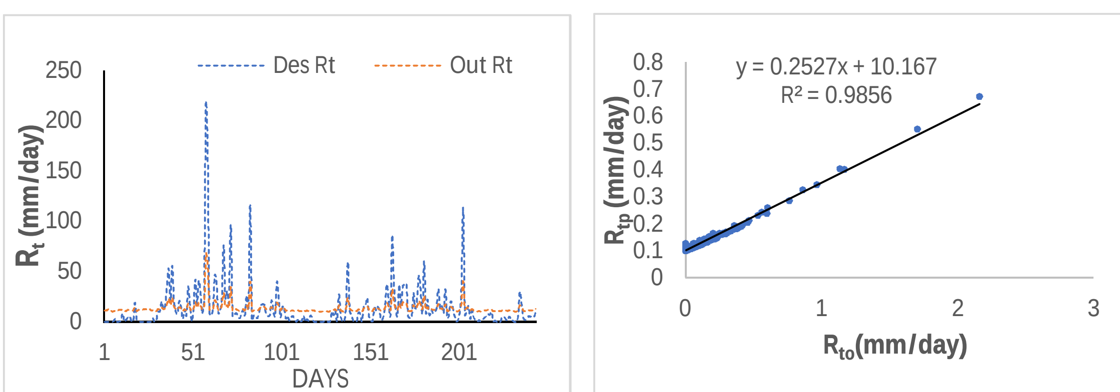


Figure 3. Comparison of observed (Rto) and predicted (Rtp) line and corresponding scatter plot by MLP -12(3-10-1) during testing period

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VII July 2021- Available at www.ijraset.com

B. Rainfall Modelling using MLR

Universal processes of forecasting rainfall amount involve Data collection, data pre-processing and data selection, Reduction of explanatory predictor, building model using regression and at the last validity check.

Table 4. Statistical indices for selected MLR models during testing phase for Bhopal

Model No.	Regression equation	Statistical index			
		MSE	CE	r	R^2
M15	R_t = =-92.1581+(-0.74912* Tmax)+(3.572505* Tmin)+(1.237285* WS)+(51.63094*RH)	307.3596	0.2267	0.4762	0.2267
M14	R_t = -81.2635 +(-1.01223 * Tmax) + (3.758351 * Tmin) + (46.58151 * RH)	308.2144	0.2246	0.4739	0.2246
M13	R_t = -124.412 + (3.370882 * Tmin) + (1.93115 * WS) + (67.27404 * RH)	308.7244	0.2233	0.4726	0.2233
M9	R_t = -125.255 + (3.594456 * Tmin) + (68.13873 * RH)	311.3173	0.2168	0.4656	0.2168
M11	R_t = 17.3554+(-2.62265* Tmax)+(3.251558* Tmin)+(-0.35754* WS)	318.6250	0.1984	0.4454	0.1984
M5	R_t = 17.29444 + (-2.59623 * Tmax) + (3.179729 * Tmin)	318.7067	0.1982	0.4452	0.1982
M10	R_t = -23.8903+(3.312827*WS)+(35.8916*RH)	336.6954	0.1529	0.3911	0.1529
M6	R_t = 49.56889+(-1.37308* Tmax)+(2.03768*WS)	343.3922	0.1361	0.3689	0.1361
M7	R_t = 11.05681+(-0.50432* Tmax)+(22.17762*RH)	343.9187	0.1348	0.3671	0.1348
M4	R_t = -13.4019+(33.70029*RH)	344.7062	0.1328	0.3644	0.1328

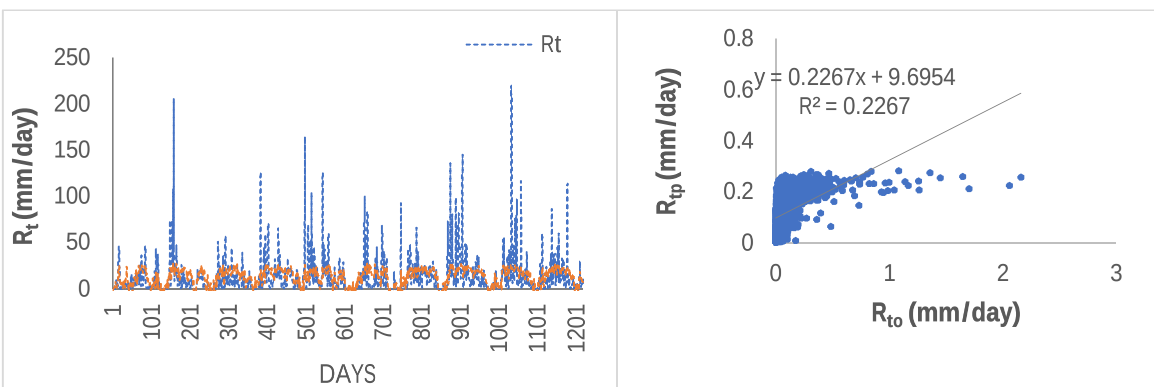


Figure 5. Comparison of observed (Rto) and predicted (Rtp) line and corresponding scatter plot by MLR-15 during the testing period

