TEST Engineering & Management

March - April 2020
ISSN: 0193-4120 Page No. 11740 - 11746

Prediction of Monthly Rainfall at SENAI, Johor using Artificial Immune System and Deep Learning Neural Network

N.S.Noor Rodi[1], M.A.Malek[2], N.Zaini[3], Amelia Ritahani Ismail[4], M.H.M Hisham[5]
1Civil Engineering Department, College of Engineering, Universiti Tenaga Nasional, Malaysia(pe20675@student.uniten.edu.my)
2Institute of Sustainable Energy , Universiti Tenaga Nasional, Malaysia(Marlinda@uniten.edu.my)
3Civil Engineering Department, College of Engineering, Universiti Tenaga Nasional, Malaysia(Nur_Atiah@uniten.edu.my)
4 Department of Computer Science, Kulliyyah of Information and Communication Technology, Universiti Islam Antarabangsa, Kuala Lumpur, Malaysia(amelia@iium.edu.my)
5School of Education, Faculty of Social Sciences and Humanities, Universiti Teknologi Malaysia Skudai, Johor Bharu, Malaysia (hizwan@utm.my)
Corresponding Author : pe20675@student.uniten.edu.my

Article Info
Volume 83
Page Number: 11740 - 11746
Publication Issue:
March - April 2020

Article History
***Article Received**: 24 July 2019*
***Revised**: 12 September 2019*
***Accepted**: 15 February 2020*
***Publication**: 16 April 2020*

Abstract

In order to obtain good accuracy for the prediction of rainfall, this paper developed the Clonal Selection Algorithm (CSA) as a model for monthly rainfall prediction at Senai, Johor, Malaysia. CSA is one of the main algorithms in the Artificial Immune System. The results were compared with an established model for prediction which is the deep Multilayer Perceptron (MLP) algorithm. MLP is a deep learning algorithm used in the Artificial Neural Network (ANN). The algorithms were modelled using rainfall historical data with four input meteorological variables which are humidity, wind speed, pressure and temperature over the period of 1987 to 2017. The result shows that CSA obtained better prediction accuracy compared to MLP. CSA was applied successfully for the prediction of a continuous time series data with a high variable in nature.

Introduction

Rainfall is a complex atmospheric process which is environment and time dependent. Understanding the rainfall process remains a major challenge, and rainfall prediction is an essential function for water resources management. The accurate prediction of rainfall remains an important factor which has significant implications for food production, securing water supplies for major population centers, and minimizing flood risks [1]. [2][3] mentioned that reliable rainfall prediction on monthly and seasonal time scales can provide useful information for water resource management, agricultural planning, and associated crop insurance applications. Nowadays, changes in the climate affects the rainfalls pattern. The impacts of these changes include the extreme occurrence of flooding and droughts [4]. Indeed, the prediction of rainfall using a good and accurate method is indispensable in order to anticipate the impact.

Rainfall prediction remains a serious issue that has attracted concern from governments, industries, risk management entities, as well as the scientific community. Rainfall is a climatic factor that affects many human activities like agricultural production, construction, power generation, forestry, and tourism among others [5]. In order to establish the relationship between rainfall with other meteorology parameters, [6] conducted a study using both regression and

March - April 2020
ISSN: 0193-4120 Page No. 11740 - 11746

non-response analysis. The study found that the relationship between wind speed and pressure is co-dependent with temperature. In addition, the relation between wind speed and pressure within a storm was determined. In a more complex condition, the relationship among wind speed, pressure and temperature was considered near to the water's surface. Most studies on rainfall prediction use meteorological parameters to achieve their objectives. [7] Araki classified the importance of diurnal variation if wind, convective cloud and temperature give an impact to the occurrence of rainfall. Then, Kilsby et al. [8] studied the significance of humidity, temperature, wind and sunshine on rainfall prediction whereby the variables were incorporated as indicators for weather generation. The study on the relationship between minimum and maximum temperature to rainfall was established by Carrera-Hernandez et al. [9]. However, numerous problems concerning the data relationship have occurred, especially with regards to forecast errors which are responsible for distribution and observational increments in space, as well as among model variables. When the relationship in the subject data was statistically examined, the estimation of each variable as a function for the other variables was determined or identified depending on the time and on the non-functional relationship obtained. The rapid development in Johor led to additional river improvement works; thus, it is time to develop new techniques to address the issues of design and management of flood and water resources. In this study, rainfall was predicted based on the relationship between humidity, pressure, wind speed, temperature and rainfall using the Artificial Immune System and Deep Learning Neural Network.

