MALAYSIAN JOURNAL OF COMPUTING AND APPLIED MATHEMATICS (2024) 7 (2) 37-44

DOI: *https://doi.org/10.37231/myjcam.2024.7.2.130*

Article: Received 07.08.2024; Accepted 13.09.2024; Published 30.09.2024

RAINFALL DAILY PREDICTION BASED ON ARTIFICIAL NEURAL NETWORK

Zahrahtul Amani Zakaria[1,2,✉], Nurul Afiqah Natasha Abdullah[1], Raja Hasyifah Raja Bongsu[1], Nor Aida Mahiddin[1], Siti Dhalila Mohd Satar[1], Maizan Mat Amin[1]

[1]Faculty of Informatics and Computing, Universiti Sultan Zainal Abidin, 22200 Besut, Terengganu, Malaysia

[2]East Coast Environmental Research Institute, Universiti Sultan Zainal Abidin, 22200 Besut, Terengganu, Malaysia

✉zahrahtulamani@unisza.edu.my

Abstract: High rainfall often leads to flooding in certain areas, particularly affecting regions accustomed to wet seasons. Consequently, individuals living in states with high rainfall, especially those along the east coast, must exercise caution. Despite this, few people fully understand the impacts of unreliable rainfall. Statistical forecasting tools for rainfall are generally inadequate for long-term predictions due to the dynamic nature of climate factors. However, the use of Artificial Neural Networks (ANN) for rainfall prediction has advanced significantly. Among the most widely employed methods for rain forecasting, ANN has shown promise. The primary objective is to collect rainfall data and utilize ANN to predict daily rainfall. Additionally, the model's performance is evaluated by comparing the predicted output to the actual output. To implement this solution, rainfall prediction is performed using the Backpropagation Neural Network (BPNN) technique, implemented as a Multi-Layer Perceptron (MLP) in the scikit-learn library, which is tailored for regression tasks. The forecast performance is quantified by measuring the Mean Squared Error (MSE) between the predicted and actual outputs. Both BPNN and MLP are well-established neural network models for rainfall modelling and prediction. These models employ predictive algorithms to achieve high accuracy in rainfall forecasting.

Keywords: Rainfall prediction, Artificial Neural Networks, Backpropagation Neural Network, Multi-Layer Perceptron, East coast regions

1. INTRODUCTION

Malaysia is a country in Southeast Asia comprising different races with various skin colors, religions, traditions, and cultures. The geography of Malaysia consists of major land masses separated by water: Peninsular Malaysia to the West and East Malaysia to the East, together with numerous small islands surrounding those land masses. Malaysia faces two monsoon wind seasons: the Southwest Monsoon, from late May to September, and the Northeast Monsoon, from October to March.

Although Malaysia is not prone to typhoons or tsunamis, there are annual floods in certain areas due to the rainfall period that only ends at the end of every year. The extended rainfall period brought more severe flooding to East Malaysia and the east coast. In terms of natural disasters, floods are the most frequent. This thus led to the use of the Kemaman Station Case Study in Terengganu, Malaysia, to predict rainfall daily concerning the frequency of the monsoon winds affecting the east coast.

Some people still overlook the risk that comes with erratic rains. Unpredictable daily rainfall patterns can present significant challenges, especially in planning projects where the predictability of the weather is significant. Planting and harvesting are done according to the rainfall estimate. Still, unpredictable rainfall patterns can put crop failures, poorer yields, and resultant financial losses. Moreover, unstable and heavy rains can also cause landslides and flash flooding, which can be very dangerous for the people and may damage the infrastructure, including roads, bridges, and houses. Weather-dependent industries that are disrupted because of losing revenue and jobs are tourist and outdoor events. Making investments in state-of-the-art meteorological technologies and early warning systems, as well as enhancing the accuracy of daily rainfall forecasts, are crucial to tracking [1].

2. RESEARCH METHOD AND PROCESS

A. The Artificial Neural Network

The Artificial Neural Network is an engineering notion borrowed from knowledge about the design of the human nervous system [2]. Brain nerve cells are utilised as the primary information processing unit in parallel and instant processing within the central nervous system of human beings. Furthermore, the process of training the ANN has many types and uses, including Perceptron, Backpropagation, Self-Organizing Map (SOM), and Delta [3,4,5]. Therefore, the study proposes a Multi-Layer Perceptron Regression (MLPR) algorithm in rainfall data prediction based on the nonlinear trends of historical data for a more accurate prediction result with minimal error. The MLPR is one of the MLPR and is a supervised algorithm that trains on input-output pairs and will learn complex relationships in the data [6,7]. Another form is the Back Propagation Neural Network, which adjusts weights by backpropagation to minimise prediction errors.

B. Backpropagation Neural Network

On the other hand, the BPNN (Figure 1) is not explicitly separated in the code but is conceptually represented by the MLPR. It's one of the basic neural network trainings using backpropagation to minimise the error function by modifying weights in the network. BPNN, in the code, shall physically be an instance of the MLPR model that emphasises using back-propagation while training the neural network. It is basically a training process for propagating the error backward through the network and updating weights iteratively to reduce error [3,8]. In this context, the implementation of BPNN in the code vividly explains backpropagation's use in training neural networks to perform certain regression tasks, for example, predicting daily rainfall based on historical data.

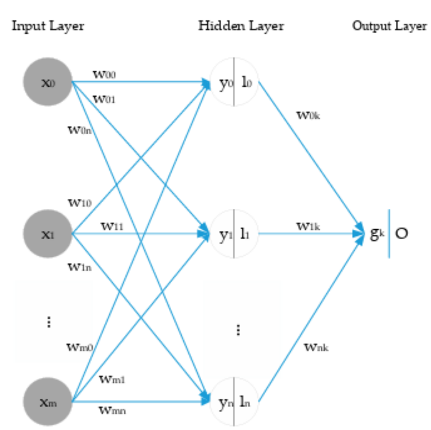


Figure 1: A Structure of BPNN architecture

After being trained, the BPNN uses the learned weights and activation functions to pass inputs through the network and compute outputs, which it then uses to make predictions on fresh data. Because of their capacity to anticipate outcomes, BPNNs are useful for a wide range of tasks, including regression and classification. BPNNs provide a flexible and potent tool for resolving complicated issues in industries like engineering, healthcare, and finance by effectively learning from data [9,10,11].

C. The Multi-Layer Perceptron Regression

Multiple layers of interconnected neurons make up the Multi-Layer Perceptron Regression (MLPR), a neural network architecture intended for regression applications. An input layer, one or more hidden layers, and an output layer are some examples of these layers. Each neuron in the network communicates with neighbouring layers through its connections, with the strength of each connection determined by a weight. The network modifies connection weights during MLPR training to approximate a function that maps input features to continuous output values. Using optimization methods such as stochastic gradient descent, incorrect gradients must be propagated backward for weight changes [12].

In MLPR, the weighted sum of inputs is subjected to an activation function to calculate the neuron output. Sigmoid, tanh, and ReLU are popular activation functions that are selected according to the specifications of the issue and the properties of the data. While MLPR is very good at capturing intricate relationships, it requires careful hyperparameter tweaking regarding layer count, neuron amount, and activation functions to achieve optimal performance. After being trained, MLPR processes inputs through the network using acquired weights and activation functions to predict output values for fresh data in an efficient manner [13].

D. Measure Square Error

The Mean Square Error (MSE) of an estimator is a measure involving the average of error squares, that is, the average squared difference between the estimated values and the actual value. It is a risk function corresponding to the expected value of the squared error loss. It is always non harmful, and the values close to zero are better. The MSE is the second moment of the error (about the origin) and thus incorporates both the estimator's variance and its bias. The formula for MSE is shown in Equation (1).

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^{2} \tag{1}$$

where y_i represents the actual values and $\hat{y}_i$ represents the estimated values.

3. EXPERIMENTAL

A. Determining training and test samples

To get an accurate prediction in the coming year using BPNN and MLPR, rainfall data were divided into two parts: training data and testing data [14]. A total of 50 annual daily maximum rainfall records from 1970 to 2020 were used as a case study, as illustrated by the bar chart in Figure 2. The precipitation was most significant in the year 2020, but it had also risen over time. This may indicate general increases over time, which could have significant implications for infrastructure, farming, and water resources. These, in turn, influence critical environmental issues that warrant further investigation, including statistical testing to assess the significance and potential causes of the observed trend.

