P-ISSN 2586-9000
E-ISSN 2586-9027
Homepage : https://tci-thaijo.org/index.php/SciTechAsia
Science & Technology Asia
Vol.25 No.4 October - December 2020
Page: [30-44]
Original research article

Improving Monthly Rainfall Forecast Model by Input Selection Technique using Deep Neural Network

Shekhar Mahat, Krisada Chaiyasarn, Uruya Weesakul*

Department of Civil Engineering, Faculty of Engineering, Thammasat University, Pathum Thani 12120, Thailand

Received 22 August 2019; Received in revised form 24 October 2019
Accepted 19 November 2019; Available online 24 December 2020

ABSTRACT

Long term monthly rainfall forecasting is essential for appropriate river basin planning and management. A number of rainfalls forecast models are developed using a variety of approaches. It is, however, difficult to forecast rainfall with high accuracy particularly for long time ahead due to its variations in both time and space. A recent monthly rainfall forecast model using Deep Neural Network was developed (MRDNN Model); however, accuracy of forecast is acceptable for one-month leading time of forecast. This study aims to improve accuracy of forecast of such model by using DNN coupling with technique of selection of most predictive variables. A river basin in eastern region of Thailand was selected as study area. Monthly rainfall data at Pluak Daeng station from 1991 to 2010 were used for training in the DNN model, while time series from 2011 to 2016 were used for model validation. Stochastic efficiency was used as criteria to evaluate accuracy of forecast. By using input selection technique, Large Scale Atmospheric Variables (LAV) as predictors in the model are reduced from 32 variables to be 8 variables, i.e., Air Temperature, Sea Surface Temperature, Precipitable Water, Specific Humidity and Outgoing Longwave Radiation, providing more accuracy of forecast with 72% of stochastic efficiency for one-month lead time. However, the proposed DNN monthly rainfall forecast model can provide an acceptable accuracy of forecast about 70% up to for 3-months leading time of forecast. Further improvement of model will be studied in order to obtain more accuracy of forecast longer leading time of forecast.

Keywords: Deep neural network; Monthly rainfall forecasting; Input selection technique; Large scale atmospheric variables; Thailand

***Corresponding author:** wuruya@engr.tu.ac.th

1. Introduction

Accurate rainfall forecast, as essential information for efficient watershed management and planning, is still one of the tedious tasks in hydrological cycle due to nonlinear nature both in space and time [1, 2]. A variety of computational techniques were adopted in rainfall forecast models in order to enhance the accuracy of forecast in last three decades. Particularly in the last decade, more sophisticated models were developed to increase their capability to capture uncertainty due to high variation of rainfall in both time and space as an impact of climate change. Over the time, traditional black box models using statistical approaches such as multiple regression models, Autoregressive Moving Average (ARMA) and Autoregressive Integral Moving Average (ARIMA) were used for rainfall forecast. For example, ARIMA model was applied to forecast monthly rainfall in Indiana, Illinois and Kentucky and found that accuracy of forecast is acceptable [3]. The ARIMA and ARMA model was used in 31 rainfall stations to forecast rainfall for agriculture water allocation planning in Thailand. The outcomes of the study reveal that relative error from 8% to 27% achieved which was acceptable [4]. However, prime drawbacks of conventional methods were that these models act as linear in nature and had limited ability to capture non stationaries and non-linearity in hydrology [5].

Another type of black box model, like Artificial Neural Network (ANN) model, was widely applied for rainfall forecast in last two decades. For example, the ANN was applied to forecast short-term rainfall in Italy. Comparison between ANN and ARMA model in forecasting of short-term rainfall for real time flood forecasting were conducted in the study. The results of the study revealed that ANN model provided more accurate forecasted rainfall than ARMA model, particularly for forecasting of rainfall with leading time form 1-6 hours [6]. ANN model was used to forecast annual rainfall in Chao Phraya river basin, located in the central part of Thailand. The result of model simulation showed model capability in forecasting of annual rainfall in this region with excellent accuracy of 96% for leading time of one-year [7]. The ANN and linear regression model were compared for daily rainfall forecast in Sao Paulo State, Brazil. Based on the results of model simulation, ANN performed better than linear regression model [8]. The ANN model was also applied for rainfall forecast in turkey. Comparison of forecasted rainfall between ANN model and multiple regression model indicated that ANN provided more accuracy in forecast [9]. Performance of ANN model was investigated for long-term rainfall forecast using large scale climate models. The study compared performance of ANN model with Multiple regression model and found that ANN forecast better over the multi regression. The reason behind this hypothetical theory was that multiple regression analysis was a linear in nature while an ANN is a nonlinear [10]. Capability of ANN was explored in prediction of monthly rainfall for four months leading time in Athens. The result of study revealed that ANN had low ability of prediction in the case of maximum rain intensity [11]. The ANN was applied for flood forecasting at Dangola station in Nile River. The result of model simulations revealed that ANN was able to predict the flood forecasting for Dongola station of 1998 August and September flood in satisfactorily level [12].

