Environment and Ecology Research 11(6): 963-972, 2023 http://www.hrpub.org
DOI: 10.13189/eer.2023.110607

Rainwater Quality Assessment Based on Artificial Neural Network Using Mathematical Models

Zuhal Abdul Hadi Hamza[1], Ahmed Makki Al-Sulaiman[2,*]

[1]Department of Civil Engineering, University of Basrah, Iraq
[2]Department of Civil Engineering, University of Qadisiyah, Iraq

Received August 5, 2023; Revised October 24, 2023; Accepted November 19, 2023

Cite This Paper in the Following Citation Styles

(a): *[1] Zuhal Abdul Hadi Hamza, Ahmed Makki Al-Sulaiman , "Rainwater Quality Assessment Based on Artificial Neural Network Using Mathematical Models," Environment and Ecology Research, Vol. 11, No. 6, pp. 963 - 972, 2023. DOI: 10.13189/eer.2023.110607.*

(b): *Zuhal Abdul Hadi Hamza, Ahmed Makki Al-Sulaiman (2023). Rainwater Quality Assessment Based on Artificial Neural Network Using Mathematical Models. Environment and Ecology Research, 11(6), 963 - 972. DOI: 10.13189/eer.2023.110607.*

Abstract One of the most crucial and difficult duties jobs performed by meteorological agencies in the world is the forecasting of weather, particularly rainfall. Furthermore, it is a difficult process that calls for knowledge from many different specialist domains. In this study, a model based on an artificial neural network (ARNN) is suggested as a method to forecast successive rainfalls according to Meteorological Station based on analyses of previous rainfall data. It is for 11 years from 2010-2021 at the Hay Al Hussian in Basrah. Based on three informative meteorological factors, the feed forward neural networks with back propagation algorithm are used for learning and predicting. The created models have been trained, validated, and tested using observations of temperature, wind speed, and relative humidity. The research discovered that the neural network ARNN organized (3-30-1) was capable of long-term rainfall forecasting at the study region with one hidden layer and layers (3-25-5-1) were capable of long-term rainfall forecasting in this region with two hidden layers. The model was able to learn the events that it had been trained to recognize, according to the results, which were backed up by strong correlation coefficient (R) values and low mean squared errors (MSE) values. The root MSE was 0.0016187, and R Value was found to be 0.997 for one hidden layer and the root MSE was 8.4731E-8, and R value was found to be 1 for two hidden layers.

Keywords Rainfall, Artificial Neural Network (ANN), Water Quality Assessment

1. Introduction

Water is the major survival of all living things on earth, including humans, animals, and plants. The hydrological cycle involves a number of intricate processes, each of which is crucial for the management and planning of water resource projects. In addition to managing the water supplies, the process of rainfall is essential to the existence of plants and animals. The majority of the fresh water on earth is dispersed by rainfall. In addition to providing water for hydroelectric power facilities and farmland irrigation, it also offers favorable circumstances for a variety of ecosystems.

Solar radiation, temperature, humidity, vapor pressure deficit, atmospheric pressure, wind, and relative humidity are the most frequent and significant elements that affect rainfall. A better improved prediction model is necessary for an early detection that can reduce risks to life individuals and property and to better manage water resources. Forecasting rainfall is crucial since low and erratic rainfall may have various implications on crops and farms as well as damage to property [1]. With the aid of accurate rainfall prediction, a significant portion of the rainwater that is currently not being used should be made use of for a variety of purposes, which may lead to proper irrigation, sufficient hydropower generation, as well as help in minimizing the losses caused by drought and flood.

Inadequate rainfall has a severe negative impact on the aquatic ecology, water supply, and water quality. Effective water use would be achievable if rainfall could be predicted several months in advance, if possible. A difficult task in practical water resources management is projecting rainfall accurately.

Models for predicting rainfall are divided into two types: physical models and models based on data-driven. Physical models take into account the all-significant laws and pertinent physical processes that have an impact on the process of rainfall. Models based on data-driven employ data from the past, in order to forecast the future. Multiple linear regression (MLR), auto-regressive integrated moving average (ARIMA), support vector machine (SVM), and artificial neural networks (ARNN) are among the most well-known and often used data-driven models that were taken into consideration for the prediction of rainfall. Many data-driven models were taken into consideration for the rainfall prediction [2, 3, 4, 5, and 6].

In many ways, the process of precipitation is quite important. For studies like the balance of water, design of irrigation systems, and management of water resources, these are the fundamental elements of the hydrologic cycle. Predicting rainfall has emerged as one of the greatest and most difficult technological and scientific challenges in the modern era. Rainfall and climate are typically very complex, not linear events that require sophisticated computerized and simulation modeling to accurately anticipate. Such non-linear systems can be predicted using an artificial neural network (ARNN).

ARNNs have been shown to be very effective tools in dealing with a variety of extremely complex situations, due to their capacity to simulate non-linear systems without the need for any assumptions.

