Forecasting Rainfall of Monsoon Season with Artificial Neural Network

Noor-E-Zannat, Raju Mondal and Munnujahan Ara
Mathematics Discipline
Science Engineering and Technology School
Khulna University, Khulna-9208
Bangladesh.
zannatnoor143@gmail.com, rajumondal1929@gmail.com, munnujahan@math.ku.ac.bd

Abstract

Now a day, worldwide meteorology is one of the most concerning issues. Among the meteorological parameters, rainfall is one of the most relevant characteristics of the climate of Bangladesh. About 80% rainfalls occurs here during monsoon season (June - October). Rainfall is a bless to many people like farmers but, sometimes heavy rainfall causes a great damage to many people specially who lives in the Southern part of Bangladesh. In urban areas, it has also a great impact on gutter system, traffic and many other activities and also causes hamper to people who travel from one place to another by plane. To predict the state of the atmosphere, rainfall prediction is one of the most important applications of science and technology. It is significant to determine the rainfall for effective uses of water resources, crop productivity, flood and also for the benefits of the airports to fly the plane. We consider Jashore as our study area of Bangladesh. Hence, we will forecast the rainfall of monsoon season of this area by using Artificial Neural Network (ANN) model and finally compare results with Multiple Linear Regressions (MLR) model to see which model gives the most relevant results.

Keywords: Artificial Neural Network, Rainfall, Monsoon Season, Multiple Linear Regressions.

1. Introduction

Weather forecasting is the dispensation of scientific knowledge for practical purpose to predict the condition of the atmosphere for a given location and time. Rainfall is one of them which is very important and complicated weather phenomenon. It has many impacts on both urban and rural areas such as in the urban areas, it has impact on gutter system, traffic and many other activities (Hung et al, 2009) and also has impact on agricultural activities in rural areas of many country like Bangladesh. Bangladesh is a land of natural calamities. Every year it faces a lot of natural disaster like storm, thunderstorm, flood, earthquake etc. Sometimes flood occurs because of heavy rainfall. We have so many important works in our everyday life. To reduce the consumption of time, sometimes people who have much money travel from one place to another by plane. Because of bad weather like heavy rainfall, sometimes these flights are delayed. Also, in the water journey, these bad weather causes much sufferings to a lot of people. Heavy rainfall occurs in the monsoon season most of the time in Bangladesh. So, it is necessary to all kind of people to know the more accurate rainfall value.

Bangladesh belongs to the Asian monsoon system. The monsoon season in Bangladesh stays from the month June to mid-October (Ahmed & Kim, 2003). About 80% rainfall occurs during the monsoon season. That's why we will discuss only the monsoon season into this paper. To forecast rainfall, Artificial Neural Network (ANN) Model is a very helpful model. In recent, many researches are doing work on ANN model. So, we will use this model in our paper. Since, ANN model is one kind of regression model, that's why we will compare it with Multiple Linear Regressions Model.

1.1 Objectives

Many researchers have done their work on rainfall and also forecast it by using different models like ANN model, MLR model etc. But, for Bangladesh since the mostly rainfall is seen on the monsoon season and on the southern part of Bangladesh. So, we will find out the rainfall of our one selected area on the monsoon season and will compare the two models like ANN model, MLR model. So, the objectives of our study are given below:

1) to find weather variables which has effects on rainfall.
2) to find the forecasted value of rainfall during monsoon season by ANN model of 1 station: Jashore.
3) to compare the forecasted rainfall of ANN and MLR model.

2. Literature Review

A monsoon is traditionally a seasonal reversing wind which corresponds to change in precipitation (Ramage, 1971), but now this term is used to describe the seasonal changes of atmosphere and also precipitation which is associated with asymmetric heating of land and sea (Zuidema et al, 2007). Monsoon season continues from June to October accompanied by rain in Bangladesh. It has different impacts on the local weather of different places. The Bay of Bengal branch of southwest summer monsoon flows over the Bay of Bengal, picking up more moisture from it, heading towards North-East India and Bangel and also it hits the coastal district Cox's Bazar of Bangladesh. Hence, the Cox's Bazar is the first district to receive rain from the southwest summer monsoon. Then, it passes across Bangladesh and hits the Eastern Himalayas and in the northeast India, Bangladesh and West Bengal; it causes a heavy rainfall.