Artificial Immune System

Artificial Immune System (AIS) is a kind of computational intelligence system which was inspired by the information processing mechanism of the biological immune system. AIS, also known as immune computation, is a fast-developing research area in the computational intelligence community. Over past two decades, researchers aimed to develop immune-based models and techniques to solve complex computational or engineering problems as mentioned by Castro and Tmmis [10]. The function of a biological immune system is to protect the body from foreign molecules known as antigens. This system has a great pattern recognition capability that may be used to distinguish between foreign cells entering the body (non-self or antigen) and the body cells (self). Visually, it can be described as in Figure 1.

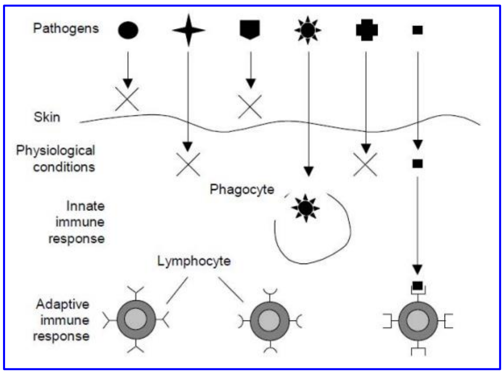

Figure 1 Multi layer structure in immune system[11]

Immune systems have many characteristics such as uniqueness, autonomy, recognition of foreigners, distributed detection, and noise tolerance [11]. The immune system is a host defense system comprising many biological structures and processes within an organism that protects against diseases. An immune system must detect a wide variety of antigens, known as pathogens, from viruses to parasitic worms, and distinguish them from the organism's own healthy tissue to make sure it can function properly [12]. One of the main concepts in AIS is the clonal selection theory. This theory was introduced by Burnet F.M in 1957 [13]. It was used to explain the basic response of an adaptive immune system to an antigenic stimulus. It establishes the idea that only cells capable of recognizing an antigen will proliferate while other cells are selected against it. Clonal selection operates on both B and T cells. B cells, when their antibodies bind with an antigen, are activated and differentiated into plasma or memory cells. Prior to this process, clones of B cells are produced and undergo somatic hyper mutation. As a result, diversity is introduced into the B cell population. Plasma cells produce antigen-specific antibodies that work against the antigen. Memory cells remain with the host and promote a rapid secondary response [10] as explained in Figure 2.

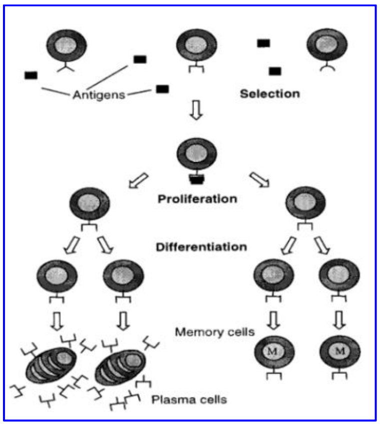

Figure 2 Clonal selection principle concept[10]

There are not many studies which apply artificial immunity algorithms to the problem of rainfall prediction. Xiang Weiguo performed weather prediction using AIS which used the rainfall and meteorology historical data and compared the results with the Artificial Neural Network (ANN) algorithm.[14]. Burcu performed a study on protein secondary structure prediction using clonal selection and multilayer perceptron [15]. Jose Pedro Alves [16] used AIS to predict building total energy consumption.

Deep learning neural network

Deep learning is becoming a more exciting application in several domains of bioinformatics [17]. Deep Learning Neural Networks (DLNN) has been widely used in many fields and it is an important method for machine learning. This algorithm is inspired by the structure of the mammalian visual system which contains many layers of neural networks. It processes the message from the retina to the visual center layer by layer; sequentially extracts edge feature, part feature, shape feature; and eventually forms an abstract concept [18]. In general, the depth of DLNN is greater than or equal to 4. For example, a MLP with more than 1 hidden layer is a DNN framework. DNN able to find distributed expression of data by extracting the feature layer by layer and combine low-level features to form high-level features . Compared with shallow neural network, DLNN has better feature expression and ability to model complex mapping [19].

