# Rainfall Classification for Flood Prediction Using Meteorology Data of Kuching, Sarawak, Malaysia: Backpropagation vs Radial Basis Function Neural Network

Soo See Chai, Wei Keat Wong, and Kok Luong Goh

*Abstract*—**Rainfall is often defined by stochastic process due to its random characteristics, i.e. space and time dependent and it is therefore, not easy to predict. In general, rainfall is a highly non-linear and complicated phenomenon. In order to acquire an accurate prediction, advanced computer modeling and simulation is required. Artificial Neural Network (ANN) has been successfully used to predict the behavior of such non-linear system. Among the different types of ANN models used, Backpropagation Network (BPN) and Radial Basis Function Networks (RBFN) are the two common ANN models that had produced valuable results. However, there was no study conducted to research on which, among these two methods, is the better model for rainfall forecast. Therefore, this study will fill this gap by comparing the capabilities of these two ANN models in rainfall forecast using metrological data from year 2009 to 2013 obtained from Malaysian Meteorological Department for Kuching, Sarawak, Malaysia. From the research, it is concluded that, BPN (MSE≈0.16, R≈0.86) performs better as compared to RBFN (MSE≈0.22, R≈0.82). The strengths and weaknesses of these models are also presented in this paper.**

*Index Terms*—**Artificial neural network, backpropagation, radial basis function, rainfall, classification.**

## I. INTRODUCTION

Weather forecasting is a complicated procedure yet the most essential and vital process for the mankind nowadays, because it severely affect hu-man activities. Highly accurate weather forecast could help to prevent casualties and damages. Amongst all the weather happenings, floods are the leading cause of natural disaster death world-wide and were responsible for 6.8 million death in the 20th century [1].

The flood-prone areas in Malaysia is approximately 29,000km2 and the population on these areas is more than 4.82 million people, which is 22% of the whole Malaysian [2]. The damage caused by flooding is estimated to be around RM915 million (£160 million). The impact of flood to Malaysia could be seen in the increase budget allocated for flood mitigations in the 5-year Malaysia Plan shown in Table I.

TABLE I: FLOOD MITIGATION EXPENDITURE IN MALAYSIA [3]

| Period | RM | Remark |
|---|---|---|
| 1971-1975 | 14 million | 2nd Malaysia Plan |
| 1976-1980 | 56 million | 3rd Malaysia Plan |
| 1981-1985 | 141 million | 4th Malaysia Plan |
| 1986-1990 | 155 million | 5th Malaysia Plan |
| 1991-1995 | 451 million | 6th Malaysia Plan |
| 1996-2000 | 845 million | 7th Malaysia Plan |
| 2001-2005 | 1.8 billion | 8th Malaysia Plan |
| 2006-2010 | 4 billion | 9th Malaysia Plan |
| 2010-2015 | 5 billion | 10th Malaysia Plan |

Rainfall is one of the most significant parameters in hydrological model. Due to the importance of rainfall, large numbers of attempts have been made to predict rainfall accurately using various traditional statistical and numerical methods. However, due to the nonlinear characteristics of rainfall, the prediction accuracies obtained by these methods were still not at satisfactory level. Recent advancements entitled Data Driven Modeling (DDM) which encompasses computational intelligence has emerged. DDM is based on analyzing the data about the system, in particular finding connections between the system variables (input, internal and output variables) without explicit knowledge of the physical behavior of the system. Artificial Neural Network (ANN) algorithms have been successfully applied in rainfall classification. Among the different ANN algorithms applied, the Backpropagation Neural Network (BPN) and Radial Basis Function Network (RBFN) are the two most commonly used in rainfall prediction and yield satisfactory results [4]. However, there is no research on verifying which method applied better accuracy result as comparing to the other.