For weather forecasting and climatology, we need some statistical data which has been measured by the meteorological department. In this study, we have used some meteorological parameters in our model. They are:

a) Precipitation
b) Temperature
c) Wind Speed
d) Mean Sea-Level Pressure
e) Relative Humidity
f) Cloud Cover

3. Methodologies

In this study, we used Artificial Neural Network (ANN) and Multiple Linear Regressions (MLR). So, the description of these models is given below:

3.1. Artificial Neural Network (ANN) Model

The ANN is a simplified mathematical model of a natural network. It is a computational non-linear model that is able to perform task like classification, prediction, decision making, visualization and others by considering examples. ANN is the component of artificial intelligence that simulates the functioning of a human brain. It is a directed graph where a vertex corresponds to a neuron and an edge to a synapse.

An ANN is based on artificial neurons which is a collection of connected units or nodes and also model the neurons in a biological brain. Each connection, can transmit a signal to other neurons like the synapses of the biological brain. A signal is received by an artificial neuron, then processes it and can signal neurons connected to it, where the "signal" at a connection is a real number, and there is also some non-linear function of the sum of its inputs by which the output of each neuron is computed. The connections are called edges. Neurons and edges typically have a weight that adjusts as learning proceeds. The weight increases or decreases the strength of the signal at a connection. Neurons may have a threshold such that a signal is sent only if the aggregate signal crosses that threshold. Typically, neurons are aggregated into layers. Different layers may perform different transformations on their inputs. Signals travel from the first layer (the input layer), to the last layer (the output layer), possibly after traversing the layers multiple times. (Wikipedia).

Neural network have been applied to model many non-linear hydrologic processes such as rainfall-runoff (Hsu et al, 1995) (Shamseldin, 1997), stream flow (Zealand et al, 1999), groundwater management (Rogers & Dowla, 1994), water quality simulation (Maier & Dandy, 1999) and rainfall forecasting. (Hung et al, 2009). ANN model mainly has two parts: Training and Testing. In the training part we have to train the data and in the testing part, we have to test some data according to the training data. In training part, ANN adjust the values of the weights and the biases.

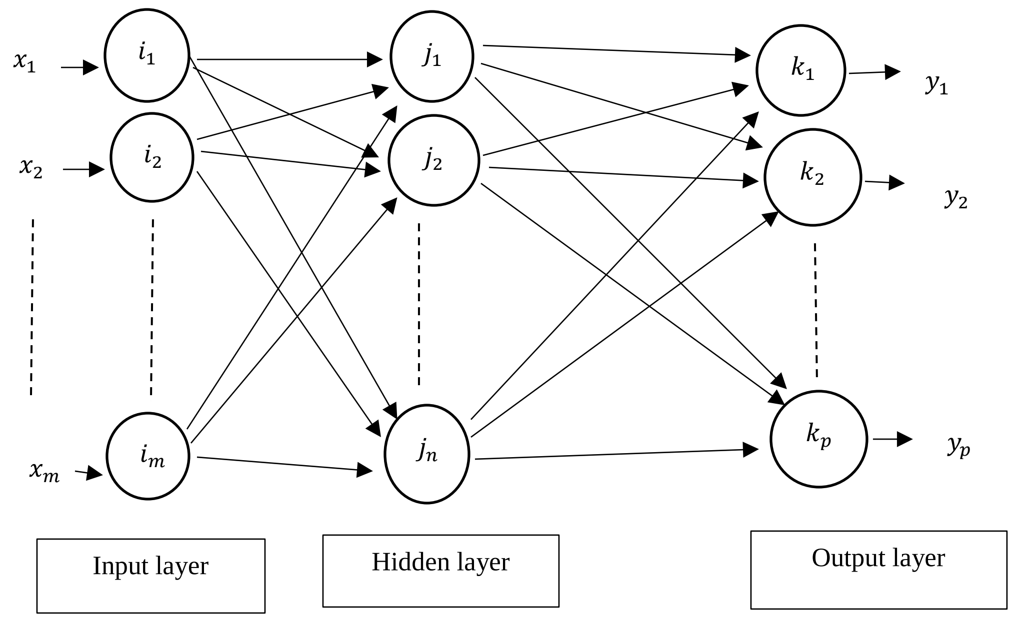


Figure 3.1: ANN Processing System

Let, $x_1, x_2, \ldots, x_m$ are input signals in the input nodes $i_1, i_2, \ldots, i_m$ respectively. These inputs are in the input layer. In the single hidden layer, $j_1, j_2, \ldots, j_n$ are the nodes of hidden layer. Now, w_{11} is the weight from i_1 to j_1, w_{12} is the weight from i_1 to $j_2, \ldots,$ w_{1n} is the weight from i_1 to j_n. Similarly, w_{21} is the weight from i_2 to j_1, w_{22} is the weight from i_2 to $j_2, \ldots,$ w_{2n} is the weight from i_2 to j_n and so on. $\Theta_1, \Theta_2, \ldots, \Theta_n$ are the biases of each nodes of the hidden layer. Then the weighted sums from input layer to each nodes of hidden layer are,

$$v_{ij} = \theta_j + \sum_{j=1}^{n} w_{ij} x_i$$

In each hidden layer, there remains an activation function through which the weighted sum passes. In most of the papers, sigmoid function is commonly used as an activation function. So, the input of each hidden nodes will be the weighted sum and the output of each hidden nodes will be:

$$h_j = \frac{1}{1 + e^{-v_{ij}}}$$

If in the output layer, $k_1, k_2, \ldots, k_p$ are the nodes of output layer and w_{j1} is the weight from j_1 to k_1, w_{j2} is the weight from j_1 to $k_2, \ldots,$ w_{jp} is the weight from j_1 to k_p and so on, and also . $\Theta_1, \Theta_2, \ldots, \Theta_p$ are the biases of each nodes of the output layer then the weighted sum from each nodes of hidden layer to each nodes of output layer is:

$$v_{jk} = \theta_k + \sum_{k=1}^{p} w_{jk} h_j$$

So, the input of each output nodes will be the weighted sum and the output of each output nodes will be:

$$y_k = \frac{1}{1 + e^{-v_{jk}}}$$

Weight and bias estimation are the main work of the training part of the network. Commonly, the initial value of the weights and the biases are considered as 0. Training has two principles:

1) Iterative error minimization: neuron-specific errors are used for updating weights.
2) Back-propagation: It computes errors from last layer to first layer.

The equations for weight estimations are respectively:

For output nodes, $w_{jk}^{new} = w_{jk}^{old} + lr * (error) * y_k$

$$error = y_k(1 - y_k) * (target\ output - y_k)$$

Where, y_k is the obtained output before training the network.

For hidden nodes, $w_{ij}^{new} = w_{ij}^{old} + lr * (error) * y_k$

$$error = y_k(1 - y_k) * \sum_i error_i\, w_{ij}^{old}$$

Once these equations are executed, then the weights are estimated by the following equation:

$$w_{jk}^{new} = w_{jk}^{old} + (1 - m) * lr * error * y_k + m(w_{jk}^{old} - w_{jk}^{older})$$

Where, lr is the learning rate which ranges from 0 to 1 and m is the momentum which ranges from 0 to 1.
This process will be continued until the weights are estimated with minimum error.
When the training is over, the model tests the system with some data to test whether it gives almost the same outputs which are observed in many systems. This is the system by which the model works.
How many data are tested or trained is not fixed. But most of the researchers divided the data set into two parts: for the number of data to be trained is, $N_{tr} = (\frac{1}{4} \div \frac{1}{3})N$, where N=total numbers of data. The number of data to be tested is, $N_{ts} = (N - N_{tr})$ (Madić et al, 2012).
The number of neurons in hidden layer should be less than twice the number of the input nodes (Rehman, 2018). This number can vary with the problem. So, first we have to consider our problem.

3.2 Multiple Linear Regression (MLR) Model

In real life, most of the problems doesn't depend only one variable. In that case, multiple linear regression is used. In multiple linear regression, there are more than one independent variable and one dependent variable. The general form of MLR is,

$$Y = a_0 + a_1X_1 + a_2X_2 + a_3X_3 + a_4X_4 + a_5X_5 + \cdots + a_nX_n$$

Here, the parameters are obtained by training data and the variables are extracted from the dataset.

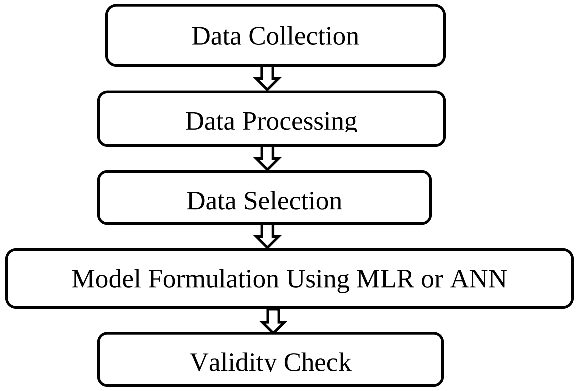
4. Data Collection

We collected daily data of the six mentioned parameters of monsoon season (June-October) of Jashore District from 1979 to 2018 from Bangladesh Meteorological Department, Dhaka, Bangladesh. Among these, the data of 1989, 1999 are missing.

5. Results and Discussion

5.1 Numerical Results

In this section, we have to follow some steps. They are given below with a chart:



The first most important thing is data collection, which we mentioned into the above and it maintains the dataset in the form of excel sheet on daily basis and we used data of 37 years as training and data of 1 year as testing in our study. The second step is data processing. In this step, we cleaned the noisy and irrelevant data. The third step is data selection. Here, we selected the data which are correlated to our analysis. Into the fourth and fifth steps, we formulated model and checked the validity of it respectively.

The multiple linear regression (MLR) model for the forecasted rainfall is,
$Y = |580.5091 - 3.3343X_1 + 0.1361X_2 + 41.2677X_3 - 520.757X_4 + 119.7838X_5|$…………(1)
Where,
Y =Predicted Rainfall
X_1 =Average Temperature
X_2 =Wind Speed
X_3 =Humidity
X_4 =Mean Sea Level Pressure
X_5 =Cloud Cover

We use on the above model modulus because some of the predicted values awe are getting are negative. But it is quite impossible for monsoon season because negative rainfall happens because of draught. So, by converting the negatives values into positive, we can get the most accurate predicted value than the previous.
For artificial neural network, we took different number of neurons into the hidden layer and found out that 7 neurons give more accurate results than the others. So, we used 7 neurons or nodes into the hidden layers.
The artificial neural network (ANN) model for the forecasted rainfall is,
$E_i = W_{1i}X_1 + W_{2i}X_2 + W_{3i}X_3 + W_{4i}X_4 + W_{5i}X_5 + W_{6i}$…………….(2)
Where, i=1, 2, 3, …, 7
The values of the weights of equation (2) is given into the following table 5.1.1.

Table 5.1.1: Weights of equation (2)

	W_{1i}	W_{2i}	W_{3i}	W_{4i}	W_{5i}	W_{6i}
1	1.046	1.046	-0.058	-0.831	-0.193	-1.101
2	-0.119	1.344	-0.394	-1.012	0.059	-1.353
3	0.465	2.575	0.342	-1.387	1.081	-1.482
4	-0.080	-0.331	-0.155	-0.534	-0.558	-0.171
5	-0.241	0.773	-0.450	-1.015	-0.190	-0.850
6	0.350	1.852	0.452	-1.020	0.911	-2.086
7	-0.159	0.801	-0.492	-0.772	0.231	-1.688

Now, F_i (i=1, 2, 3,…,7) can be calculated by sigmoid function by using the following equation:
$F_i = \frac{1}{1+e^{E_i}}$………………..(3)
Y=0.904-(1.208*F_1)-(1.518*F_2)-(3.026*F_3)+(0.405*F_4)-(1.066*F_5)-(2.555*F_6)-(1.399*F_7)…………(4)
By putting the values of our mentioned parameters into the above equations, we can get the predicted value of rainfall.

5.2 Graphical Results

Here, we have shown the graphical representation of the actual rainfall compared with the MLR predicted rainfall and also ANN predicted rainfall.

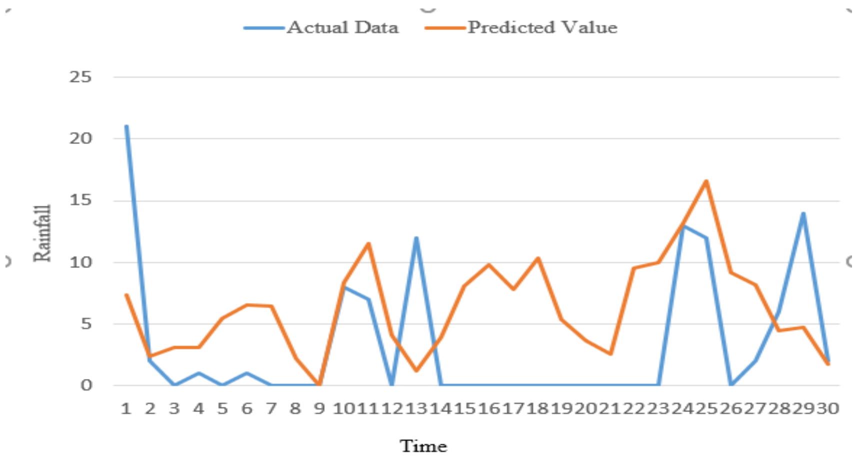


Figure 5.2.1: Comparison between actual data and MLR predicted value of rainfall (June, 2018)

In Figure 5.2.1, we showed how MLR predicted rainfall differs from actual rainfall at the month June.

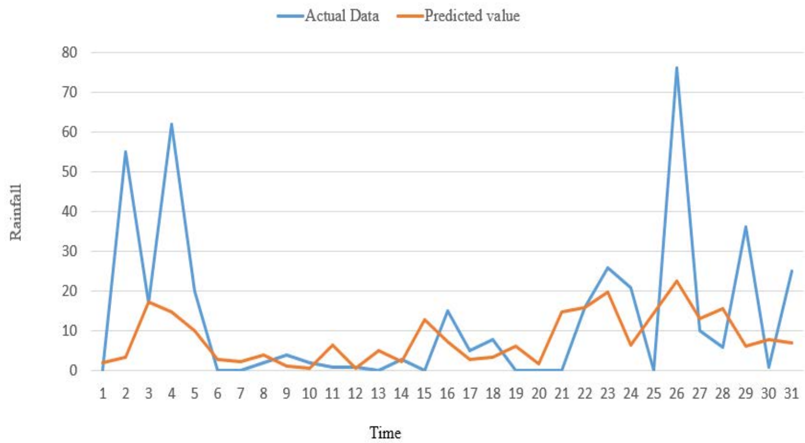


Figure 5.2.2: Comparison between actual data and MLR predicted value of rainfall (July, 2018)

In Figure 5.2.2, we showed how MLR predicted rainfall differs from actual rainfall at the month July.

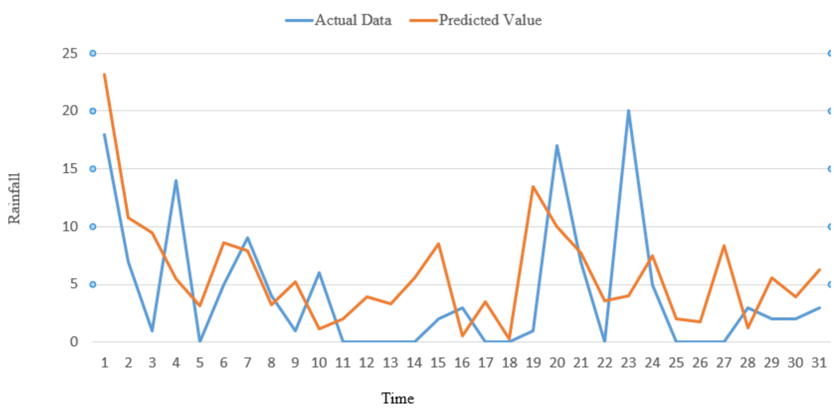


Figure 5.2.3: Comparison between actual data and MLR predicted value of rainfall (August, 2018)

In Figure 5.2.3, we showed how MLR predicted rainfall differs from actual rainfall at the month August.

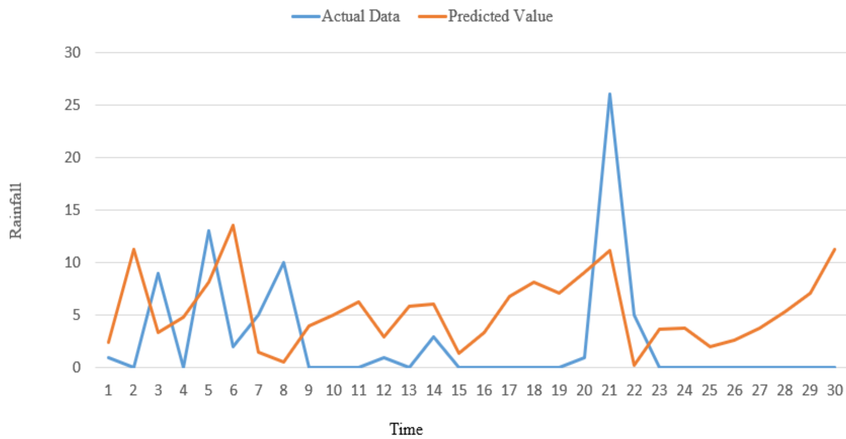


Figure 5.2.4: Comparison between actual data and MLR predicted value of rainfall (September, 2018)

In Figure 5.2.4, we showed how MLR predicted rainfall differs from actual rainfall at the month September.

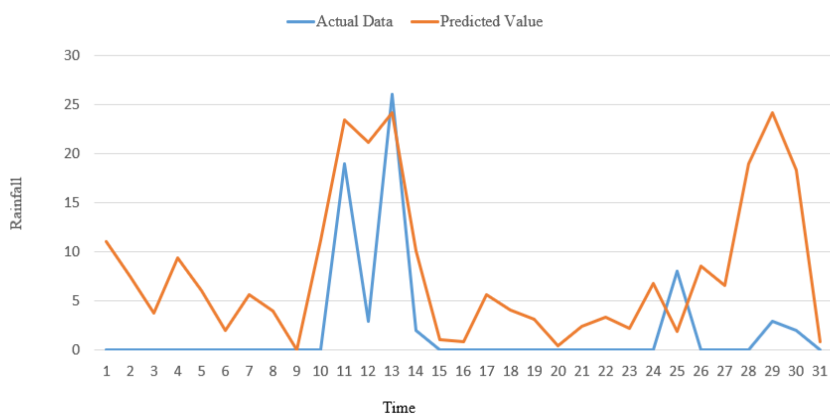


Figure 5.2.5: Comparison between actual data and MLR predicted value of rainfall (October, 2018)

In Figure 5.2.5, we showed how MLR predicted rainfall differs from actual rainfall at the month October.

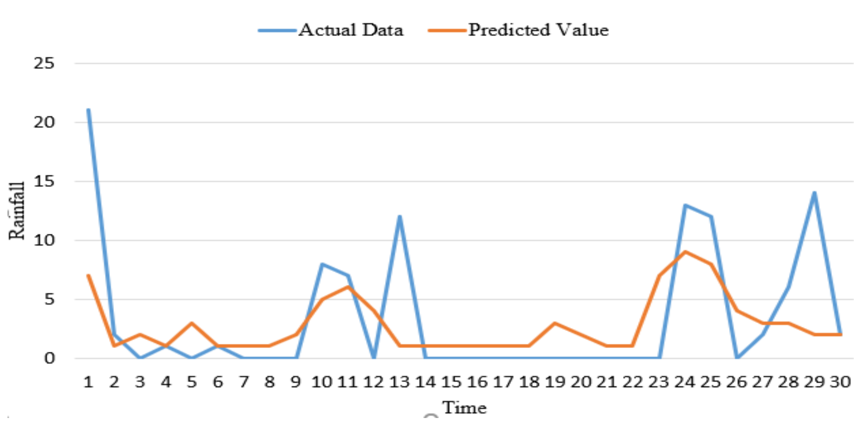


Figure 5.2.6: Comparison between actual data and ANN predicted value of rainfall (June, 2018)

In Figure 5.2.6, we showed how ANN predicted rainfall differs from actual rainfall at the month June.

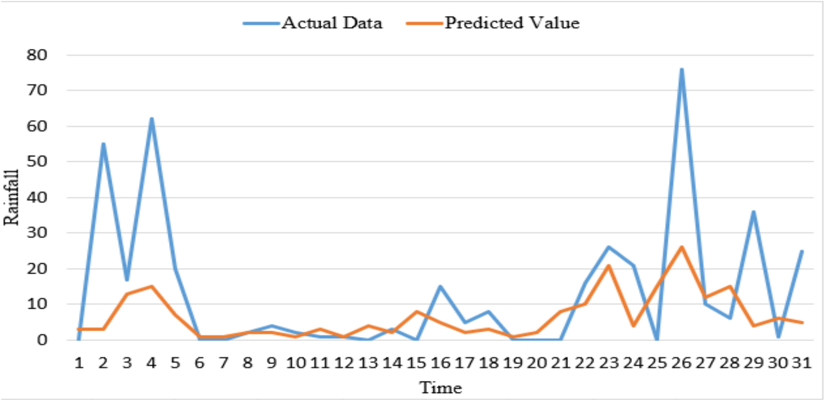


Figure 5.2.7: Comparison between actual data and ANN predicted value of rainfall (July, 2018)

In Figure 5.2.7, we showed how ANN predicted rainfall differs from actual rainfall at the month July.

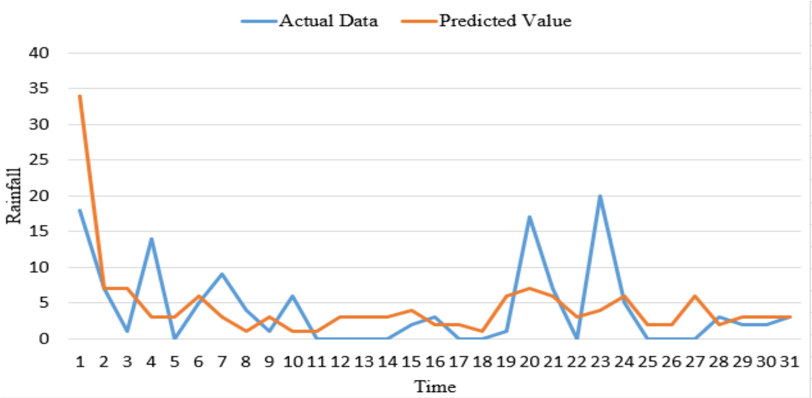


Figure 5.2.8: Comparison between actual data and ANN predicted value of rainfall (August, 2018)

In Figure 5.2.8, we showed how ANN predicted rainfall differs from actual rainfall at the month August.