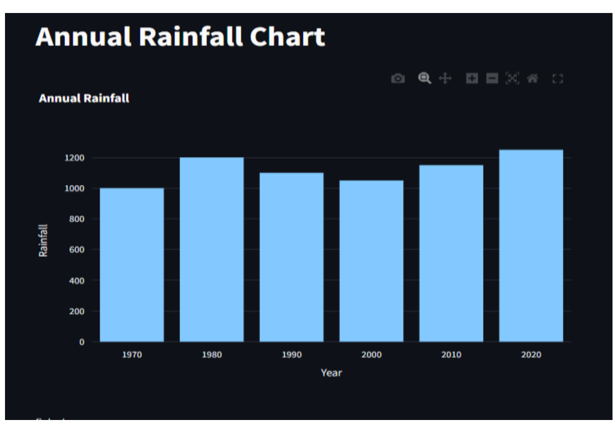
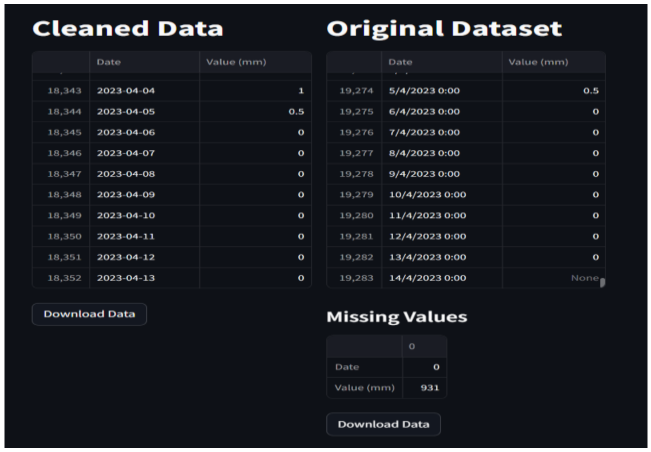


Figure 2: Bar chart of clean rainfall data of Kemaman Station (1970-2020)



Cleaned Data

	Date	Value (mm)
18,343	2023-04-04	1
18,344	2023-04-05	0.5
18,345	2023-04-06	0
18,346	2023-04-07	0
18,347	2023-04-08	0
18,348	2023-04-09	0
18,349	2023-04-10	0
18,350	2023-04-11	0
18,351	2023-04-12	0
18,352	2023-04-13	0

Download Data

Original Dataset

	Date	Value (mm)
19,274	5/4/2023 0:00	0.5
19,275	6/4/2023 0:00	0
19,276	7/4/2023 0:00	0
19,277	8/4/2023 0:00	0
19,278	9/4/2023 0:00	0
19,279	10/4/2023 0:00	0
19,280	11/4/2023 0:00	0
19,281	12/4/2023 0:00	0
19,282	13/4/2023 0:00	0
19,283	14/4/2023 0:00	None

Missing Values

	0
Date	0
Value (mm)	931

Download Data

Figure 3: Rainfall data after cleaning data and real data

Figure 3 presents both the original dataset and a cleaned version, highlighting the process of data preparation. The cleaned dataset likely involves removing outliers, inconsistencies, or missing values. This data pre-processing step is crucial for ensuring the accuracy and reliability of any subsequent analysis or machine learning models. This demonstrates the application of machine learning techniques in environmental science research, aiming to predict future rainfall patterns

B. Flow of MLPR and BPNN for Prediction

The MLPR relies on the same principles as other neural networks, utilizing the backpropagation algorithm for training [15]. Key formulas include:

***Backpropagation Neural Network (BPNN)*:**

Step 0: Initiation of all the weights
Step 1: If the termination condition is not fulfilled
Step 2: For each pair of training data [8], following phases were involved:

***Phase 1:* Forward Pass**

$$a^{(l+1)} = \sigma(W^l a^l + b^1) \tag{2}$$

where $a^{(l)}$ is the activation of layer l, $W^{(l)}$ are the weights, $b^{(l)}$ are the biases and σ is the activation function.

***Phase 2:* Backward Pass**

$$\delta^{(l)} = (W^{(l+1)})^T \delta^{(l+1)} \odot \sigma'(z(l)) \tag{3}$$

where $\delta^{(l)}$ is the error term for layer l, $\sigma^{(l)}$ is the derivative of the activation function and $z^{(l)}$ is the weighted input to layer l.