The emergence of ARNN, which occupied the attention of researchers because of its ability to deal with complex non- linear problems better than other statistical methods, especially the ARNN technique that relies on the backpropagation algorithm, it was possible to successfully identify the internal dynamics of the rainy times series, and predict them for the longest period, because it could be done a month or a year ago, and there is a surplus of evidence in the literature on the superiority of using ARNN over other statistical methods, especially in the field of rain forecasting.

ARNN can be utilized for prediction because they are capable of studying and identifying the trend from a large historical set of data. Compared to statistical and mathematical models, ARNN is more accurate [7]. ARNN, a form approach based on data built, is based on the idea of biological neurons. ARNN can be used to examine the relationship between weather-related variables and rainfall ARNN can be used to examine the relationship between meteorological factors and rainfall. An adaptive system, or ARNN, can change its structure in response to input coming from the outside or from within the network as it goes through its learning phase.

The emergence of ARNN occupied the interest of researchers because of its ability to deal with complex nonlinear problems better than other statistical methods, especially the ARNN technique. It is that relies on the backpropagation algorithm, which was possible to successfully determine the internal dynamics from the series of rainy times, as well as predict them for the longest period. Since it can be performed a month or a year in advance, there is an excess of evidence in the literature of the superiority of using ARNN in the field of rain forecasting over other statistical methods [8, 9, 10].

The neurons of a neural network are interconnected, and when one neuron produces an output, others take that input to produce the final output. When given an example of a collection of input data with predetermined results or output, the network learns. The factors that determine their weight are changed (either manually or automatically by a computer algorithm), and these weights for connection store the information needed to make the final outcome closer to the predetermined result [11].

An (ARNN) is a type of data processing system that has an architecture similar to the cerebral cortex of the brain and is composed of numerous simple, highly interconnected processing components. An ANN has three layers: input, hidden, and output, each of which has a variety of artificial neurons. To create neural networks with various computing capabilities, a wide range of network designs and learning techniques can be combined. Because they can accommodate significant parameter fluctuations, ANNs in hydrology are more robust than any computational approaches or modeling.

MATLAB software (version 9.70.1190202 (R2019)) is used to model and evaluate ANN. Math Works created the multi-paradigm programming language and proprietary environment known as MATLAB. Matrix manipulation, visualizations of functions and data, implementing algorithms, interface for user construction, and communication with other programming languages are all possible with MATLAB.

Perform data analysis, algorithm development, and model and application development using MATLAB. Control systems, signal processing, test and measurement, computational finance, and computational biology are examples of MATLAB applications.

So, the system was created to organize huge amounts of data into matrices and treat this data as individual arrays rather than collections of data. With this it is relatively easy to perform complex operations on the data and very complicated to get it incorrect.

Figure 1. The location of the study area

2. Study Area

Basrah Governorate is located in the south of Iraq. Monthly rainfall data were collected from Hay Al-Hussien District Meteorological Station located at latitude 53.13' 29 30" N and longitude 33.92' 46^0 47 E, and at an altitude of 2.4 m above sea level.

Due to the importance of water and the increasing need for it, especially at the present time as a result of the increase in population and the rise in global temperature, which leads to an increase in the amount of evaporation from water bodies, the issue of accurate rain forecasting has become the focus of attention of researchers for many decades. The aim is to obtain a future outlook and achieve an integrated and effective water resources management.

The primary objectives of this study are to collect and analyze climatic data related to precipitation. Determine this in creating an artificial neural network (ARNN) model that predicts precipitation. Figure 1 shows the location of the study area.

3. Methodology and Mathematical Model

3.1. Artificial Neural Network (ANN)

A computer system's artificial neural network (ARNN) is a part designed to resemble how the human brain processes and evaluates input. It serves as the cornerstone of artificial intelligence (AI) and searches for solutions to issues that are inaccessible or challenging by human or statistical criteria. As more data become available, the self-learning capabilities of ARNNs allow them to provide better results [11].

The nodes joined by connections create an ARNN. A layer is formed by grouping several nodes with comparable properties. A layer can be thought of as a set of nodes that are related to other layers or to the surrounding environment but not to each other.

The basic building block of a neural network is the artificial neuron, which is commonly referred to as a "perceptron". It is a mathematical procedure based on a simplified model of biological neurons. It can also be considered as an immediate binary logical gate. The basic purposes of each artificial neuron are:

1. Using the input layer, neurons and nodes take the input values.
2. Weigh each one individually and multiply it with delivery weights.
3. To create an output, the products are added, added, and added with a bias value before being provided by the activation functions (transfer function).

The ability of ARNN to model the relationship between a set of input and output variables plays an important role in ANN. Since ARNN is a nonlinear modeling tool that does not need an explicit mathematical formula for the relationship between input and output variables, as well as, between various types of artificial neural networks, especially the backpropagation algorithm, it has received a lot of attention in the field of weather forecasting [12].

There are three components of ARNN, through which ARNN is distinguished from the others, namely the form of interconnection between neurons, the method that determines the weights of these interconnections which are called training, learning and algorithms, in addition to the type of activation function used [13].