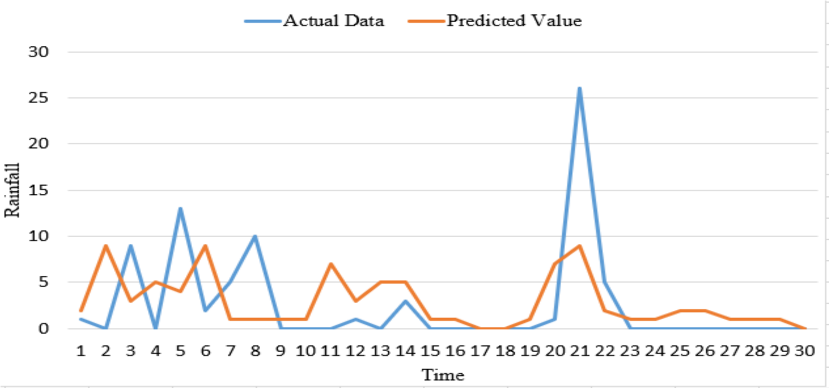


Figure 5.2.9: Comparison between actual data and ANN predicted value of rainfall (September, 2018)

In Figure 5.2.9, we showed how ANN predicted rainfall differs from actual rainfall at the month September.

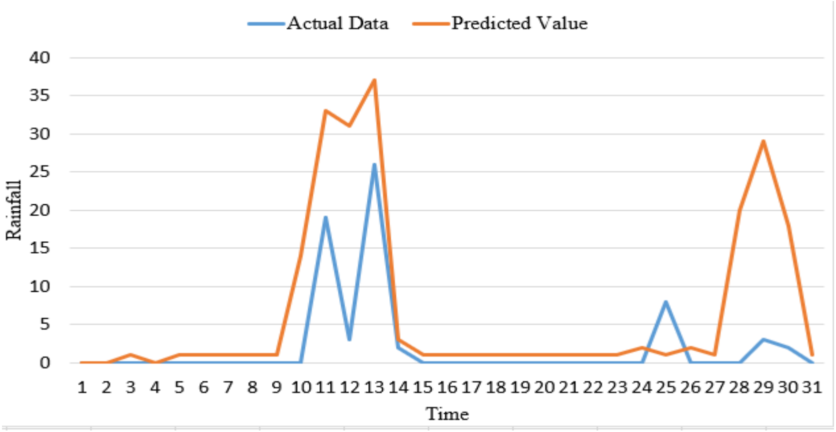


Figure 5.2.10: Comparison between actual data and ANN predicted value of rainfall (October, 2018)

In Figure 5.2.10, we showed how ANN predicted rainfall differs from actual rainfall at the month October.

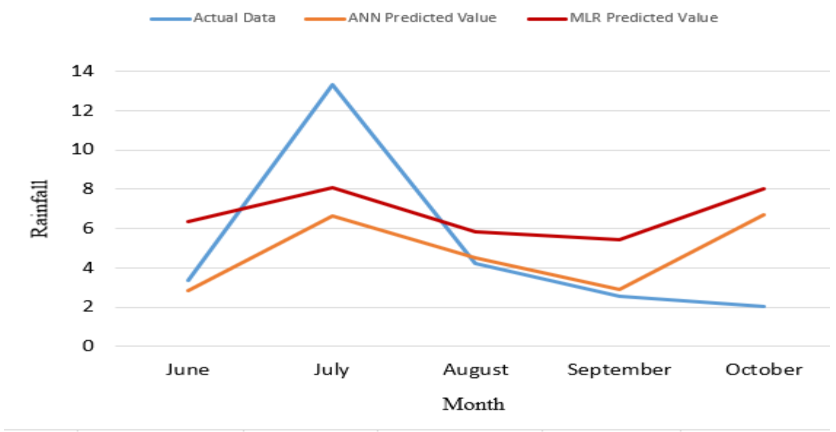


Figure 5.2.11: Comparison between actual and ANN predicted rainfall and MLR predicted rainfall of monsoon season

In Figure 5.2.11, we showed a comparison between the actual and ANN predicted rainfall and MLR predicted rainfall of monsoon season and found that ANN gave most accurate predicted value of rainfall than MLR.

5.3 Proposed Improvements

Here, we took the daily data that's why we couldn't make the comparison with the non-linear regression model because there were some 0 value into some data. But in the nearest future, we will take the average data of each month and will also compare the actual rainfall with non-linear regression model and other forecasted model also. We will also find out the average predicted rainfall of different regions and also show, where the rainfall is low and where it is high.

5.4 Validation

For the purpose of comparison analysis between the multiple linear regression model and ANN model, we found out the error between the actual value and the predicted value. The root mean square error (RMSE) is one of the best error estimation methods. The formula for RMSE is,

$$RMSE = \sqrt{\frac{\sum(actual\ calue - predicted\ value)^2}{n}}$$

By finding the RMSE error, we found that, in our study ANN has error 9.27 where, MLR has error 10.279, which told that ANN gives less error than MLR model in Rainfall forecasting.