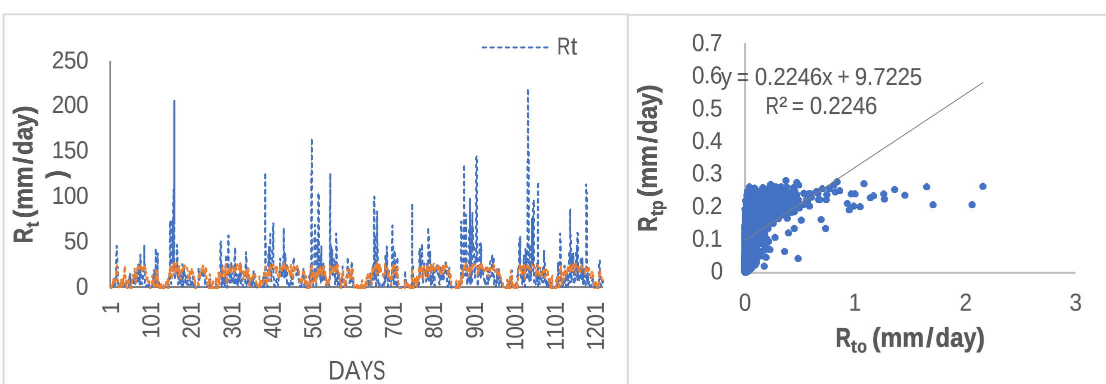


Figure 6. Comparison of observed (Rto) and predicted (Rtp) runoff and corresponding scatter plot by MLR-14 during the testing period

V. CONCLUSION

In this study, we attempted to forecast the daily rainfall on the basis of Multilayer Perceptron (MLP) and Multiple Linear Regression (MLR) techniques for Bhopal, Madhya Pradesh. Daily weather data were collected from the site swat.tanu.edu.

Fifteen MLP models and MLR were selected based on the performance evaluation indices during testing period. The following conclusions were drawn from the results in this study:

A. On the basis of lower MSE value and higher CE and r values, MLP-3 model was found to be the best model followed by MLP-12 model. According to the MLP-3 model, current day's wind speed depends on current day's rainfall.

B. The MLP model with input of maximum temperature, minimum temperature, wind speed and relative humidity was found to be the best for prediction of rainfall.

C. It was clearly evident the Multi Linear Regression Method is not suitable for the dataset under this study.

REFERENCES

[1] Ahmed AN, Elashafie A, Karim O, Jaffar O (2009). Evaluation the efficiency of Radial Basis Function Neural Network for Prediction of water quality parameters. Eng. Intelligent Syst., 17(4): 221-231

[2] Badaoui, A. Abdallaoui and S. Chabaa (2016). Optimization numerical the neural architectures by performance indicator with LM learning algorithms J. Mater. Environ. Sci. 8(1) 169-179.

[3] Chavan P. M, S. K. Ukarande 2012 Comparison of Rainfall- Runoff Model using ANN and MLR. Water Res. Manage., 23(11): 2289-2315.

[4] Deka P, Chandramouli V (2005). Fuzzy neural network model for hydrologic flow routing. J. Hydrol. Eng., 10(4): 302-14.

[5] El-Shaflie A, Reda TM, Noureldin A (2007). A Neuro-Fuzzy Model for inflow Forecasting of the Nile River at Aswan High Dam. Water Res. Manage., 21(3): 533-556.

[6] Lin GF, Chen LH (2005). Application of an artificial neural network to typhoon rainfall forecasting. Hydrol. Process, 19: 1825-37.