What distinguishes ARNN most is its ability to distinguish and generalize, in order to reach this benefit. The data during the calibration period are divided into three groups through training, validation and testing to the training set to reach a network with the lowest level of error, firstly, the validation set, which is used to improve the performance of the neural network on previously unrecognized input samples during the training phase, secondly. The test set is used for the purpose of ensuring the compatibility of performance to train the network, thirdly.

These operations (verification, testing) aim to reach what guarantees the performance of the network, to prevent excessive training of the proposed network, and to determine the weights that link the network inputs to its outputs in the best way.

Whereas, the ARNN architecture consists of three layers: the input layer that contains the aggregated data, the output layer that generates the computed information, and one or more hidden layers that connect the input and output levels as shown in Figure 2.

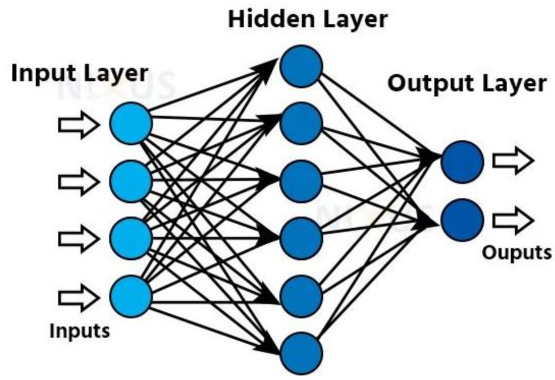

Figure 2. Artificial neural net works

3.2. Regression Analysis

Linear regression analysis of previous rainfall performed on data. It is for 11 years from 2010-2021, as shown in Table 1. set collected for the purpose of building and evaluating ANN models, in addition to predicting precipitation amount.

Table 1. The average monthly rainfall in this study

	Rain	Rain	Rain	Rain	Rain	Rain	Rain	Rain	Rain	Rain	Rain	Rain
Average monthly readings period (2010-2021)	13.4	59.7	0	7.8	3	0	0.1	0	0	54.6	0	58.6
	4.7	7.2	0	5.8	22.4	0	0	12.2	0	31.2	3.8	45.5
	43.8	0	26.5	2	10	0	0	1.6	0.6	8	13.4	0.6
	10.8	0.2	31.6	0	7.4	4.5	0	46.7	50.9	7.2	0	9.4
	6.9	0	0.3	0	7.2	11.9	0	16.3	28.9	0	4.4	0
	13	0	0	0	0	64.5	0.9	41.4	2.2	0	7.5	0
	10.3	0	18	0	18.6	21.5	22.6	5.8	9.1	0	10.1	0
	16.7	0	0	0	0	16.4	39.6	46.7	45.3	0	4.6	0
	54.7	0	0	0	7	2.8	10.1	2.5	0	51.6	0	48.6
	26.1	2.9	0	9.7	0.5	22	2.3	0	13	17.3	0	10.3
	0	6.4	0	15.5	0	0	3	0	0	8.8	0	3.2
Average	18.22	6.95	6.945	3.71	6.918	13.05	7.145	15.75	13.636	16.245	3.982	16.02
Standard Deviation	16.13	16.9	11.66	5.06	7.299	18.27	12.19	18.6	18.315	19.556	4.456	21.84

4. Data Preprocessing

- The division of data

The data for this study are split into three sections: training, testing, and validation. The model training, testing, and validation processes are the three fundamental steps that happen one at a time in the model creation process (15% of the data were utilized for testing, 15% for validation, and 70% of the data were used for training). Out of a total of 156 meteorological data points, 109 were used for training and 23 for testing and validation in order to predict rainfall.

- Data normalization

If specific pre-processing processes, such as data normalization, are carried out on the network input and target, neural network training can be carried out more effectively. Data normalization adjusts the input and the target so that they fall within a predetermined range. This procedure eliminates the data's cyclicity and aids in the network's rapid training. The data samples were adjusted between 0 and 1 using the equation, to increase the neural networks' training effectiveness.

$$x^n = \frac{x - x_m}{x_{ma} - x_m}$$

Where, x^n= Value of normalization
x = original data value
x_m = The minimum amount of data in total that will be normalized
x_{ma} = The maximum amount of data in total that will be normalized.

- Evaluation of ANN model performance

Evaluation of model performance was done on the basis of quantitative assessments; The model's quantitative prediction performance was assessed using the Mean Square Error (MSE) and correlation coefficient (R).

$$MSE = \frac{1}{M}\sum_{i=1}^{M}(R_{ob} - R_p)^2$$

Where, R_{ob}=observed data
R_p=predicted data

$$R = \frac{\sum(R_{ob} - R_{\tilde{ob}})(R_p - R_{\tilde{p}})}{\sqrt{\sum(R_{ob} - R_{\tilde{ob}})^2} + \sqrt{\sum(R_p - R_{\tilde{p}})^2}}$$

Where, $R_{\tilde{ob}}$ = mean observed data
$R_{\tilde{p}}$ = mean predicted data

5. Results and Discussion

The results of the ANN model are presented. Temperature, wind speed, relative humidity, and other meteorological variables were used as inputs, and the output of the forecast model was precipitation. The models used to predict when it will rain as mentioned earlier, use 70% of the data for training, 15% for testing, and 15% for validation. The mean squared error and correlation coefficient value were used to evaluate the effectiveness of the generated model.