## II. DATA

### A. Meteorology Data

The study area selected for this study is Kuching city, the capital city of Sarawak located in the east Malaysia. Kuching, one of the fast and highly developed cities in Malaysia, is located at the southwest of the Sarawak state with latitude 1.6019N and longitude 110.3244E. Kuching city has a cover area of 895.09 km2 and a population of 681,901 [5]. Malaysia experiences an average of 143 flood events annually between 2001 to 2015 and more than 90% of these events are flash flooding. Sarawak, a state on Borneo in Malaysia, experiences an annual rainfall of around 3000-4000mm. Kuching, the capital city of Sarawak, is established besides

Sarawak River about 30 km from the estuary. The low-lying landform, in addition to the extremely high annual rainfall, contributes to flash flooding events to the city. Therefore, accurate classification of flood is crucial for Kuching city. For the purpose of this research, a collection of historical meteorology parameters composed of daily measurement of six elements observed in Malaysian Meteorological Department were collected. These daily meteorology data consists of minimum temperature (℃), maximum temperature (℃), mean temperature (℃), mean relative humidity (%), mean wind speed (m/s), mean sea level pressure (hPa) and mean precipitation (mm) was collected from year 2009 to year 2013. The observed meteorology data has been classified into four features to ease the forecasting process that is light precipitation, moderate precipitation, heavy precipitation, and very heavy precipitation. The classification of these was based on the research report done by [6] which heavily dependent on the precipitation data. Table II shows the classification of data based on the precipitation.

TABLE II: RAINFALL EVENT CLASSIFICATION

| Precipitation (mm) | Rainfall Event Classification |
|---|---|
| Light Precipitation | < 10mm |
| Moderate Precipitation | 11mm – 30mm |
| Heavy Precipitation | 31mm – 50mm |
| Very Heavy Precipitation | > 51mm |

### B. Data Pre-processing

The existence of noise and data missing will affect the performance of ANN result [7]. Therefore, before applying the data into the ANN models, data checking was done for incomplete data. Missing data which was confirmed by the Sarawak Meteorology Station were removed from the dataset. The dataset was next normalized. Normalization is done to ensure good results could be obtained and to prevent numerical difficulties occurs during performing calculation [8]. Moreover, normalization speeds up the training process of the ANN and reduces the likelihood of the ANN getting stuck in local minima [9]. For this research, the meteorology data was normalized to fall between -1 and 1.

### III. METHODOLOGY

Six inputs: daily minimum temperature (℃), maximum temperature (℃), mean temperature (℃), mean relative humidity (%), mean wind speed (m/s), mean sea level pressure (hPa) were used to train and test the network in classify the rainfall event. The networks were trained using 59 months (ranging from 1st Jan 2009 to 30th Nov 2013) and test with the data of 1 month (ranging from 1st Dec 2013 to 31st Dec 2013).

Artificial neural network (ANN) model in this study learns through a set of examples. During the learning process, the ANN model will build a model through its weights to generalize and predict the outputs from inputs that was previously supplied through its inputs. The architecture of the ANN model includes the input, hidden and output layers. The ANN model "learns" by adjusting the weights between the

nodes. At the end of the training process, the ANN model should be able to predict a target value from a given input value [10]. For the ANN model to learn, the data was divided into: training, validation and testing sets through random sampling. The "trained" ANN model is considered to be able to obtain an acceptable error when it is used for other similar situation [9]. This is verified through the testing set. The validation set was used to verify that the ANN model is producing acceptable error during the training stage.

### A. BPN Model

According to [11], single hidden layer feedforward network can approximate any measureable function arbitrarily well regardless of the activation function, the dimension of the input space and the input space environment. Therefore, for the purpose of this study, one hidden layer backpropagation neural network is used. The Levenberg-Marquardt back propagation learning algorithm in Matlab software was used. The architecture of the BPN used was 6 inputs, 10 hidden neurons and 1 output as shown in Fig 1.
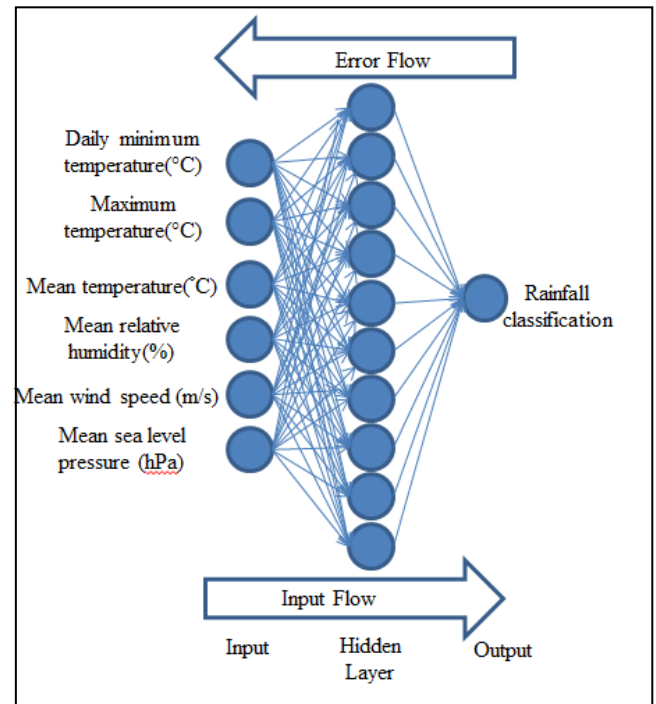


Fig 1. Architecture of the BPN model used.

Backpropagation algorithm is a supervised learning algorithm which adapts the synopsis weights, in order to minimize the Mean Square Error (MSE) between the actual and predicted output from the input vector. As shown in Fig. 1, the algorithm includes a forward and backward phase. In the forward phase, the input vector travels layer to layer from input to output, generating certain response (output). During the backward phase, the error signals travelled from output layer to the input layer. Adjustment of the network parameters is done during the backward phase in order to minimize the MSE values.

### B. RBFN Model

Radial functions are class of functions which could be applied in any sort of model (linear or non-linear) and any sort of network (single- or multi-layer) [12]. According to [13],

the single-weight layer network (the input component is feed-forward to the basis functions whose outputs are linearly combined with weights into the network outputs) is associated with the traditional RBFN model. It is proven that RBFN with one hidden layer can approximate any function; as a result they are called universal approximators. Therefore, in this study, the single-weight layer RBFN model with Gaussian learning algorithm is used (Fig. 2). In the RBFN model, the input units distribute the values to the hidden layer without multiplying them with weights, i.e., the input units were distribute uniformly to the hidden layer.
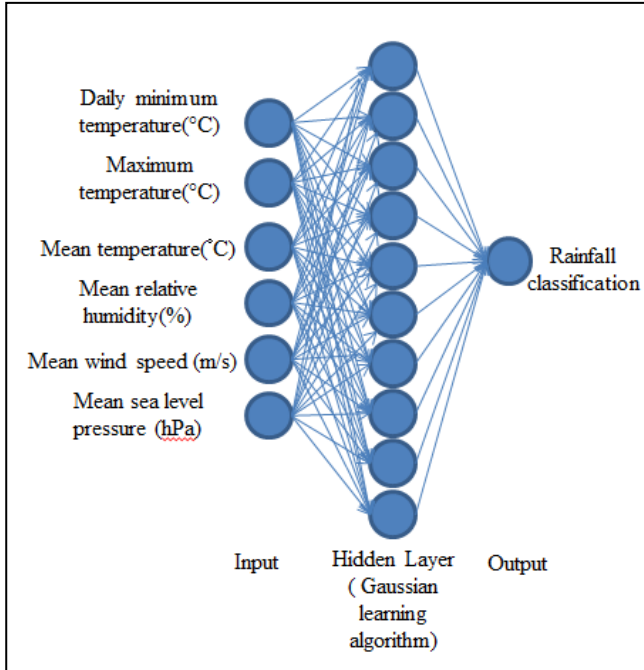


Fig. 2. Architecture of RBFN used.

Both BPN and RBFN models are trained with supervised learning algorithms. The characteristics of the BPN and RBFN models are mostly similar to the feed forward neural network whereby both of the models perform linear representations and weights summations. However, for RBFN, faster training could be achieved due to the local transformation performed in the RBFN model.

## IV. RESULTS AND ANALYSIS

TABLE III row 1 shows the results obtained using 59 months data for training and 1 month unseen data for testing. The results obtained show that BPN with 10 hidden neuron predicted rainfall with higher accuracy (MSE=0.16) and closer relationships (R=0.86) comparing to RBN which produced the accuracy with MSE=0.22 and relationship between the actual and predicted data, R=0.82.

From the results, BPN performs better comparing to RBFN for rainfall classification. However, it was noticed that, during the training and testing processes of both networks, BPN produced results which were not consistent and required repeated training and validation in order to obtain a better accuracy. On the other hand, RBFN was better in generating consistent results. The inconsistent results observed in BPN are due to random assignment of the weights in the network as comparing to RBFN which only involves the adjustment of the "spread" of the network. During each of the BPN training,

a random weight is set up and assigned into the network to adjust the network learning rate. Due to the random assignation of random weight, the network produces random training performance, thus repeated training of the network model is required in order to produce a decent result. During the RBFN training, by adjusting the 'spread' manually in the specification of network properties, the network training and testing performance can be optimized. Therefore, the network produces a consistence result throughout the training process. Consistence results here are referred as outputs which were similar or the same being generated each time when the network is tested.