***Phase 3:* Weight Update**

$$W^{(l)} = W^{(l)} - \eta \frac{aL}{aW^{(l)}} \tag{4}$$

$$b^{(l)} = b^{(l)} - \eta \frac{aL}{ab^{(l)}} \tag{5}$$

where η is the learning rate and L is the lost function (e.g., mean squared error)

***Phase 4:* Model Selection and Training**

Specifically, MLPR and BPNN for rainfall prediction. These models are instantiated using scikit-learn's MLP Regressor with different hyperparameters for experimentation:

Multi-Layer Perceptron Regressor (MLPR):

$$y = f(W_2 \cdot \sigma(W1 \cdot x + b_1) + b_2) \tag{6}$$

where σ is the activation function (ReLU in this case), W_1 and W_2 are the weights and b_1 and b_2 are the biases.

Linear Regression:

$$y = \beta_0 + \beta_1 X + \epsilon \tag{7}$$

where β_0 is the intercept, β_1 is the slope and ϵ is the error term.

The backpropagation algorithm iteratively updates the weights using the gradient descent method. The steps collectively allow the MLPR to learn from historical data and make accurate predictions about future rainfall, contributing significantly to the research's objectives of enhancing rainfall prediction accuracy through advanced machine learning techniques.

C. Measure Square Error (MSE) for Training and Testing

Neural Network Training with Keras:

Function train_model that utilizes the Keras library to create and train a neural network model with a specified architecture and number of epochs. The model architecture can be customised by defining the number of layers and neurons in each layer.

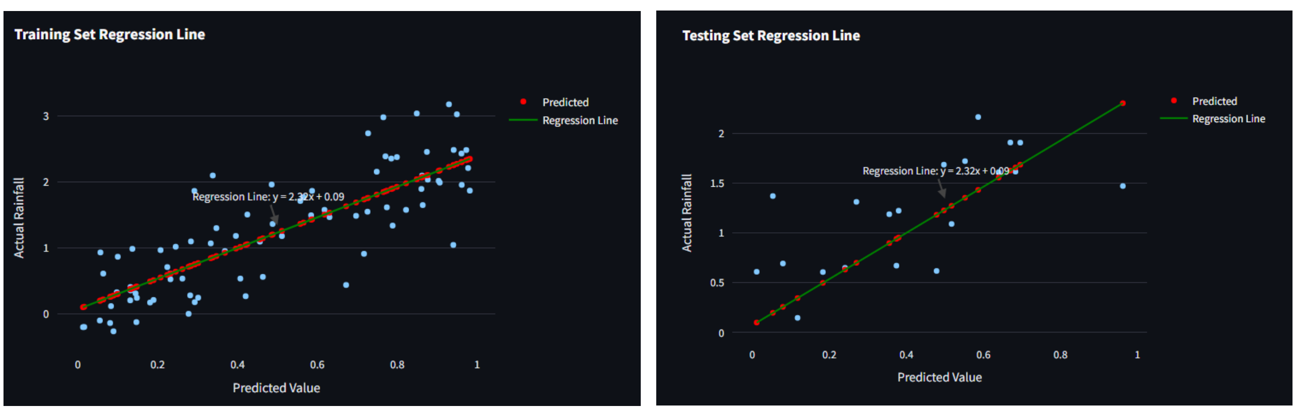


Figure 4: Testing Set Regression Line and Training Set Regression Line

The training set in Figure 4 is essential to the development of a model because it provides the basis for teaching the model to identify relationships, patterns, and representations that are present in the data. The output of the MSE on training set with 0.26 as shown in Figure 5. Thus, the model successfully captures the underlying structure of the training data when the MSE on the training set is low. On the other hand, an abnormally low MSE can be a sign of overfitting, in which the model learns to remember the training set but finds it difficult to generalise to new data. Therefore, the objective is to optimise model performance while avoiding overfitting and achieving a balanced MSE.