5.1. Model Architecture for ANN

There are many numbers of hidden layers, hidden neurons, transfer functions, etc. in an ANN model architect. Even though a general structure can be followed based on previously successful applications, there are no set guidelines for creating an ANN model. Levenberg-Marquardt's learning rule and the Feed Forward Back Propagation (FFBP) model architecture were employed in this work.

The biggest worry while resolving actual problems based on Multilayer feed forward networks has been the number of hidden layers in the architecture. The trial and error method was the most accurate and straight forward approach of choosing the ideal number of concealed layers. For the purpose of choosing the best ANN model, five models were built and compared.

5.2. Distribution of Nodes and Layers

There may be one or many hidden layers depending on the nonlinearity of the problem. So, the decision about the number of hidden units involves a trade-off between the model's required flexibility and the length of time required to train the network. It is typical to start with a minimal number of hidden units and gradually expand it until the network's approximation quality is adequate. Unfortunately, every time the network's structure is changed, the network must be completely retrained. The number of nodes in a hidden layer or layers has a significant impact on the network training results.

The number of hidden layers and nodes in each hidden layer has a common effect on over-fittings, predictions, and outputs of neural networks. In order to select the right number of hidden layers and nodes for each hidden layer, a trial-and-error method is used.

The following guidelines are used to determine how many nodes should be included in the hidden layer [14]:

1. For both training patterns and testing patterns, the output network parameters' maximum error should be as small as possible.
2. There should be as few training epochs as possible.

In this study, the network is tested with configurations of one and two hidden layers, each with an increasing number of nodes (s). The best topology is first established using a single hidden layer with an activation function that is a linear (purelin) function in the output layer and a hyperbolic tangent (tansig) function in the hidden layer.

The performance of various topologies for both training and testing is illustrated in Fig. 3. When different numbers of nodes (5,10,15,20,25,30,35,40,45) are explored. According to this graph, the network with 30 nodes in the hidden layer performs better than others during training (MSE=0.0016723).

Then, two hidden layers are utilized, with the first hidden layer's activation function being a hyperbolic tangent function (tansig). The second hidden layer and output layer's activation functions are linear (purelin) functions.

There are five nodes in each buried layer, varying in number. Fig 4. displays how well different network topologies performed throughout training. The network that has first and second hidden layers with 25 nodes each performs the best for both training and testing (MSE=8.47e-8).