Indrastanti [20] used multilayer perceptron as part of the deep learning method to predict flood events based on rainfall time series data and water levels in a weir. By using this MLP, the study obtained the MAPE value of 3.64%, which means the error generated in the system built is 3.64% compared with the real value used in testing. The result was compared to the multiple regression linear, which showed that MLP had a better result in predicting water elevation level.

Parallel with the increasing popularity of deep learning in recent years, the number of applications in the prediction of rainfall has also increased day by day. Yen. M et al used the deep learning echo state network for hourly rainfall prediction [21]. Advanced weather forecasting using deep learning was studied by A. Subashini et al. As the result, LSTM neural network has given substantial results with high accuracy among other weather forecasting techniques [22].

Study area and data collection

In this paper, historical meteorology data were collected from 1987 to 2017 for rainfall prediction. The proposed deep learning model was tested using different parameters in order to get the best output behaviors. The historical monthly data for Senai station was taken from the Malaysia Meteorological Department. Thirty years of monthly data for five weather parameters were collected, which include humidity, wind speed, temperature, pressure and rainfall. Rainfall variability is affected by various atmospheric conditions. These variables were selected because they are interdependent and influence precipitation. Temperature is a measure of the ability of the atmosphere and water to receive and transfer heat from other bodies. Humidity is the water vapor content of the air. As a normal relative, if the humidity of the air below the cloud is low, the evaporation can significantly reduce the amount of water in falling raindrops, therefore reducing the rainfall.

Methodology for CSA

Figure 3 shows the flowchart of the clonal selection algorithm used in this study.

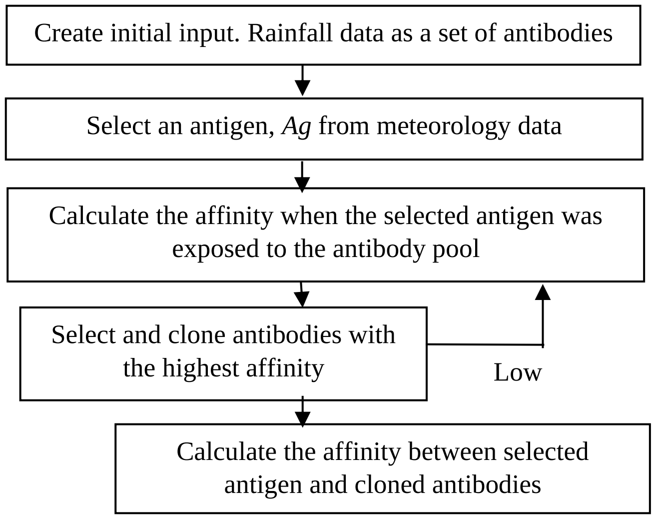

Figure 3 Flow chart of CSA

The algorithm code for clonal selection was written in Matlab R2017a. The initialization of data begins with the population of antibodies, P which uses real values of historical rainfall. Then, a set of antigens was prepared using real values of weather parameters (humidity, pressure, temperature and wind speed) at the same period.

The initial data can be form as

$$P = [\ Ab_1, Ab_2, \ldots Ab_n\ ]$$
Equation 1

$$Ag = [\ Ag_1, Ag_2, \ldots Ag_n\ ]$$
Equation 2

Training is the first stage for data analysis. For this stage, the raw meteorological data were directly used as the antibodies and antigen and were then run in a clonal selection algorithm. Then, the antigens from the parameters randomly attack the antibodies to create the bonding. The bonding between antibodies and antigens were calculated and named as the affinity measure.

Then, the affinity measure for the cloning process was determined using Euclidean Distance, D equation. The equation is as shown in Equation 3.

$$D = \sqrt{\sum_{i=1}^{L}(Ab_i - Ag_i)^2}$$
Equation 3

Where,

Ab is antibodies

Ag is antigen

The testing process of the proposed prediction model begins once the cloned Abs were generated. Then, the second measurement of affinity values between the selected Ag and cloned Abs (get from previous step) were calculated. The selected Ags were obtained from ten numbers for each meteorological parameter that was calculated in the training process. This was to ensure that an accurate model verification is performed. In this study, ten numbers of cross validations were used to ensure the validity of the data. In order to obtain the selected Ag, the forecasting rate (β) obtained from the training process was used again in this stage. The hyper mutation mechanism was carried out at each antibody in a population of P which is submitted to a mutation that is inversely proportional to the affinity. The percentage of accuracy for the proposed model is calculated. This marks the end of the testing process. Lastly, the prediction process was performed using the final cloned Abs.

Methodology for Multilayer Perceptron

The results from CSA were compared with the deep learning neural network, which is a deep multilayer perceptron. Although deep learning is quite popular in recent years, actually it is not a new topic. It is possible to solve a problem with increasing complexity via Artificial Neural Networks (ANN) by increasing the number of layers. Although the increasing number of layers refers to a deeper decision mechanism, it also means more processor power. Technological development in recent years has enabled studies with deep learning. There are five input features or parameters which are rainfall, humidity, air pressure, wind speed and temperature. In theory, MLP can have many layers but must have at least three: the input, a hidden layer, and the output layer. In the deep learning neural network, the number of hidden layers can be added to increase the bonding or connection numbers among the neurons. All the neurons in the MLP are connected to each neuron in the layer at both sides. When initializing the MLP model, the parameters of the MLP neural network and the number of hidden layer nodes are designed. In the algorithm, the hidden layer cannot be directly accessed as the input and output. Each hidden layer consists of several neurons. Neurons are connected between different layers using weight and bias. The weights used are computed from the training set. The input set needs to be sampled to obtain the training and testing sets. Figure 4 shows the MLP architecture.

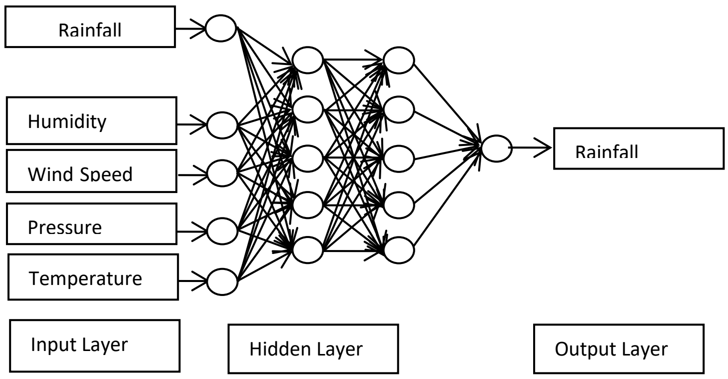

Figure 4 MLP architecture

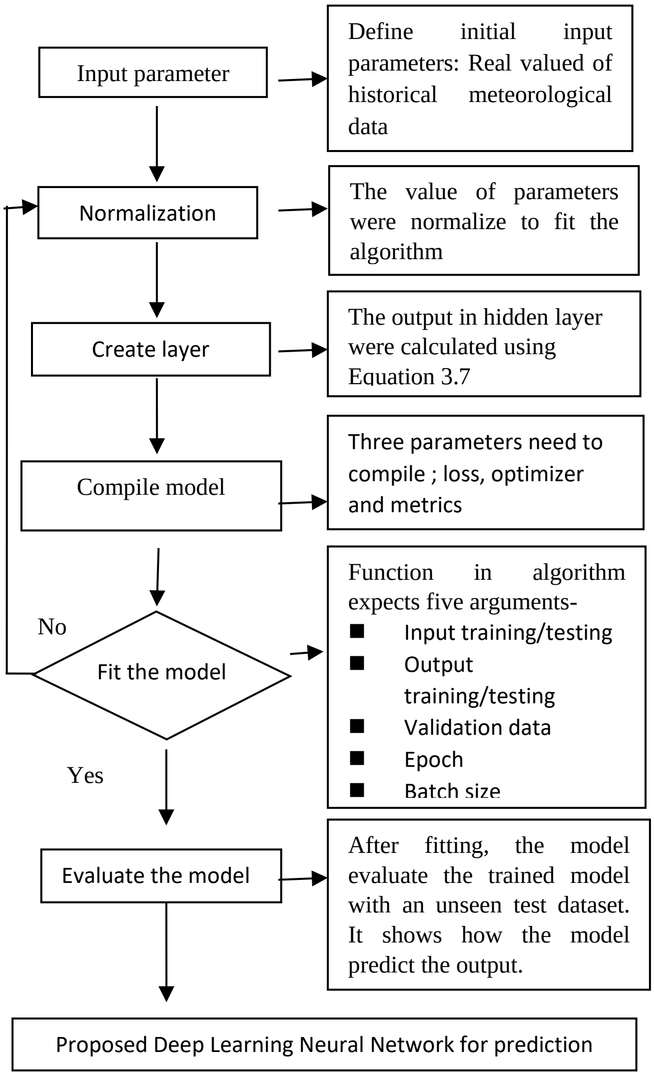

Figure 5 The flowchart of Deep Learning Neural Network

March - April 2020
ISSN: 0193-4120 Page No. 11740 - 11746

Figure 5 shows the methodology of rainfall prediction using the deep neural network algorithm. Table 1 shows the parameter setting for MLP algorithms.

Table 1 Parameter setting for MLP

Number of epochs	1000
Number of features	420
Number of hidden layer	3
Neurons in each layer	100
Number of train sample iterations	10
Activation function	ReLU

The data processing was divided to 70% for training and 30% for validation and testing. K fold cross validation technique was used for validation, and the sampling was done randomly into k equal sample. Out of all these k sub samples, one sub sample was retained for testing and the rest of the k-1 sub samples were used for training the model. For optimization, once all the iterations were completed for each sub sample, the results from the individual iterations need to be combined to produce one single result. Equation 4 shows how to obtain the output of neuron *j* in the hidden layer

$$H_j = f\left(\sum_{i=1}^{n} w_{ji} x_i + b_i\right)$$

Equation 4

Where w_{ij} and b_i are the weights and biases of the hidden layer neurons. Then, f(.) is the output layer neuron activation function. The activation function used for this model is the Rectified Linear Unit (ReLU). It can be calculated using Equation 5

$$f(x) = \frac{1}{1+e^{-x}}$$

Equation 5

Statistical evaluation of model performance

The correctness and accuracy of the model can be checked using the Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE) and Percentage Accuracy (PA) function. RMSE measures the difference between the values calculated by the model and the actual values. Meanwhile, MAPE is calculated by using absolute error in each period divided by the observed values that are apparent for that period. Next is the average absolute percentage error. This approach is useful when the size of the prediction variable is important in evaluating the prediction accuracy. MAPE indicates how much error exists in predicting the compared with the real value [23].

Result and discussion

The training process of the sample data set is the main factor that affects the learning and generalization abilities of the population. The training results are shown in graphical comparisons to further prove the reliability of the proposed model. The proposed CSA based rainfall prediction model was trained through RMSE with ten (10) runs for each number of detectors tested from 1987 to 2008 in order to get the best number of detectors. The number of detectors run were 100, 200, 300, 400, 500, 600, 700, 800, 900 and 1000. The result is presented in Figure 6.

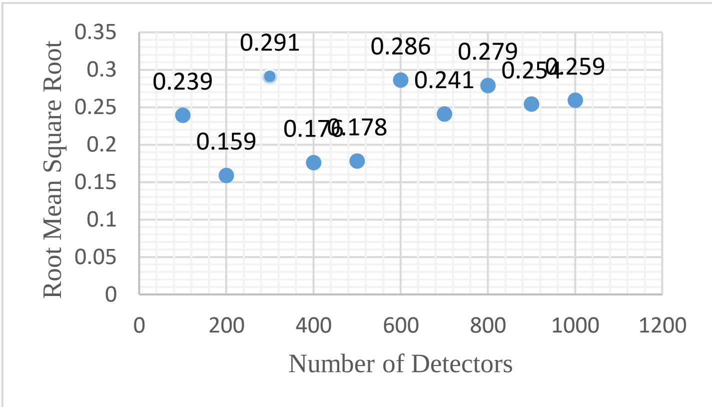

Figure 6 RMSE against number of detectors for CSA

Figure 6 shows the RMSE results against ten detectors from the simulation in the training stage. According to the graph, the lowest RMSE value obtained was 0.159 which is for 200 iterations. The detector's number for the lowest value of RMSE was used as the number of detectors for the next simulation testing. In MLP, the training result was observed by changing the number of hidden layers and epochs. The best number of layers used for this was 10. Dropout is an effective regularization technique to avoid overfitting in the neural network. By neglecting 50% of neurons of each hidden layer from the input parameters, it reduces the memorization of the data [15].

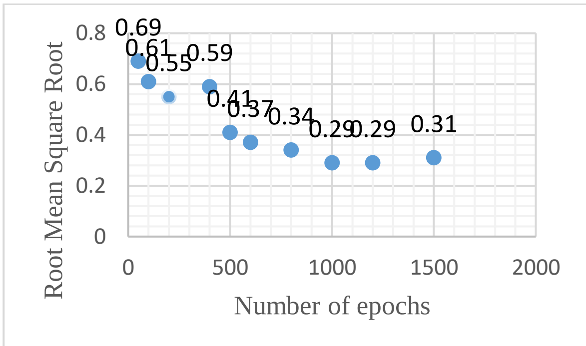

Figure 7 RMSE against number of epochs for deep MLP

TEST Engineering & Management
March - April 2020
ISSN: 0193-4120 Page No. 11740 - 11746

The simulations were run and tested using different numbers of epochs. Based on the result in Figure 7, the RMSE results were put against ten epochs. According to the graph, the lowest RMSE value was obtained is 0.29 which is for 1000 and 1200 iterations. A thousand epochs were used as the number of epochs for the next simulation in testing. After the learning training was over, 30% of the remaining meteorological data of Senai (not used in system training learning) was used to validate the prediction performance of the system towards the rainfall data to make comparisons with the deep MLP system based on the neural network.

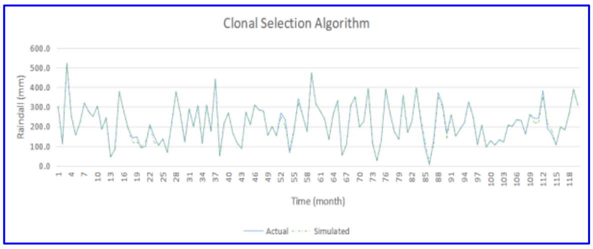

Figure 8 The comparison between actual and simulated rainfall at testing stage using CSA

Figure 8 shows an almost similar trend between actual and simulated values for rainfall using CSA. From the simulation, the similarity percentage between actual and predicted rainfall is 89%. The highest value of simulated rainfall was 521mm.

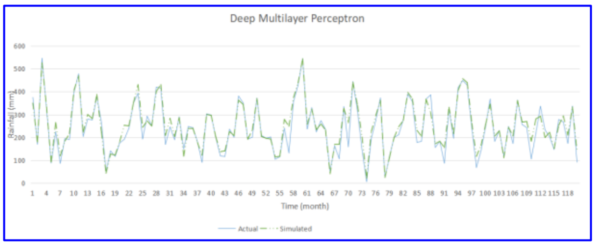

Figure 9 The comparison between actual and simulated rainfall at testing stage using MLP

Figure 9 shows the trend between actual and simulated values for rainfall using deep MLP. From the simulation, the similarity percentage between actual and predicted rainfall is 79%. The highest value of simulated rainfall was 512mm.

Table 2 shows the comparison result between CSA and deep MLP. From the result, CSA obtained a higher percentage of accuracy, which is 89% compared to MLP 79%. The RMSE and MAPE for CSA also produced better results compared to MLP.

Table 2 Comparison result between CSA and Deep MLP

	Clonal Selection Algorithm	Deep Multilayer Perceptron
RMSE(%)	14.95	18.65
MAPE(%)	25.65	30.34
PA(%)	89	79

Conclusion

In this study, an algorithm inspired from the clonal selection theory approach was proposed to predict monthly rainfall for the Senai area, located in Johor, Malaysia. The prediction result was then compared with MLP, which is a deep learning technique. During the training stage, the CSA came out with 200 iterations as the best model as it gave the lowest RMSE value of 0.159. Then, for the MLP, the number of epochs chosen was 1000. The results obtained indicate that the accuracy for CSA at 89% was higher compared to MLP at 79%. This study highlighted the ability of CSA in predicting rainfall. As MLP is well known as a good rainfall prediction tool, this study concludes that an algorithm inspired by clonal selection theory in the artificial immune system can be applied successfully for the prediction of continuous time series data such as monthly rainfall data which is highly variable in nature.

Acknowledgement

The authors would like to acknowledge technical and financial support from Yayasan Tenaga Nasional (YTN) Scholarship, UNITEN RMC Internal Research Grant: RJO 10517919/iRMC/Publication and access to data from Malaysia Meteorological Department (MetMsia).