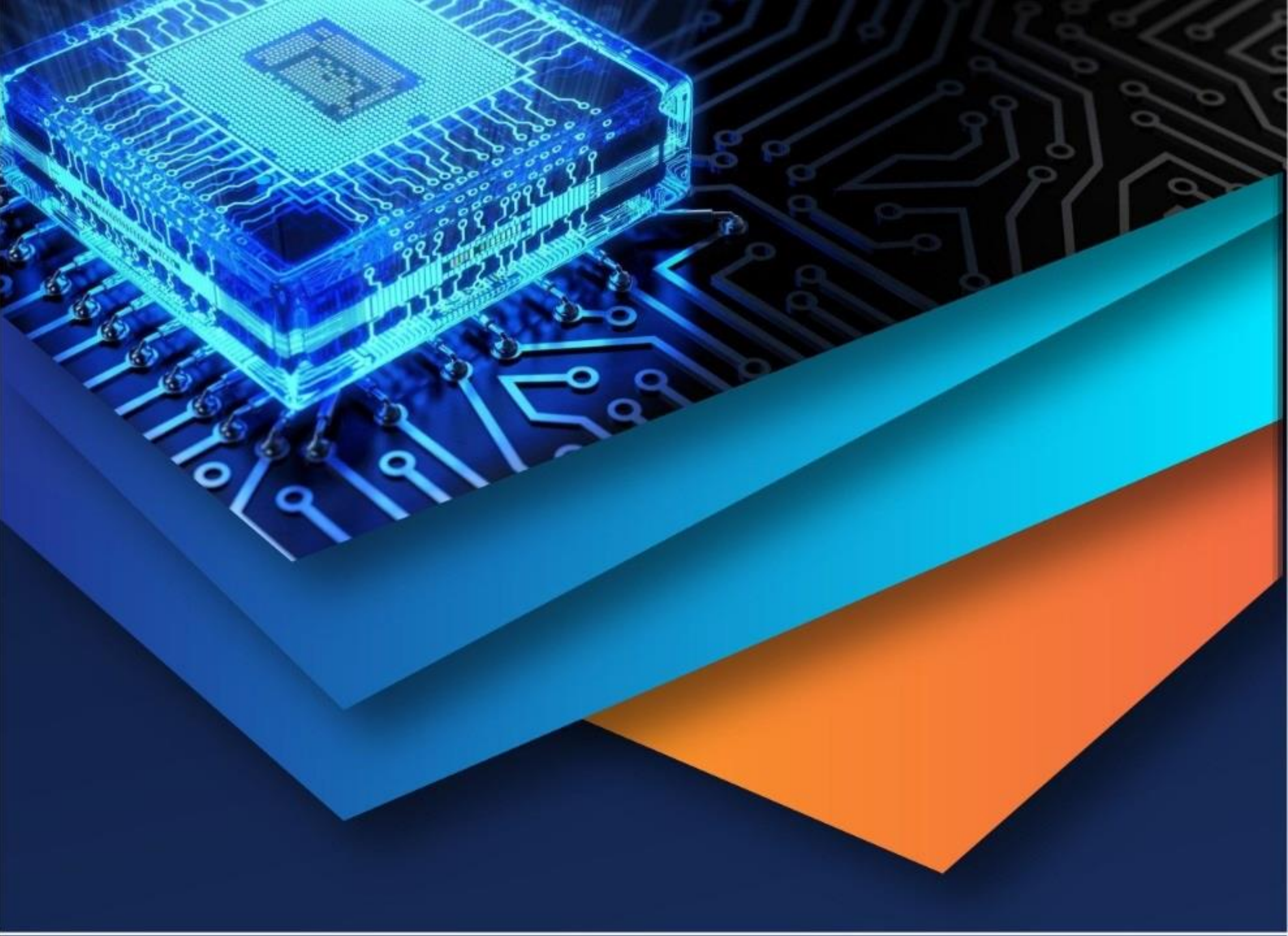
[7] Luk KC, Ball JE, Sharma A (2001). An application of artificial Neural networks for rainfall forecasting. Math. Compute. Modell. 33: 683-93.

[8] Malik Anurag, kumar Anil, Singh P.Rajesh, 2019 Application of heuristic approaches for hydrologicaldrought using multi-scalar streamflow drought index. 22, 4-9.

[9] Malik, A. and kumar , A. 2015 " Co-active Neuro-Fuzzy Inference System (CANFIS) and Multiple Linear Regression (MLR) based Suspended Sediment Modelling" ,Journal of Indian Water Resources Society, Vol 35.

[10] Najah A, Elshafie A, Karim OA, Jaffar O (2009). Prediction of johor river water quality parameters using artificial neural networks. Euro. J. Sci. Res., 28(3): 422-435.

[11] Samkele S. Tfwala and Yu-Min Wang (2016). Estimating Sediment Discharge Using Sediment Rating Curves and Artificial Neural Networks in the Shiwen River, Taiwan http://www.mdpi.com/journal/water Water 8, 53; doi:10.3390.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY