Rainfall prediction using artificial neural network and sequential modelling

[1]Prasun Kumar, [2]Shubham Kumar, [3]Pappu Kumar, [4]Dr. Mani Bhushan

[1]Student civil engineering department at RRSD college of Engineering, Begusarai, Bihar, India,
[2]Student civil engineering department at RRSD college of Engineering, Begusarai, Bihar, India,
[3]Research Scholar, NIT Patna
[4]Assistant professor, Department of Civil Engineering RRSD College of Engineering, Begusarai, Bihar

Abstract: The multilayer artificial neural network is mostly used in Prediction related projects using Back Propagation Algorithm as it's easy to use and due to its high accuracy result. There are two stages in its processing and learning cycle, the first one brings forth the input, and the other one matches the output by changing its weights. Three different ANN model proceed with Back Propagation Neural Network, Layer Recurrent Network, and Cascaded Back Propagation Network is developed, and the best training function and adaptive learning function is determined among these three models. Weather, financial Prediction, Face recognition, signature detection, and character recognition are some applications of the Feed Forward Neural Network. The objective of this paper is to search the best training function and adaptive learning function for the best result and best network with the Least Mean Square Error (MSE). The region of Patna (Bihar) has been selected to analyse the rainfall data, and using an artificial neural network with various weather indices, different characteristics of the hidden neurons in the system are used to study with the two-layer model used for training.

***Keywords*: Rainfall Prediction, Patna, Back Propagation Algorithm, Layer Recurrent Network, Cascaded Back Propagation Mean square error.**

I. INTRODUCTION

There is a vital role of agriculture in the growth of the Indian economy and the growth of the economy depends on agriculture. In the sector of the Indian economy, the Prediction of rainfall is essential for agricultural development in the country. Prediction of rainfall is an essential and arduous operational liability to be carried out by meteorological departments in the world. A tremendous amount of earlier and actual data from previous data records is required for Prediction. The work is challenging as all conclusions are to be taken in the feature of uncertainty. The consequences of precipitation (mainly rainfall) for our civilization are vast. Nowadays, Artificial Neural Networks are used extensively in various aspects of engineering as it can be used to model linear as well as non-linear systems without taking into consideration the multiple assumptions that are used in many statically methods. The historical data are used for Prediction for the ability of examining and determining for the use of artificial neural network for forecasting. Statical and mathematical model has not much accuracy as ANN has. The above model works on the principle of biological neurons. ANN can analyze meteorological parameters and rainfall. Mostly used method for training neural networks is the Back propagation method. It is overseen that for learning and making up training sets, it requires a datasheet of the required output from several inputs. It is also applicable for the Feed-Forward network, different characteristics of the hidden neurons in the network is used to study with two-layer model used for training. The objective of this project is to develop ANN model using Back Propagation algorithm, Layer Recurrent Network and Cascaded Back Propagation algorithm to predict rainfall in Patna region of Bihar and to find the best adaptive learning function and best training function among these with least Mean Square Error.

II. STUDY AREA

The region of Patna (Bihar) has been selected to analyse using artificial neural network models with various weather indices; Patna is one of the 38th districts of Bihar. Patna is located on the bank of the holy river Ganga in the Eastern part of India. The length of the city is approximately 35 km (22 miles) and width nearly 16 to 18 km (9.9 to 11.2 miles). The average elevation of Patna is about 174 ft (53 m). Cartographic coordinates of Patna are 25.6 degrees North and 85.1 degrees East. The primary monsoon season begins in June, and the rainiest months are July and August. The temperature here averages 26.0 degrees Celsius and Precipitation here is about 1031 mm (40.6 inches) per year.

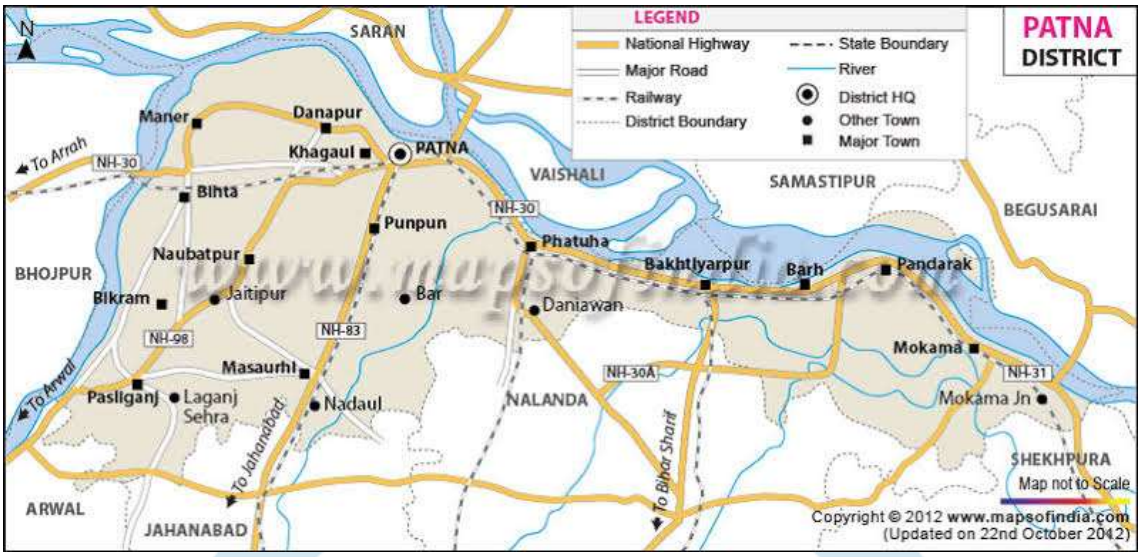


Fig 1. Map of the study area

III. LITERATURE REVIEW

[1] Initiated the concept of using ANN in weather forecasting and developed an ANN model for weather forecasting. Adaptive systems called Adaline for pattern recognition is used in this method. In the last few decades, in ANN huge increment in the field of application of ANN has opened up new approaches in Prediction including phenomena related to environment. [2] This research work presents the comparison of the performance of ANN with multi linear Regression in estimating missing rainfall data and Cyprus. Researchers developed a model using which series of rainfall can be generated for those places where the time series is discontinued or where the archive started recently using forward extension and back extension concept respectively. ANN is used to estimate the daily rainfall at a particular site in Cyprus termed as target station with the help of daily rainfall data obtained from surrounding sites, which had a large and sufficient archival data from observation networks. Suspected data can also be checked using the records from neighboring stations. The concept of using neural network is in contrast to the traditional multiple linear regression method. [3] In this research work ANN with Scaled Conjugate Gradient Algorithm (ANN-SCGA) is used and another neural network called Evolving fuzzy Neural Network (EfuNN) is used to predicate the rainfall time series. Monthly rainfall data as input data were used to train the model. The researchers investigated 87 years of data of Kerala, India. It is found that neuro-fuzzy networks were efficient as it showed better performance with minimum error compared to the neural network approaches.
[4] In this research work, a prediction model was developed using the concept of ANN to predict monsoon rainfall. The model was based on the Backpropagation learning technique. In its execution, the input data were rainfall data of summer monsoon months, and average summer rainfall data of the same year were used as target data. After training and testing, it was found that a network having 11 nodes in the hidden layer is more efficient for rainfall forecasting. After that, a comparison between the performance of the neural network having 11 hidden nodes with a three-layered model and the implementation of the asymptotic regression technique was made. Finally, they concluded that the eleven hidden nodes with a three-layered model are more efficient than the asymptotic regression technique. [5] In this research work, an ANN model was developed for rainfall prediction for Bangkok, Thailand, with a lead time of 1-6 hrs. An ANN model was developed with four years of hourly data from seventy-five rain gauge stations from surrounding areas. They finally used that model for real-time analysis of flood management and rainfall prediction. In this research paperwork, a prediction model for Udupi, Karnataka, was developed with average monthly rainfall data as input and output data. In Udupi, the Main monsoon was identified from May to October with light rainfall in April and November [2]. The researchers collected and manipulated the rainfall data of 8 months (April to November) from 1960 to 2010. Three different algorithms: Back Propagation Algorithm (BPA), Layer Recurrent Network (LRN), and Cascaded Back Propagation (CBP), were tested. The researchers concluded that out of these three networks, BPA gave the best result with high accuracy and least MSE. Finally, they concluded that BPA is the premier algorithm among these three networks.

IV. THEORETICAL BACKGROUND

An Artificial Neural Network is a computational structure. The motivation behind neural networks came from the way the human brain works, i.e., motivated by the study of biological neural processing. It accepts input data and generates appropriate output after interpretation [6]. Adaptation is realized by adjustment of the internal network connections by applying a certain algorithm. The objective of ANN is to process the data in a way that is previously trained to generate the appropriate and relevant result.

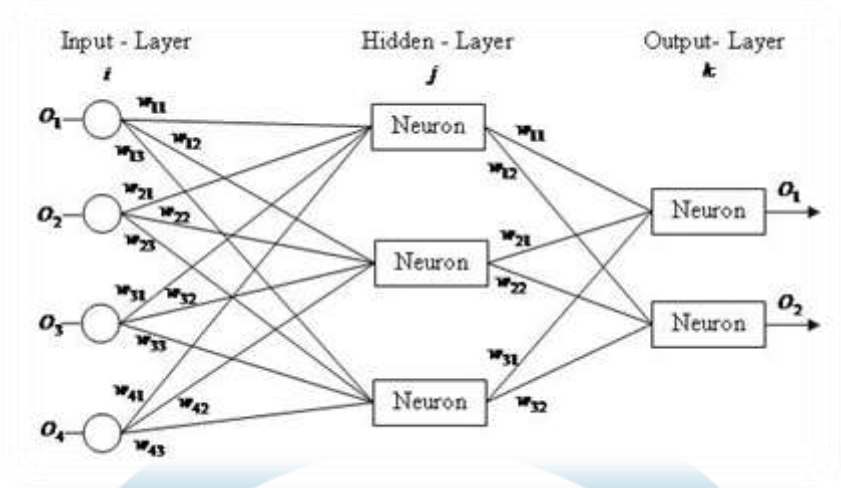


Fig 2. Layout of feed-forward neural network

Algorithms used in this project are Back Propagation Algorithm, Layer Recurrent Network, and Cascaded Back Propagation Algorithm [7].

V. METHODOLOGY

There are four sequential steps in the model building process; they are as follow:

a. Data Collection
b. Data Normalization
c. Training, Testing, and Validation
d. Comparison of Data Obtained

1.1.1 Data Collection

There are some research data for input and target for Patna, Bihar. The input parameters are average temperature, average relative humidity, moderate pressure, and average wind speed. The output parameter is average rainfall from May to October.

Rainfall can be of two types:

a. Seasonal Rainfall
b. Non-seasonal Rainfall

The months from June to September are recognized as main seasonal monsoon and the rest of the months are non-seasonal rainfall, sometimes with light rainfall in May and October. So, data for Six months from May to October from 1980 to 2019 are considered. The data are collected from the IMD website and the National Centers for Environmental Prediction (NCEP) or http://www.soda-pro.com.

It is so hard to forecast daily basis data and maintain accuracy; that's why the monthly average of these data is taken into consideration. These parameters should be authentic to generate minimum Mean Square Error. Computation and analysis of hourly data or daily basis data is a challenging work to keep the purity of these parameters that may lead to fallacy in training, testing, and validation.

1.1.2 Data Normalization

Input and Output data has different units; hence it has to be normalized. The mean/average of these data (i.e., temperature, relative humidity, pressure, wind speed, and rainfall) is formulated for Normalization.

The formula used in Normalization:

$$\textbf{Mean (M)} = \frac{\textbf{Sum of all entries}}{\textbf{Total number of entries}}$$

And

$$\textbf{Standard Deviation (SD)} = \sqrt{\frac{\sum(x-\mu)^2}{N}}$$

The standard deviation (SD) and the mean (M) of each parameter are calculated. After calculating the Standard Deviation and Mean of each parameter, the values of each weather indices taken as parameters are normalized using the following formula.

$$\textbf{Normalized value} = \frac{(x-M)}{SD}$$

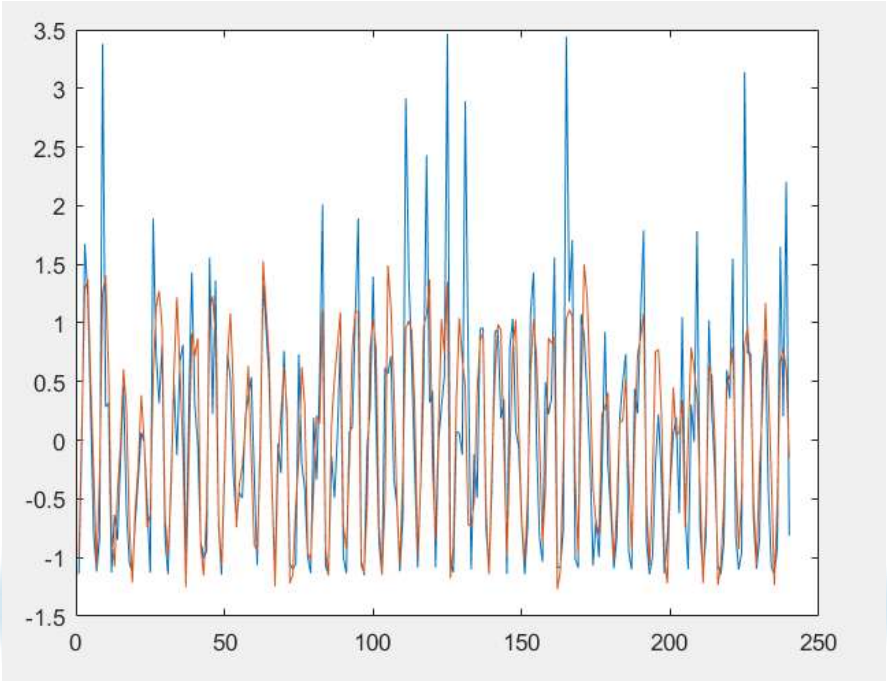
It helps in preserving the relationship between the actual data values.

1.1.3 Training, Testing and Validation

After obtaining the normalized data and creating a database, the data are imported in the workspace of Matlab to train the networks with different network properties and adaptive learning functions. Various combinations are made by changing network properties, adaptive learning functions, and the number of neurons. The proposed model is two-layered. The next step is to train the network in Matlab. Training is done after the completion of data training when the error is between acceptable limits. Training, Testing, and Validation are done using NFTOOL and NNTOOL in Matlab [8].

1.1.4 Comparison of Data Obtained

All the results are stored in the workspace after the testing and validation part is over. A graph is plotted between target and predicted outputs to make a comparison between target and predicted outcomes. Plotting a graph is the best method to compare actual result and predicted output. The accuracy of the model can be analyzed using these graphs. The chart is plotted between target and predicted output values, which shows a high degree of resemblance between mark and predicted output values. This indicates that the ANN model developed is entirely accurate in forecasting.



Graph 4.1 Comparison graph between the target data and predicted data

The above graph shows a comparison between the target and the predicted data. It has a minimum error and high accuracy.

VI. RESULT AND DISCUSSION

Two different tools were used for the implementation of the algorithm in Matlab. They are as follows:

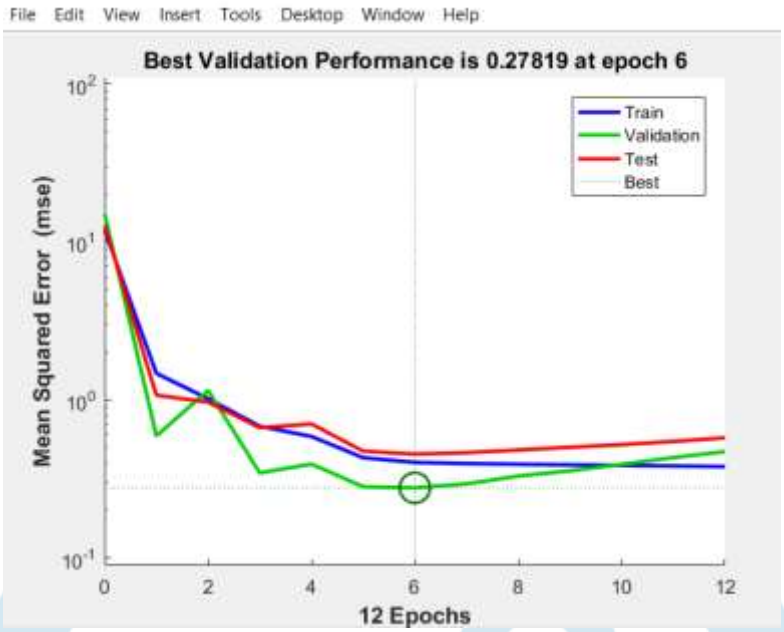
a. NFTOOL
b. NNTOOL

1. Implementation Of Back Propagation Using NFTOOL

Back Propagation Algorithm (BPA) was implemented in the NFTOOL. The performance graph and Regression are plotted to analyze the model. The MSE Obtained is:

MSE=**0.3917**

The performance can be plotted as:



Graph 6.1 Performance graph of BPA using NFTOOL

The Regression can be plotted as:

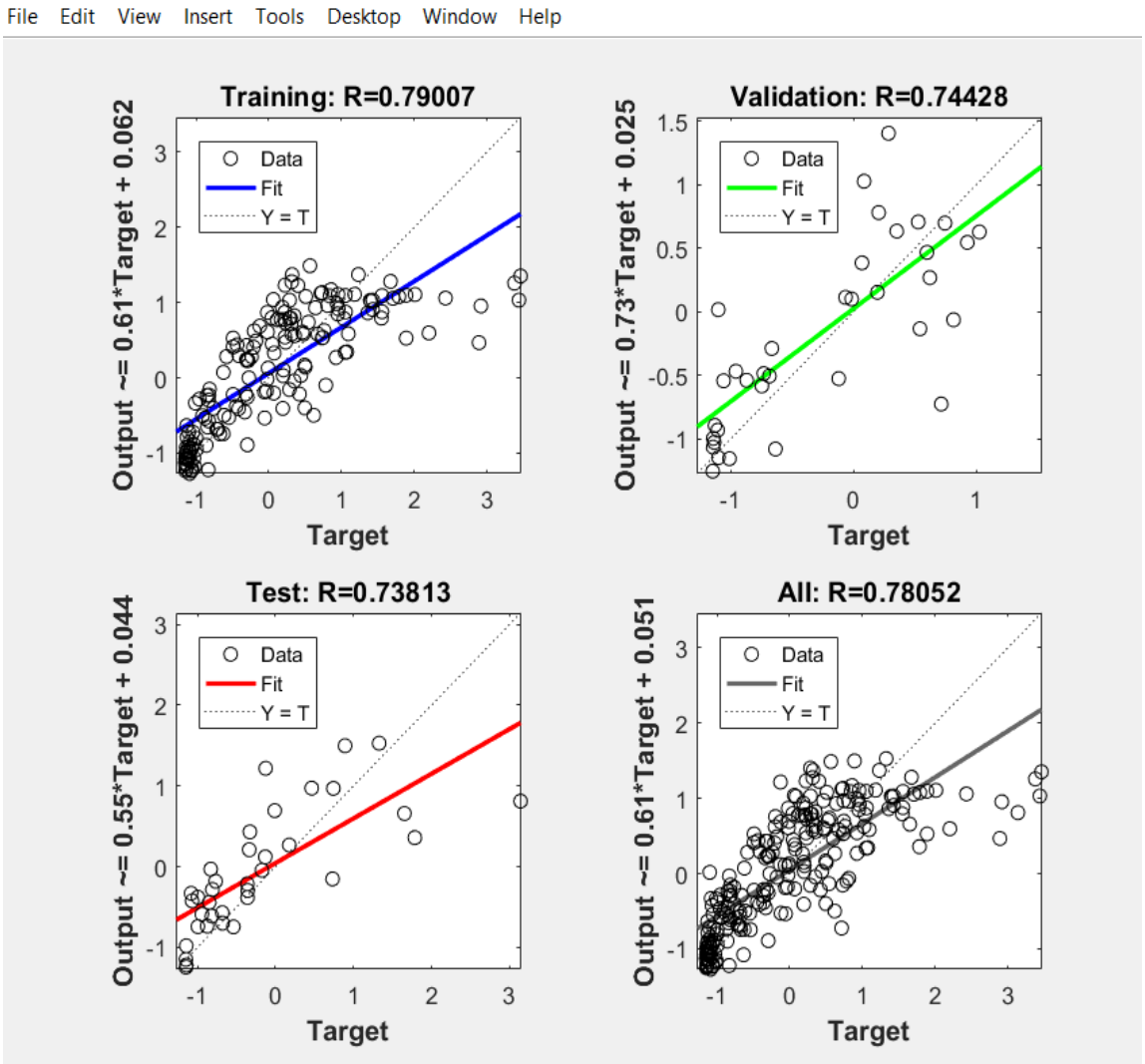


Fig 6.1 Regression using NFTOOL

The above graph showed outcomes when BPA was implemented using NFTOOL. The MSE obtained is between acceptable limits. It has a minimum error and high accuracy.

2. Implementation of Multilayer Architecture using NNTOOL

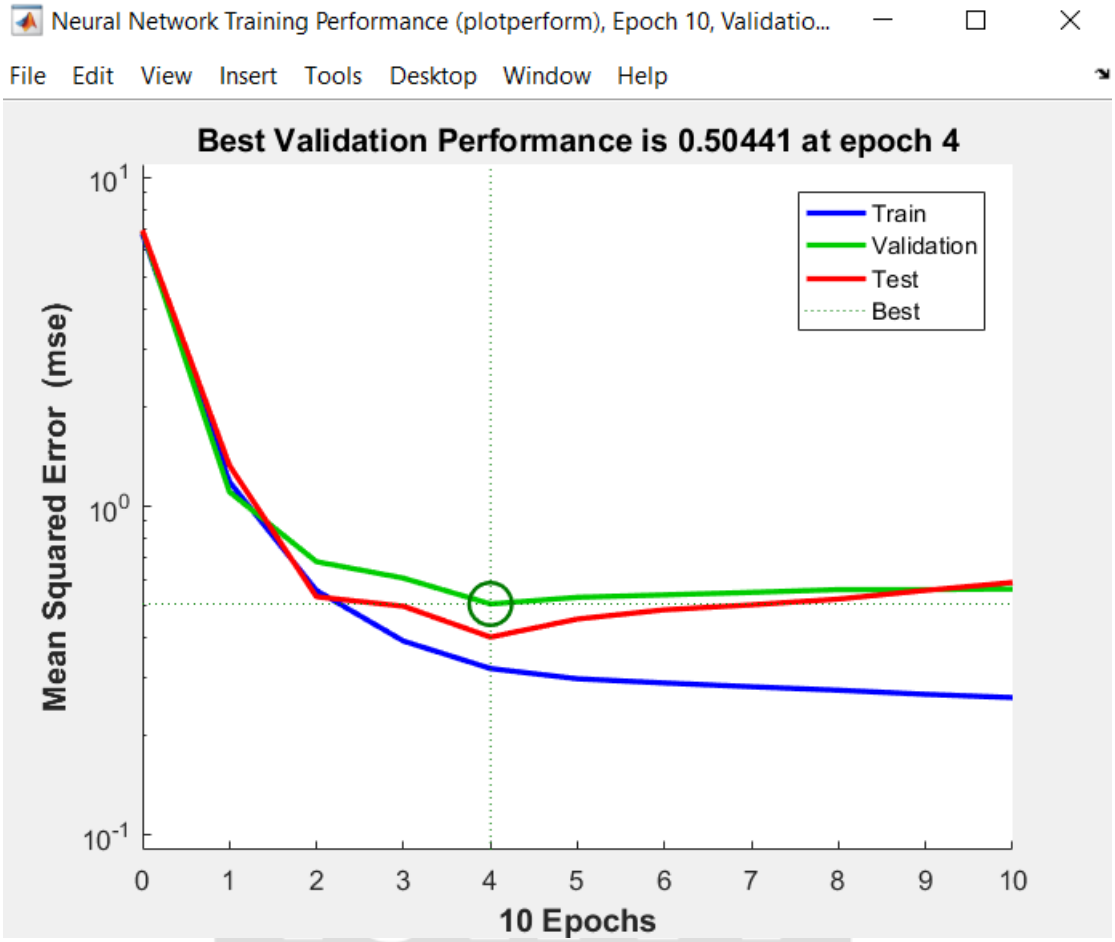
Multilayer architecture was tested using the following algorithms.

1. Back Propagation Algorithm
2. Layer Recurrent network
3. Cascaded Backpropagation

Best Training and Testing Case of **BPA**:

Training Function : TRAINLM
Adaptive Learning Function: Learned
No. Of Neurons : 20
MSE : 0.3602

The performance can be plotted as:



Graph 6.2 Performance graph of Back-propagation for case 7

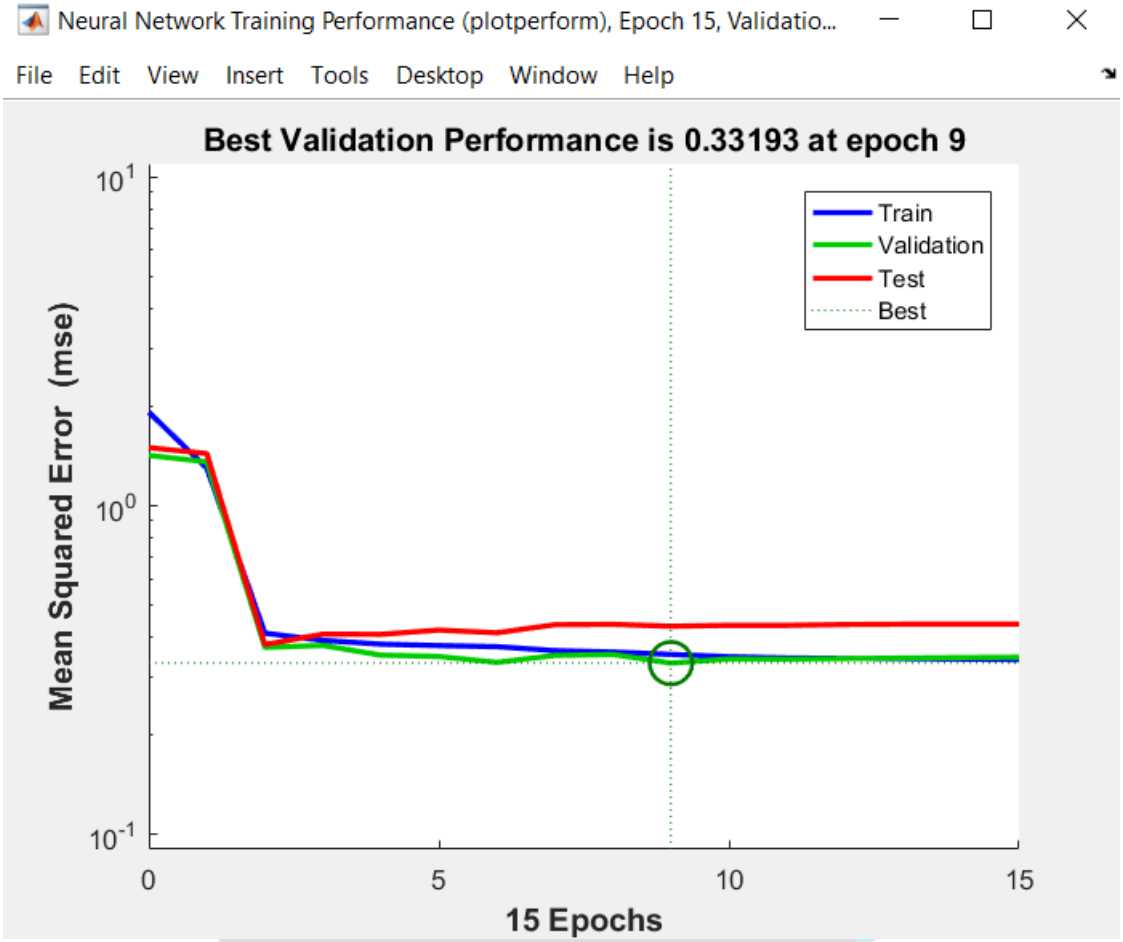
Back propagation algorithm was implemented using NNTOOL with input and Target data using two different training functions (TRAINLM & TRAINRP) and two different adaptive learning functions (Learngdm & Learngd) with 10 & 20 numbers of neurons. The MSE obtained in each case are shown in Table 6.1.

TABLE 6.1.BACK-PROPAGATION TESTING CASES

Case	Training Function	Adaptive Learning Function	No. Of Neurons	MSE
1	TRAINLM	Learngdm	10	0.4079
2	TRAINRP	Learngdm	20	0.4844
3	TRAINLM	Learngd	10	0.3848
4	TRAINRP	Learngd	20	0.5313
5	TRAINLM	Learngdm	20	0.3892
6	TRAINRP	Learngdm	10	0.3984
7	TRAINLM	Learngd	20	**0.3602**
8	TRAINRP	Learngd	10	0.4492

Case 7 with "TRAINLM" as a training Function and "Learngd" as an adaptive learning function with 20 numbers of neurons has the least MSE. So, case 7 is the premier case of BPA of this model when implemented in NNTOOL. The performance graph for the best possibility of Back-Propagation is shown in graph 6.2, and the best performance is observed at epoch 4.

Best Training and Testing Case of **LRN**:
Training Function : TRAINLM
Adaptive Learning Function : Learngd
No. Of Neurons : 20
MSE : 0.3612
The performance can be plotted as:



Graph 6.3 Performance graph of Layer-Recurrent for case 7

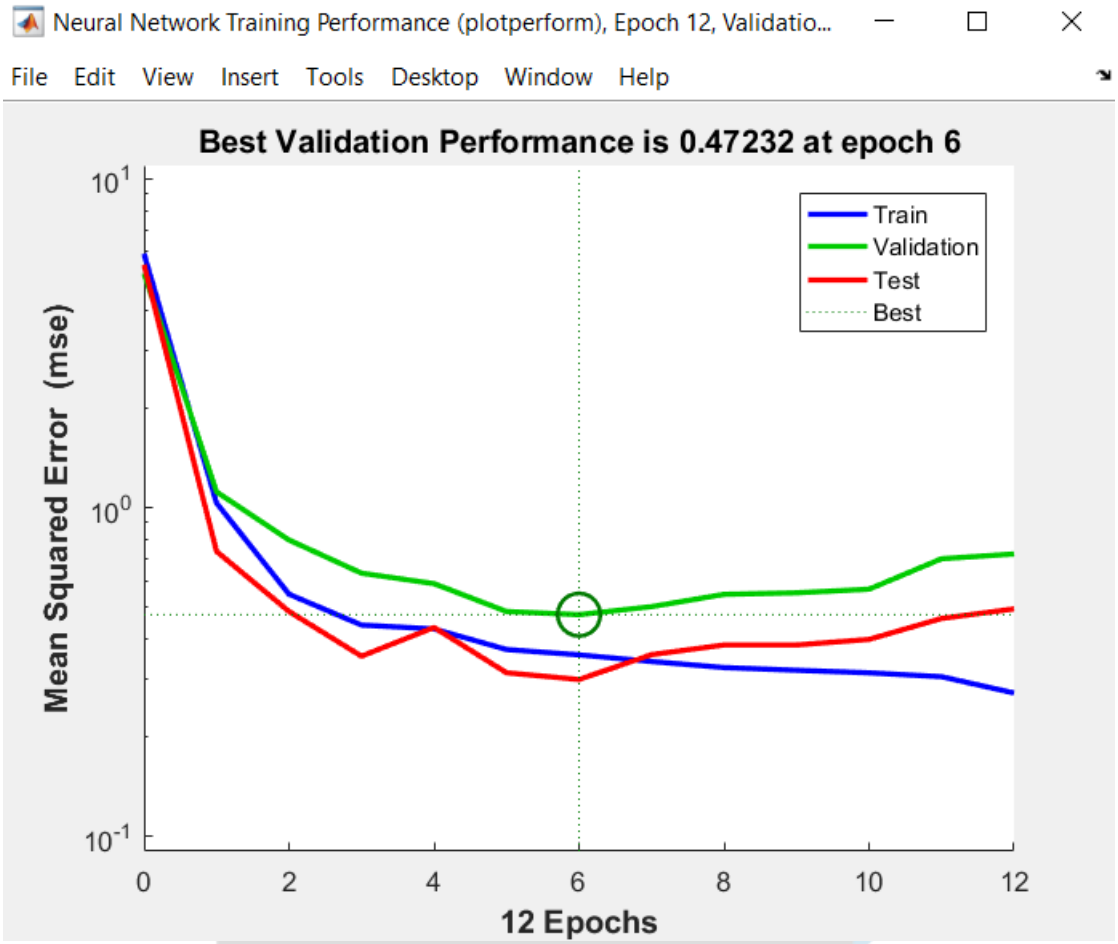
Layer Recurrent Network was implemented using NNTOOL with input and Target data using two different training functions (TRAINLM & TRAINRP) and two other adaptive learning functions (Learngdm & Learngd) with 10 & 20 numbers of neurons. The MSE obtained in each case are shown in Table 6.2.