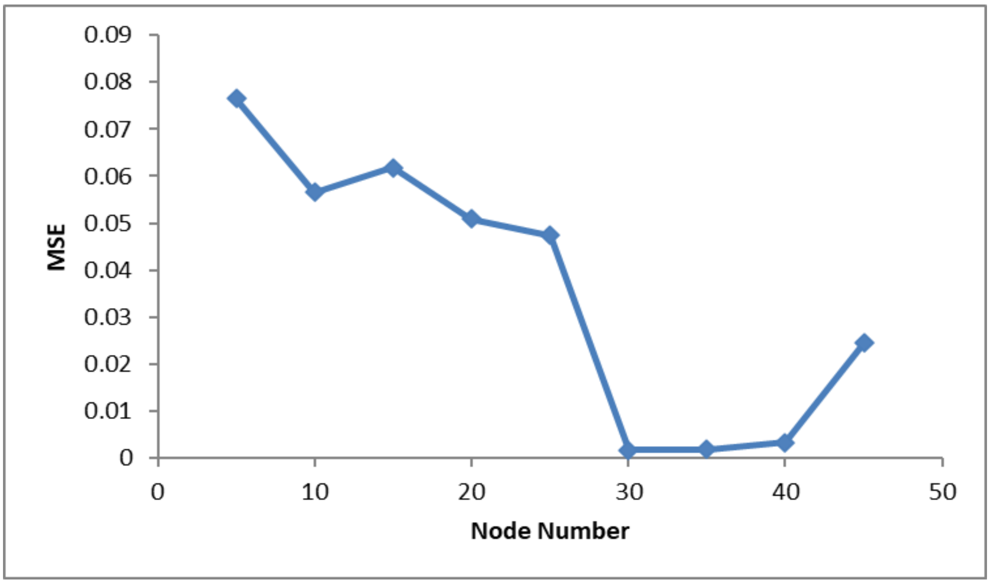

Figure 3. Performance of Network with one hidden layer

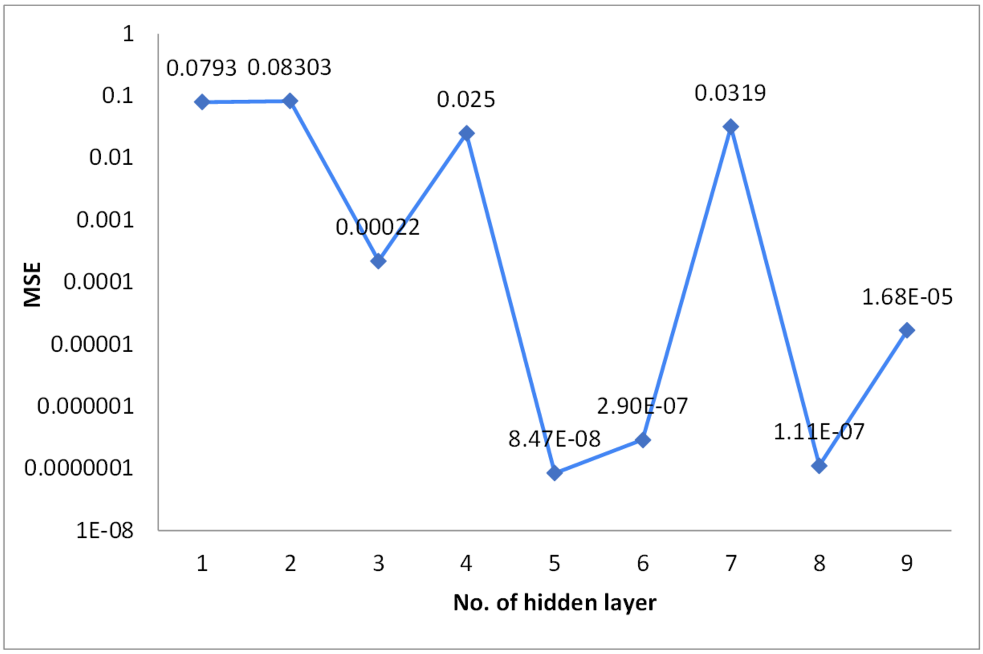

Figure 4. Performance of Network with two hidden layers

5.3. Training Algorithm

The neural networks were trained using training algorithms that try to minimize the value of the performance function, that is (MSE). Table 2 shows the performance of these algorithms in terms of the MSE and the number of (Epoch), as the latter gives an indication of the speed of convergence of the algorithms. Neural network toolboxes are different names function that updates weights and bias value for training:

- Train BFGS quasi-newton method (Train BFGS).
- Train conjugate gradient back propagation (Train CGBP).
- Train gradient descent back propagation (Train GDB).
- Train gradient descent with adaptive learning rate (Train GDA).
- Train gradient descent with momentum back propagation (Train GDM).
- Train gradient descent with momentum and an adaptive learning rate (Train GDMA).
- Train Levenberg-Marquardt optimization (Train LMO).
- Train one-step secant method. (Train OSS).
- Train resilient back propagation algorithm. (Train RBP).
- Train scaled conjugate gradient back propagation. (Train SCG).

A comparison was made between several ANNs that have the same specifications in the MATLAB software package.

The performance of Train LMO algorithms in terms of accuracy and convergence speed in all test cases. The results in Table 2 show that the performance of ANN is able to find a good set of weights and biases. There are three advantages of using a good algorithm. Firstly, good accuracy leads to better performance of an ANN. Secondly, faster convergence leads to lower computation time, Thirdly, its low speed of performance makes it easier and more reliable for ANN.

Table 2. MSE for the Networks with different types of training algorithms

Training Algorithms	MSE	Epochs
Train BFGS	0.2009	1
Train CGBP	0.083128	11
Train GDB	0.09399	100
Train GDA	0.10057	15
Train GDM	0.19433	100
Train GDMA	0.058454	70
Train LMO	0.0016723	93
Train OSS	0.17722	1
Train RBP	0.059143	65
Train SCG	0.083209	15

The performance of the ANN model, that the ANN with 30 hidden layer neurons gives the best performance 0.0016732, and value R for the three sets of training, validation and testing 99.5, 99.5, and 99.7, respectively, as shown in Table 3 and figures 5 and 6, respectively.

Table 3. The performance of ANN model with one hidden layer

No. of hidden layer	Correlation R%				MSE
	Train	Validate	Test	Total	
5	60.9	70.3	49.3	59.5	0.07646
10	72.04	78.7	71.6	72.2	0.05661
15	76.2	78.1	68.9	74.4	0.06187
20	83.1	81.98	88.3	83.4	0.050873
25	62.9	83.3	46.6	62.1	0.0474
30	99.5	99.5	99.7	99.5	0.0016723
35	99.5	99.4	99.1	99.4	0.001925
40	98.9	99.	99.1	98.9	0.003306
45	91.7	92.6	92.6	85.2	0.02447