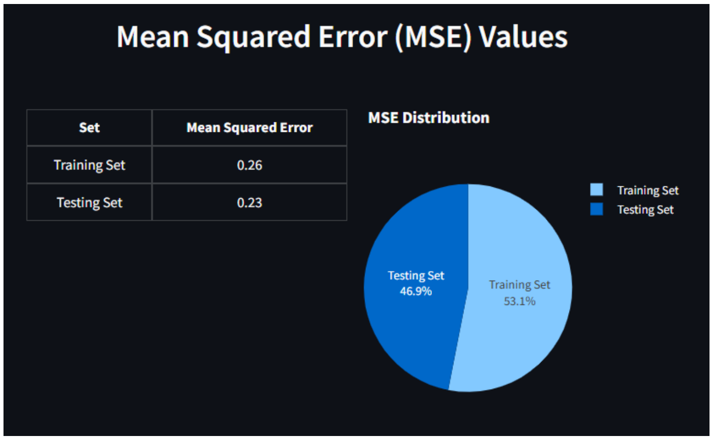


Figure 5: MSE for Training and Testing Set

The testing set is used to evaluate how well the model generalises to new, unknown data; it consists of data that was not seen during model training. The output of the MSE testing set with 0.23 as shown in Figure 5. This small value MSE on the test set suggests that the model can accurately predict new instances. A noticeably higher MSE on the testing set in contrast to the training set can indicate overfitting or other shortcomings in the model. To ensure the model's resilience and dependability in practical applications, the goal is to achieve a low and consistent MSE throughout training and testing sets. This workflow is beneficial for researchers looking to apply artificial neural networks to environmental data for predictive analysis.

4. RESULT AND DISCUSSIONS

In this research, Table 1 evaluated the performance of different MLPR architectures to come up with the best model that would work well for the prediction. Therefore, MSE for both training and testing on various epochs concerning different architectural configurations are considered. Model 1 shows two hidden layers containing 50 and 10 neurons were used, and the model was trained for 500 epochs. Thus, training data gave MSE of 0.2466, while testing resulted in 0.2591. It represents fair accuracy but probably underfitting since the number of epochs was relatively low. Model 2 increased the epochs to 1000 and changed the number of neurons in the neuron's configuration to [50,20]. Then, this model slightly improved to train MSE 0.2408 and test MSE 0.2491. These results suggest that while further training is helpful, it doesn't bother much in generalising. Model 3 had the number of epochs increased to 1500, with almost similar neuron configurations as those of Model 2, and gave the best training MSE of 0.23994, with an improved testing MSE at 0.2466. These results thus suggest that increased training time brings better performance but must always be balanced against overfitting. Experiments demonstrate that performance can be increased by increasing the number of epochs and tuning of the architecture but at a decreasing rate. Model 3 showed the best trade-off between training and testing errors with 1500 epochs and a more profound architecture; hence, it is considered the most robust among the three models tested.

Table 1: Result of MSE Training and Testing for Different Models

MODEL	MLPR	ARCHITECTURE	EPOCH	MSE TRAINING	MSE TESTING
1	MLP Regressor	2 hidden layer(s) with [50, 10] neuron(s)	500	0.2466	0.2559
2	MLP Regressor	2 hidden layer(s) with [50, 20] neuron(s)	1000	0.2408	0.2491
3	MLP Regressor	2 hidden layer(s) with [50, 20] neuron(s)	1500	0.2394	0.2466

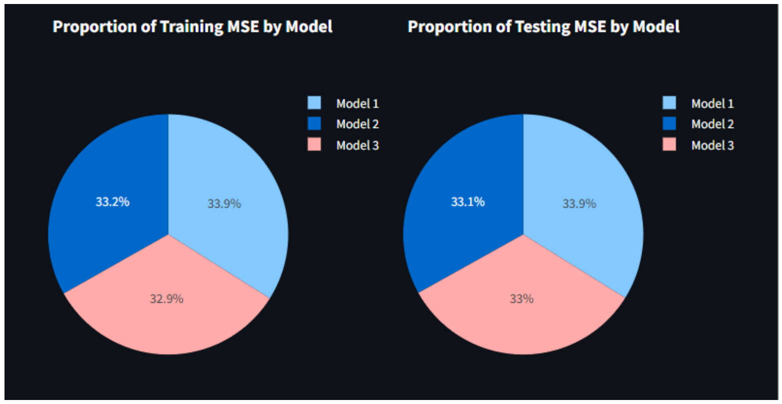


Figure 6: Pie Chart of Training and Testing MSE

Figure 6 shown of the proportion of Training and Testing MSE by Model. Each slice corresponds to one of the three models, indicating which performs comparatively well. Lower proportions in both train and test MSEs are better; these proportions are very close across models, further supporting the point that all models are doing relatively close to each other, some a bit better than others. These results underline the careful tuning of the architecture and duration of training that needs to be performed in any MLPR model if optimal performance is to be reached. Further research into more architectures and other hyperparameters would further refine these findings, probably generalising them for application on other predictive tasks. This study on various configurations of the MLPR model, the importance is therefore brought to the fore of model tuning to reliable predictive performance, offers valuable insights for future applications in similar domains.