TABLE 6.2 LAYER RECURRENT NETWORK TESTING CASES

Case	Training Function	Adaptive Learning Function	No. Of Neurons	MSE
1	TRAINLM	Learngdm	10	0.3798
2	TRAINRP	Learngdm	20	0.3881
3	TRAINLM	Learngd	10	0.3991
4	TRAINRP	Learngd	20	0.5046
5	TRAINLM	Learngdm	20	0.3743
6	TRAINRP	Learngdm	10	0.4492
7	TRAINLM	Learngd	20	**0.3612**
8	TRAINRP	Learngd	10	0.3862

Case 7 with "TRAINLM" as training Function and "Learngd" as adaptive learning function with 20 numbers of neurons has least MSE. So, case 7 is the premier case of LRN of this model when implemented in NNTOOL. Performance graph for best case of Layer Recurrent Network is shown in graph 6.3 and the best performance is observed at epoch 9.

Best Training and Testing Case of **CBP**:
Training Function : TRAINLM
Adaptive Learning Function : Learngdm
No. Of Neurons : 20
MSE : 0.3653
The performance can be plotted as:



Graph 6.4 Performance graph of Cascaded Back-Propagation for case 5

Cascaded Back-Propagation Algorithm was implemented using NNTOOL with input and Target data using two different training functions (TRAINLM & TRAINRP) and two different adaptive learning functions (Learngdm & Learngd) with 10 & 20 numbers of neurons. The MSE obtained in each case are shown in Table 6.3.

TABLE 6.3 CASCADED BACK-PROPAGATION TESTING CASES

Case	Training Function	Adaptive Learning Function	No. Of Neurons	MSE
1	TRAINLM	Learngdm	10	0.3882
2	TRAINRP	Learngdm	20	0.4380
3	TRAINLM	Learngd	10	0.3931
4	TRAINRP	Learngd	20	0.4031
5	TRAINLM	Learngdm	20	**0.3653**
6	TRAINRP	Learngdm	10	0.5125
7	TRAINLM	Learngd	20	0.3965
8	TRAINRP	Learngd	10	0.3896

Case 5 with "TRAINLM" as training Function and "Learngdm" as adaptive learning function with 20 numbers of neurons has least MSE. So, case 5 is the premier case of CBP of this model when implemented in NNTOOL. Performance graph for best case of Cascaded Back-Propagation Algorithm is shown in graph 6.4 and the best performance is observed at epoch 6.

VII. CONCLUSION

From this research work, it has found that when the algorithm is implemented using NNTOOL gave the best results as compared to NFTOOL, i.e., Multilayer algorithm is better than a single layer algorithm in terms of performance. It is noticed that "TRAINLM" is the best training function as it shows results with the least error. It is observed that "Learngdm" is the premier adaptive learning

function as it also shows results with the least error. It is observed that the network, when constructed with "Learngd" as an adaptive learning function, takes a bit little more time to process and train the system. When the number of neurons is increased from 10 to 20, the results gets enhanced and we came to a conclusion that n increasing number of the neuron the error in that process decreases and it gives the best result. During data manipulation, it is observed that if the data is not normalized, then it provides a product with maximum error, and accuracy in such cases is beyond expectations; hence Normalization is an important step to be followed before fitting it to the network.

- Different networks with the same training functions and adaptive learning functions having the same number of neurons give different results.
- The value of MSE decreases as the number of neurons increases in an ANN.
- Backpropagation shows the best result among these three networks.
- Learngdm is observed to be the best adaptive learning function as it shows results with the least MSE.
- Learngd is a bit time-consuming.
- The best training function observed is TRAINLM.
- A multilayer algorithm is better than a Single-layer algorithm in terms of performance.
- NNTOOL is observed to be the best for training the data as it gives an option to execute the algorithm rather than BPA.
- It is necessary to normalize high order Input and Output data for the best results.

References

[1] M. J.-C. Hu, "Application of the adaline system to weather forecasting," Department of Electrical Engineering, Stanford University, 1964.

[2] K. Abhishek, A. Kumar, R. Ranjan, and S. Kumar, "A rainfall prediction model using artificial neural network," in *2012 IEEE Control and System Graduate Research Colloquium*, 2012, pp. 82-87.

[3] A. Abraham and B. Nath, "A neuro-fuzzy approach for modelling electricity demand in Victoria," *Applied Soft Computing*, vol. 1, pp. 127-138, 2001.

[4] S. Chattopadhyay and G. Chattopadhyay, "Identification of the best hidden layer size for three-layered neural net in predicting monsoon rainfall in India," *Journal of Hydroinformatics*, vol. 10, pp. 181-188, 2008.

[5] N. Q. Hung, M. S. Babel, S. Weesakul, and N. Tripathi, "An artificial neural network model for rainfall forecasting in Bangkok, Thailand," *Hydrology & Earth System Sciences*, vol. 13, 2009.

[6] "Groundwater Solute Transport Modeling:Effects of Transport and Fate," *International Journal of Recent Technology and Engineering*, vol. 8, pp. 3541-3544, 2019.

[7] "<A9178058119.pdf>."

[8] D. R. Nayak, A. Mahapatra, and P. Mishra, "A survey on rainfall prediction using artificial neural network," *International Journal of Computer Applications*, vol. 72, 2013.