In addition, the coefficient of correlation between the target value and output value is defined into table 5.4.1. From the above two graphs and correlation coefficients, it is clear that ANN can forecast more accurate predicted values than MLR. Correlation Coefficients of each parameters with the predicted value is given into the table 5.4.2.

Table 5.4.1: Correlation coefficient between the actual rainfall, MLR predicted rainfall and ANN predicted rainfall

	MLR Predicted Rainfall	ANN Predicted Rainfall
Actual Rainfall	0.386	0.444

Table 5.4.2: Correlation coefficient between the actual rainfall and ANN predicted rainfall

SI. No	ANN Predicted Rainfall
Average Temperature	-0.57496
Wind Speed	0.044189
Humidity	0.705477
Mean Sea Level Pressure	-0.063861
Cloud Cover	0.626201

6. Conclusion

The objective of this paper is to develop a model that can be used to predict daily rainfall by using various daily meteorological variables in Jashore District, Bangladesh. For this purpose, methods of MLR and ANN were used. The results obtained with these models were compared to each other and with the measured data and found out that ANN model gives the most accurate result. The advantage of this model is that having the required various daily meteorological variables such as daily average temperature, wind speed, humidity, mean sea level pressure and cloud cover can be predicted quickly and satisfactorily without the use of any other parameters related to rainfall.

References

Ahmed, R., & Kim, I.-K. Patterns of daily rainfall in Bangladesh during the summer monsoon season:case studies at three stations. *Physical Geography*, 295-318,2003.

Amare, M., Jensen, N. D., Shiferaw, B., & Cissé, J. D. Rainfall shocks and agricultural productivity: Implication for rural household consumption. *Agricultural Systems*, 79-89, 2018.

Hsu, K., Gupta, H., & Sorooshian, S., Artificial neural network modeling of the rainfall-runoff process. *Water Resour*, 2517-2530,1995.

Hung, N. Q., Babel, M. S., Weesakul, S., & Tripathi, N. K., An artificial neural network model for rainfall forecasting in Banhkok, Thailand. *Hydrology and Earth System Sciences*, 1413-1414,2009.

Madić, M., Marinković, V., & Radovanović, M., Mathematical modeling and optimization of surface roughness in turning of polyamide based on artificial neural network. *ISSN 1392 - 1207. Mechanika.*, col 18(5): 574-581,2012.

Maier, R., & Dandy, G. Comparison of various methods for training feed-forward neural network for salinity forecasting. *Water Resour*, 2591-2596,1999.

Ramage, C.,Monsoon. Available: https://en.wikipedia.org/wiki/Monsoon,1971.

Rehman,H.,Available:https://www.researchgate.net/post/How_to_decide_the_number_of_hidden_layers_and_nodes_in_a_hidden_layer/5a5db4ee48954c2610357754/citation/download,2018.

Rogers, L., & Dowla, F., Optimization of groundwater remediation using artificial neural networks with parallel solute transport modeling. *Water Resour*, 457-481,1994.

Shamseldin, A., Application of a neural network technique to rainfall-runoff modeling. *J. Hydrol.*, 272-294,1997.

Artificial Nueral Network, Available: https://en.wikipedia.org/wiki/Artificial_neural_network

Zealand, C., Burn, D., & Simonovic, S., Short term streamflow forecasting using artificial neural networks. *J. Hydrol*, 32-48,1999.

Zuidema,, P., Fairall, , C., Hartten,, L., Hare., J., & Wolfe, D., On air-sea interaction at the mouth of the gulf of California. *Journal of Climate*, 1649-1661,1 May, 2007.

Biographies

Noor-E-Zannat and Raju Mondal is currently a B.Sc. final year student at Mathematics Discipline under Science Engineering and Technology School, Khulna University, Bangladesh. They are doing their research work on meteorology field.

Munnujahan Ara is a Professor in the Mathematics Discipline at Khulna University, Khulna, Bangladesh. She received Ph.D. in telecommunication from university of Porto, Portugal under MAP-TELE (joint work with University of Minho, University of Porto and University of Aveiro, Portugal) programme, M.Sc. in Applied Mathematics and B.Sc. (Honors) in Mathematics from Khulna University, Bangladesh. She has published journal papers in national and international journals and attended on national and international conferences and seminars. Dr. Ara has completed different projects with Portuguese national funding agency for science, research and technology (FCT). She is a member of International Association of Engineers and also life member of Mathematics Forum, Khulna, Bangladesh, Bangladesh Physical Society, Bangladesh and Bangladesh Mathematical Society, Bangladesh.