5. CONCLUSION

The rainfall at the Kemaman Station in Terengganu was modeled using the MLPR and BPNN algorithms to predict rainfall. The lowest MSE during testing was 0.2466, achieved with an architecture of 2-50-20 and 1500 epochs, after evaluating three models with 500, 1000, and 1500 epochs. Based on the results of this study, the proposed MLPR model shows potential as a predictive algorithm due to its good accuracy. For future studies, it is recommended to compare additional ANN techniques with more parameters to enhance prediction accuracy during the optimization phase.

REFERENCES

[1] Abhishek, K., Kumar, A., Ranjan, R., & Kumar, S. (2012, July). A Rainfall Prediction Model using Artificial Neural Network. In 2012 IEEE Control and System Graduate Research Colloquium (pp. 82-87). IEEE.
[2] Geetha, G., & Selvaraj, R. S. (2011). Prediction of Monthly Rainfall in Chennai using Back Propagation Neural Network Model. International Journal of Engineering Science and Technology, 3(1).
[3] Hardwinarto, S., & Aipassa, M. (2015). Rainfall Monthly Prediction based on Artificial Neural Network: A Case Study in Tenggarong Station, East Kalimantan-Indonesia. Procedia Computer Science, 59, 142-151.
[4] Gomathy, C. K., Bala Narasimha Reddy, A., Pavan Kumar, A., & Lokesh, A. (2021). A Study on Rainfall Prediction Techniques. International Journal of Scientific Research in Engineering and Management, 5(10).
[5] Kumar, A., Deo, M. M., Jeet, P., Kumari, A., & Prakash, O. M. (2022). Daily Rainfall Prediction for Bihar using Artificial Neural Networks: Prediction of Rainfall using ANN. Journal of AgriSearch, 9(4), 320-325.
[6] Wang, W. C., Chau, K. W., Cheng, C. T., & Qiu, L. (2009). A Comparison of Performance of Several Artificial Intelligence Methods for Forecasting Monthly Discharge Time Series. Journal of hydrology, 374(3-4), 294-306.

[7] Xiang, Y., Gou, L., He, L., Xia, S., & Wang, W. (2018). A SVR–ANN Combined Model based on Ensemble EMD for Rainfall Prediction. Applied Soft Computing, 73, 874-883.
[8] RanjanNayak, D., Mahapatra, A., & Mishra, P. (2013). A Survey on Rainfall Prediction using Artificial Neural Network. International Journal of Computer Applications, 72(16), 32–40.
[9] Chen, G., Fu, K., Liang, Z., Sema, T., Li, C., Tontiwachwuthikul, P., & Idem, R. (2014). The Genetic Algorithm Based Back Propagation Neural Network for MMP Prediction in CO2-EOR Process. Fuel, 126, 202-212.
[10] Farajzadeh, J., Fard, A. F., & Lotfi, S. (2014). Modeling of Monthly Rainfall and Runoff of Urmia Lake Basin using “Feed-Forward Neural Network” and “Time Series Analysis” Model. Water Resources and Industry, 7, 38-48.
[11] Shrivastava, G., Karmakar, S., Kowar, M. K., & Guhathakurta, P. (2012). Application of Artificial Neural Networks in Weather Forecasting: A Comprehensive Literature Review. International Journal of Computer Applications, 51(18), 17-29.
[12] Charaniya, N. A., & Dudul, S. V. (2013). Design of Neural Network Models for Daily Rainfall Prediction. International Journal of Computer Applications, 61(14).
[13] Basheer, I. A., & Hajmeer, M. (2000). Artificial Neural Networks: Fundamentals, Computing, Design, and Application. Journal of Methods Microbiological Journal of Microbiological Methods, (43).
[14] Zhao, H. S., Jin, L., & Huang, X. Y. (2010). A Prediction of the Monthly Precipitation Model Based on PSO-ANN and Its Applications. 3rd International Joint Conference on Computational Sciences and Optimization, CSO 2010: Theoretical Development and Engineering Practice, 2, 476–479.
[15] Delashmit, W. H., & Manry, M. T. (2005, May). Recent Developments in Multilayer Perceptron Neural Networks. In Proceedings of The Seventh Annual Memphis Area Engineering and Science Conference, MAESC (Vol. 7, p. 33).