References

1. Xinguang He, Huade Guan,Jianxin Qin , Journal of Hydrology: A hybrid wavelet neural network model with mutual information and particle swarm optimization for forecasting monthly rainfall, 2015
2. Everingham, Y., Baillie, C., Inman-Bamber, G., Baillie, J: Forecasting water allocations for Bundaberg sugarcane farmers. Clim. Res. 36 (3), 231–239, 2008
3. Garbrecht, J.D., Zhang, X.C., Schneider, J.M., Steiner, J.L.: Utility of seasonal climate forecasts in management of winter-wheat grazing. Appl. Eng. Agric. 26 (5), 855–866, 2010
4. Kumagai, T. and T. Kume: Influences of diurnal rainfall cycle on CO2 exchange over bornean tropical rainforest. Ecological Modelling, 2012.

March - April 2020
ISSN: 0193-4120 Page No. 11740 - 11746

5. World Health Organization: Climate Change and Human Health: Risks and Responses. World Health Organization, January 2003
6. Wooten, R.D.: Statistical Analysis of the Relationship Between Wind Speed, Pressure and Temperature. Journal of Applied Science, Volume 11(15), pp.2712-2722. 2011
7. Araki, R., Yamanaka, M.D., Murata, F., Hashiguchi, H., Oku, Y., Sribimawati, T., Kudsy, M., and Renggono, F. Seasonal and Interannual : Variations of Diurnal Cycles of Wind and Cloud Activity Observed at Serpong, West Jawa, Indonesia. *Journal of Meteorological Society of Japan*, Volume 84A, pp. 171-194, 2006
8. Kilsby, C.G., Jones, P.D., Burton, A., Ford, A.C., Fowler, H.J., Harpham, C.James, P., Smith, A., and Wilby, R.L. : A Daily Weather Generator for Use in Climate Change Studies. *Journal of Environmental Modelling and Software*, Volume 22,pp. 1705-1719, 2007
9. Carrera-Hernandez, J.J., and Gaskin, S.J. Spatio Temporal Analysis of Daily Precipitation and Temperature in the Basin of Mexico. *Journal of Hydrology*, Volume 336, pp.231-249, 2007
10. Castro L. de and Timmis J: An Artificial Immune Network for Multimodal Function Optimization. Proceedings of IEEE Congress on Evolutionary Computation (CEC'02), vol. 1, pp. 699-674, May, Hawaii. 2002
11. L. N. De Castro and F. J. Von Zuben, “Artifificial immune systems: Part i–basic theory and applications,” *Universidade Estadual de Campinas, Dezembro de, Tech. Rep*, vol. 210, no. 1, 1999.
12. Pankaj Chaudhary, Kundan Kumar ,Artificial Immune System: Algorithms And Applications Review ,International Journal of Information and Communication Technology,Volume 1 Issue 1, January-June 2018
13. Burnet F.M.,: The Clonal Selection Theory of Acquired Immunity. Cambridge University Press,1959
14. Xiang Weiguo. Information Technology and Applications (IFITA), 2010 International Forum on , Vol 2 ; 386 – 389,2010
15. Prediction of Protein Secondary Structure with Clonal Selection Algorithm and Multilayer Perceptron BURCU ÇARKLI YAVUZ1 , NİLÜFER YURTAY2 , and OZHAN OZKAN3,2014
16. Jose Pedro Alves, jose Nuno Fidalgo, Classifification of Buildings Energetic Performance Using Artificial Immune Algorithms, 2019
17. Min S, Lee B, Yoon S. Deep learning in bioinformatics. Brief Bioinform.; 18(5):851–69,2017
18. T. Serre, G. Kreiman, M. Kouh, et al., “A quantitative theory of immediate visual recognition,” Progress in Brain Research, vol 165, pp. 33-56, 2007.
19. Huang Yi, Duan Xiusheng : A Study on Deep Neural Network, Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC), 2016
20. Indrastanti R. Widiasari, Lukito Edi Nugroho : Deep Learning Multilayer Perceptron (MLP) for Flood Prediction Model Using Wireless Sensor Network based Hydrology Time Series Data Mining,2017
21. Yen, M., Liu, D., Hsin, Y. *et al. :* Application of the deep learning for the prediction of rainfall in Southern Taiwan. *Sci Rep* **9,** 12774,2019
22. A. Subashini1 , S M. Thamarai2 , Dr. T. Meyyappan3,Advanced Weather Forecasting Prediction using Deep Learning,International Journal for Research in Applied Science & Engineering Technology (IJRASET),2019
23. Chris Tofallis, "A Better Measure of Relative Prediction Accuracy for Model Selection and Model Estimation," Journal of the Operational Research Society, vol. 66, pp. 1352–1362, 2015.