TABLE III: MSE AND R OBTAINED FROM BPN AND RBFN USING 50 AND 100 HIDDEN NEURONS

| No. of hidden neurons | BPN | | RBFN | |
|---|---|---|---|---|
| | MSE | R | MSE | R |
| 10 | 0.1647 | 0.8575 | 0.2206 | 0.8191 |
| 50 | 0.2065 | 0.8399 | 0.2473 | 0.7763 |
| 100 | 0.2898 | 0.7366 | 0.3035 | 0.7048 |

To verify that the number of hidden neurons used is sufficient for both the ANN models to generalize effectively from the data, a series of experiments with different hidden neurons were carried out for. For each of the BPN and RBFN model, the hidden neurons were set 50 and 100. The numbers of data used for training and testing were the same as used for 10 hidden neurons (i.e. 59 months, ranging from 1st Jan 2009 to 30th Nov 2013 and test with the data of 1 month, ranging from 1st Dec 2013 to 31st Dec 2013). For each of the ANN model with the hidden neuron of 10, 50 and 100, the network was trained and tested 15 times. The MSE obtained shows that, the increase of hidden neuron in BPN caused the accuracy obtained to deteriorated (Fig. 3. (a)). In the case of RBFN, there is no effect on the use of different number of hidden neurons on the accuracy obtained and the results was consistent (Fig. 3.(b)). TABLE III shows that, the average MSE values obtained from 50 and 100 hidden neurons were 0.2065 and 0.2898 respectively for the BPN model and for the RBFN model, the values were 0.2473 and 0.3035. In term of relationships between the output obtained and the actual value of testing data, it was found that, the correlation coefficient, R deteriorate as well (Table III). Therefore, it could be concluded that, the 10 hidden neurons are sufficient for the BPN to "learn" from the training data.

## V. CONCLUSION

From the study, it could be concluded that, BPN could classify the rainfall better as comparing to the RBFN model. With the use of 10 hidden neurons for both models, it could be seen that, the BPN could achieve an accuracy of around 33.4% better than the RBFN model. However, in order to obtain a good accuracy in BPN, repeated training and validation using the training data set is needed. This is different with RBFN as this network model would produce consistent result by adjusting the "spread" manually in the network properties. Moreover, as explained earlier, the transformation in the RBFN is performed locally and therefore, faster training could be obtained.

From the results obtained in Table III, it could be seen that, the accuracy obtained by both of the models were very closed when the number of hidden neurons increased. Yet, this is not a good choice, as there is a waste of resources in terms of computation powers when more hidden neurons are used.

This research work is beneficial for researchers who are working on backpropagation neural network or radial basis function network as the results in this paper shows the benefits and weaknesses for each of the model using same set of data.
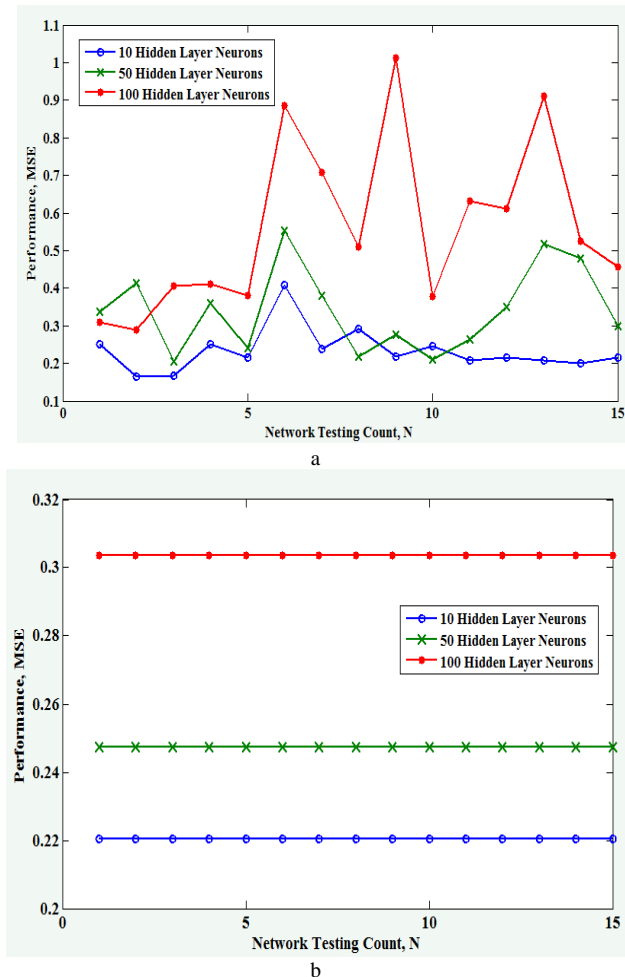


a



b

Fig. 3. MSE obtained using 10, 50 and 100 hidden neurons for (a) BPN model; (b) RBFN model.

REFERENCES

[1] S. Doocy *et al*., "The human impact of floods: A historical review of events 1980–2009 and systematic literature review PLOS Curr," *Disasters*, 2013.

[2] K. C. Luk, J. E. Ball and A. Sharma, "An application of artificial neural networks for rainfall forecasting," *Mathematical and Computer modelling*, vol. 33, no. 6, pp. 683-693, 2001.

[3] D. R. Nayak, A. Mahapatra and P. Mishra, "A survey on rainfall prediction using artificial neural network," *International Journal of Computer Applications*, vol. 72, no. 16, 2013.

[4] G. Shrivastava *et al*., "Application of artificial neural networks in weather forecasting: A comprehensive literature review," *International Journal of Computer Applications*, vol. 51, no. 18, 2012.

[5] *Demographics of Sarawak - Wikipedia, the Free Encyclopedia.* [Online]. Available: http://en.wikipedia.org/wiki/Demographics_of_Sarawak

[6] R. Suzana, T. Wardah and A. Hamid, "Radar hydrology: New Z/R relationships for Klang River Basin Malaysia based on rainfall classification," *World. Acad. Sci. Eng. Technol*, 2011.

[7] J. Sola and J. Sevilla, "Importance of input data normalization for the application of neural networks to complex industrial problems," *IEEE Transactions on Nuclear Science*, vol. 44, no. 3, pp. 1464-1468, 1997.

[8] S. Chen, Y. Wang, and I. Tsou, "Using artificial neural network approach for modelling rainfall–runoff due to typhoon," *Journal of Earth System Science*, vol. 122, no. 2, pp. 399-405, 2013.

[9] S.-S. Chai *et al*., "Use of soil moisture variability in artificial neural network retrieval of soil moisture," *Remote Sensing*, vol. 2, no. 1, pp. 166-190, 2009.

[10] B. Pradhan and S. Lee, "Regional landslide susceptibility analysis using back-propagation neural network model at Cameron Highland, Malaysia," *Landslides*, vol. 7, no. 1, pp. 13-30, 2010.

[11] R. Fierro and F. Lewis, "Multilayer Feedforward Networks are Universal Approximators," *IEEE Trans. Syst., Man, Cybern*, vol. 29, no. 6, pp. 649-654, 1999.

[12] M. J. Orr, *Introduction to Radial Basis Function Networks*, 1996, Technical Report, Center for Cognitive Science, University of Edinburgh.

[13] D. S. Broomhead and D. Lowe, "Radial basis functions, multi-variable functional interpolation and adaptive networks," 1988, DTIC Document.

**Soo See Chai** obtained her PhD (spatial science) from Curtin University of Technology, Western Australia. She is currently the senior lecturer in the Faculty of Computer Science and Information Technology (FCSIT), University of Malaysia Sarawak (UNIMAS). Her research interests include Geographic Information System (GIS), remote sensing, digital image processing and artificial intelligence.

**Wei Keat Wong** obtained his bachelor degree in computer science and information technology, majoring in software engineering from the Faculty of Computer Science and Information Technology (FCSIT), UNIMAS. Currently, he is working as validation engineer in NCS Global Technology Sdn Bhd, Penang, Malaysia.

**Kok Luong Goh** obtained his master of science in information technology from the FCSIT, UNIMAS. He is currently working as a lecturer in International College of Advaned Technology Sarawak (ICATS), Kuching.