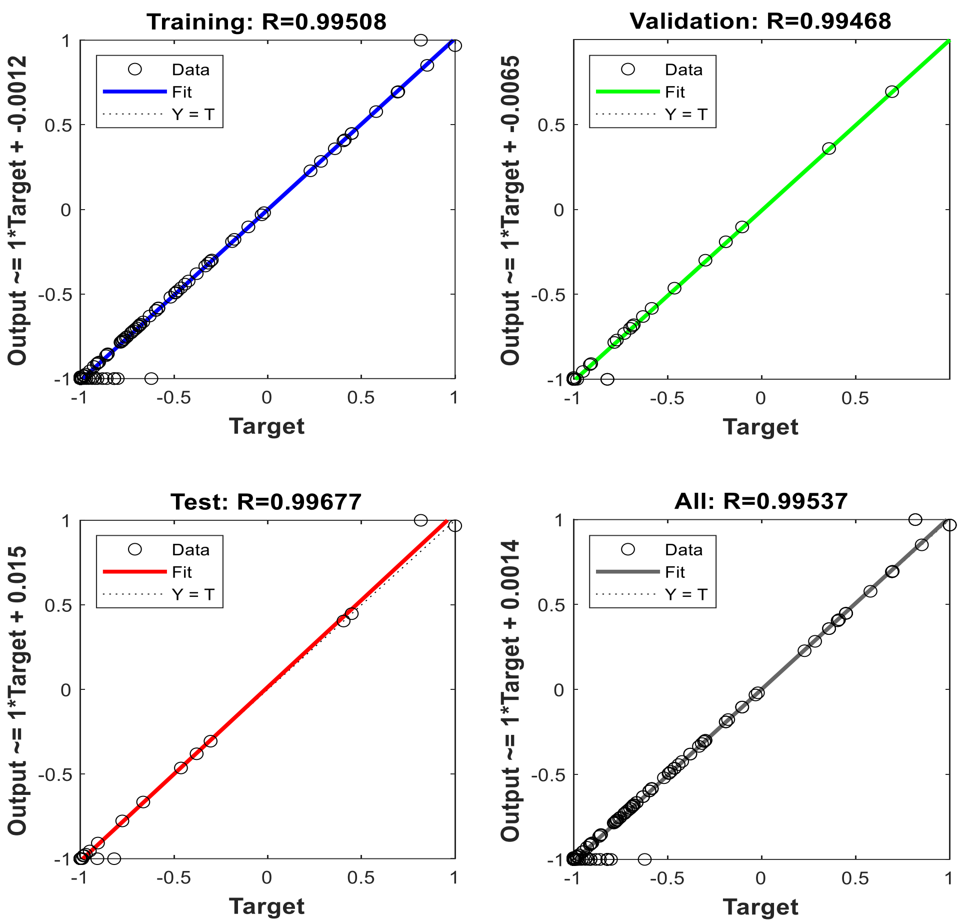

Figure 5. Regression plot for hidden layer 30

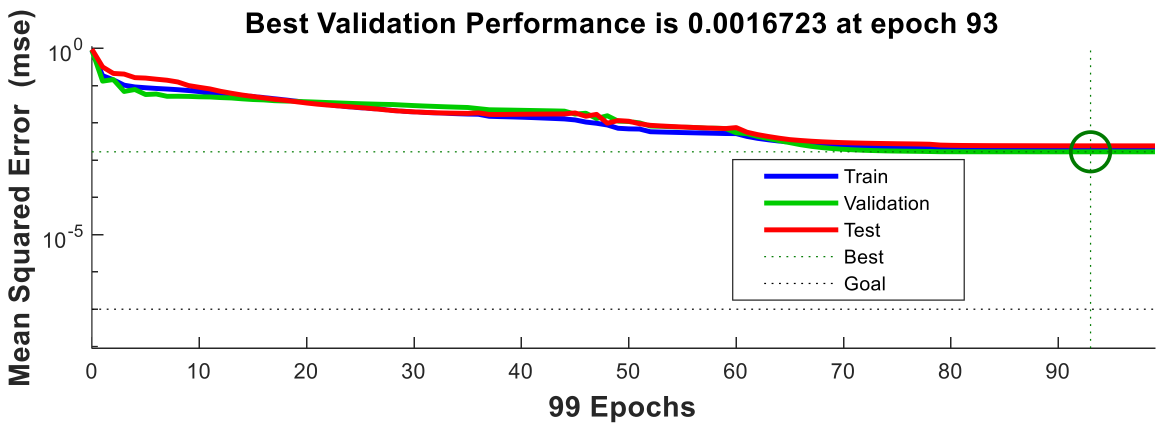

Figure 6. Performance of model for hidden layer 30

The performance of ANN model, that the ANN with (25-5) hidden layer neurons gives the best performance 8.4731E-8, and value R for the three sets of training, validation and testing equal one, as shown in figures 7 and 8, respectively.

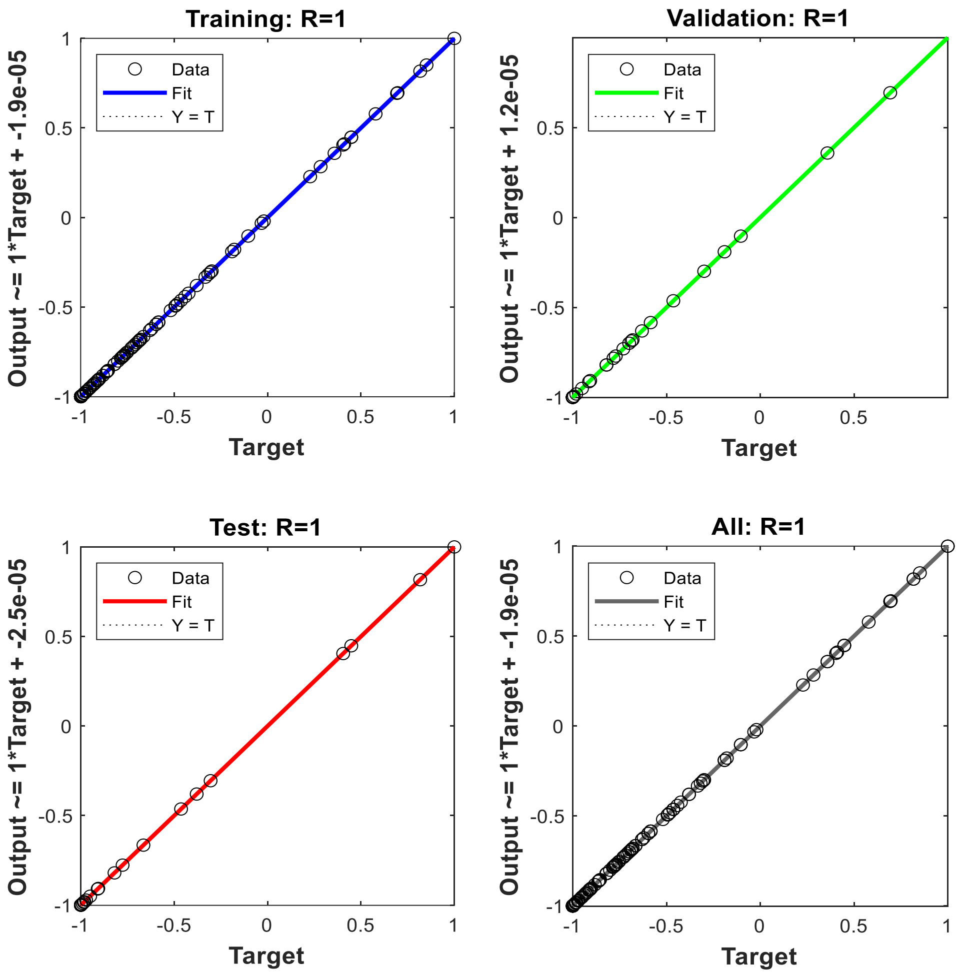

Figure 7. Regression plot for two hidden layers

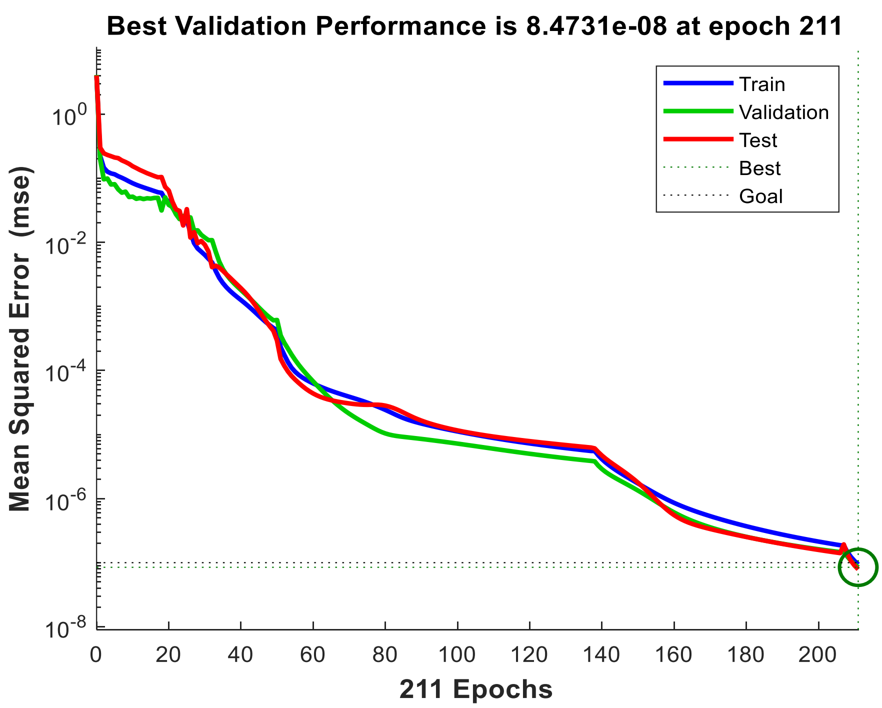

Figure 8. Performance of model for two hidden layers

6. Conclusions

Rainfall is regarded as a key component in the study of the hydrologic cycle, the knowledge of which is required for effective planning and management of water resources. It is calculated both directly and indirectly, however employing direct and indirect approaches has some drawbacks due to the nonlinear complexity of the estimate.

Artificial Neural Networks are a substitute technology for rainfall estimation that helps to solve these issues (ANN).

Mean Square Error (MSE) and Coefficient of Correlation (R) statistics were used to evaluate how well the ANN model performed. The best-fitting model is the one with the highest value (R) and lowest value (MSE). The current investigation produced the following precise findings:

1. It has been discovered that artificial neural networks (ANN) are an effective technique for creating rainfall predictions.
2. Using indicators, R, and MSE, the constructed model's quantitative performance is evaluated.
3. For one hidden layer, model (3-30-1) is found to be the best fit model for the simulation of rainfall, with R and RMSE of 99.5% and 0.0016723, respectively.
4. For two hidden layers, model (3-25-5-1) is found to be the best fit model for the simulation of rainfall, with R and RMSE of 100% and 8.4731E-8, respectively.
5. There is a linear agreement between the output of the ANN and the measured rainfall data, which indicates the possibility of using this technology with future rainfall values in the city of Basrah.

Acknowledgements

The authors appreciate the Meteorological Department for providing a wonderful opportunity to obtain the statistics as well as the University of Basrah/Civil Department with the cooperation of University of Al-Qadisiyah /Civil Department for their valuable assistance in completing this task.

REFERENCES

[1] Lee, J., Kim, C.G., Lee, J.E., Kim, N.W., and Kim, H. "Application of Artificial Neural Networks to Rainfall Forecasting in the Geum River Basin, Korea", water, vol. 10, 1448, no. 1, pp. 1-14, 1018, doi:10.3390/w10101448.

[2] Kaviarasu, K., Sujith, P., and Ayaappan, P. "Prediction of Rainfall Using Image Processing", Int. Conf. of Computational In tell. Computing Res, 2010.

[3] Agboola, A.H., Gabriel, A.J., Aliyu, E.O., and Alese, B.K., "Development of a Fuzzy Logic Based Rainfall Prediction Model", Int. J. Eng. Technol. Vol. 4, no. 3, pp. 427-435, 2013

[4] Lawal, A., Abubakar, U.Y., Danladi, H., and Gana, A.S., "Prediction of annual rainfall pattern using hidden Markov model (HMM) in Jos, Plateau State", Nigeria. J. Appl. Sci. Environ. Manage, Vol. 20, no. 3, pp. 617-622, 2016

[5] Graham, A. and Mishra, E.P., "Time series analysis model to forecast rainfall for Allahabad region", J. Pharmacognosy and Phytochemistry, vol. 6, no. 5, pp. 1418-1421, 2017.

[6] Haq, D.Z., Novitasari, D.C., Hamid, A., Ulinnuha, N., Arnita, Farida, Y., Nugraheni, R.D., Nariswari, R., Ilham, Rohayani, H., Pramulya, R., and Widjayanto, A., "Long Short-Term Memory Algorithm for Rainfall Prediction based on El-Nino and IOD data", Procedia Comput. Sci., vol. 17, no. 9, pp. 829–837, 2021

[7] Sharma, A. and Nijhawan, G., "Rainfall Prediction Using Neural Network", Int. J. Comput. Sci. Trends and Technol., vol. 3, no. 3, pp. 65-69., 2015.

[8] Abhishek, K., Ranjan, R., and Kumar, A. A, "Rainfall Prediction Model using Artificial Neural Network", IEEE Control and System Graduate Research Colloquium (ICSGRC, pp. 82-87. 2012, DOI:10.1109/ICSGRC.2012.6 287140.

[9] Kim, S., Kisi, O., Seo, Y., Singh, V.P., and Lee, C.J. "Assessement of rainfall aggregation and disaggregation using data-driven models and wavelet decomposition", Hydrol. Res., vol. 48, no. 1, pp. 99-116, 2016.

[10] Dash, Y., Mishraa, S.K., and Panigrahib, B.K., "Rainfall prediction for the Kerala state of India using artificial intelligence approaches", Comput. Electr. Eng., vol. 70, no. 1, 66-73, 2018

[11] Haykin, S., "Neural Networks and Learning Machines (3rd Ed.)', Prentice-Hall, New Jersy, vol. 47, no. 3, pp. 27-35, 2009.

[12] Vanrullen, R., "The power of the feed-forward sweep", Advances in Cognitive Psyschology, Vol. 3, No. 2, pp. 167-176, 2007,

[13] Shahin, M. A; Holger, R.M; Jaksa, M. B., "Data Division for Developing Neural Networks applied to Geotechnical Engineering', Journal of Computing in Civil Engineering, Vol. 18, no. 2, pp. 105-114, 2004,

[14] Lu, W., "Neural Network Model for Distortional Buckling Behavior of Cold Formed Steel Compression Members", Ph.D. Thesis, Helsinki University of Technology, June 2000.