The ANN was implemented for monthly rainfall forecast in Queensland, Australia. Climatic indices, monthly rainfall, atmospheric temperature, and solar radiation were used in the model to forecast monthly rainfall with 3-months leading time for 20 rainfall stations. The results revealed that ANN predicted lower root mean square error than Australian Bureau of

Meteorology's Predictive Ocean Atmosphere Model for Australia [13]. Later such ANN model was developed using input selection and optimization technique to forecast monthly rainfall in Queensland, Australia. It was observed in the study that most of the predictors in the model had high variations both in space and time. It was summarized that there was no optimal set of variables for three sites. Moreover, the ANN showed better predictive result than that of climatology to forecast monthly rainfall of 1, 2 and 3-months lead time [14]. Recently, ANN model was adopted with single month optimization technique to improve accuracy of monthly rainfall in Brisbane catchment, Queensland, Australia. The finding of the study reveals that single month optimisation technique outperforms prediction model of Predictive Atmospheric Model for Australia which is used by the Australian Bureau of Meteorology. In the study, the coefficient of correlation with 0.85 was achieved for 12-month lead time [15].

The ANN was implemented for forecasting seasonal rainfall using six ocean indices in Chao Phraya River Basin. The finding of the simulation revealed that ANN model forecasted seasonal rainfall with RMSE of 0.092 and correlation coefficient of 0. 917 for 12-months lead time. Moreover, the result showed that ANN performed quiet well to forecast rainfall even in dry season [16]. The ANN with backpropagation algorithm having two hidden layers was explored to forecast monthly rainfall in Tenggarong, East Kalimantan, Indonesia. Accuracy of the forecasting rainfall revealed that mean square error of 0.00096341 was observed for monthly rainfall forecast. It indicated that ANNs was capable to forecast rainfall for rainy season. Moreover, the ANN with backpropagation had capability as a predictive algorithm to check accuracy of forecast [17].

The ANN was applied to forecast rainfall in eastern river basin of Thailand. The main objective of the study was to test the capability of ANN model to rainfall forecast in Pluak Daeng station. The findings indicated that ANN had capability to forecast rainfall with a lead period of 3-months [18].

From the past research results, it indicated accuracy and efficiency of rainfall forecasting need to be improve using different rainfall forecast models. Deep learning techniques, for example, Recurrent Neural Network (RNN) model [19], Long Short-Term Memory (LSTM) models [20-24], Deep learning and Convolutional Neural Network model [25] can be used as alternative techniques to enhance performance and efficiency of rainfall model.

Interestingly, Deep Neural Network (DNN) had been widely used in multidisciplinary field of science such as image classification [26], object detection [27], natural language processing [28], speech recognition [29], noise robust speech recognition [30], medical class images [31], and bioinformatics [32]. The major reason of its popularity was the nature to solve the complex types of problem in real life. With the hidden layers, hidden layers compute the output features with a set of filters. It makes DNN more powerful computing techniques than simple neural network.

Recently, the RNN was performed to check the capability for forecast Indian summer monsoon. Finding of result suggested that RNN model outperformed the existing India Meteorology Department models with mean absolute error of 3.3% [33].

Artificial intelligence approaches were applied for rainfall prediction in Kerala, India. The extreme learning machine was compared with ANN and K-nearest neighbor approaches in prediction of rainfall. The extreme learning machine provided better performance in forecasting of summer monsoon and post monsoon rainfall than the other algorithms because of the numbers of hidden nodes in the hidden layer [34]. Recently, capability of the DNN in capturing uncertainty of monthly rainfall was investigated in eastern river basin in Thailand. Large atmospheric variables (LAV) such as air temperature, geopotential height, meridional wind, omega, outgoing longwave radiation, relative humidity, specific humidity, sea level pressure, sea surface temperature, zonal wind, precipitation rate and precipitable water in different atmospheric layers were used as predictors in the DNN model [35] due to their high correlation with seasonal rainfall in Thailand [36-43]. Result of forecasting revealed that DNN model was able to predict monthly rainfall from one-month to twelve months ahead; however, accuracy of forecast decreased when leading time of forecast increased. The most practical time of forecast was one month-ahead with the efficiency of forecast around 70% that the forecasted values were within the range of standard deviation of the observed ones based on DNN model [35].

This study aims to improve accuracy of forecasting by using input selection technique with DNN model to predict monthly rainfall.

2. Study Area and Data Collection

The Khlong Yai river basin, with total drainage area of about 1,800 kilometers square, was selected as study area, as shown in Fig. 1. The most efficient water resources management and planning in this region is needed due to further economic development plan in the region. Consequently, an accurate long-term monthly rainfall forecast is necessary for appropriate water management and water allocation in the area. Based on availability and reliability of rainfall data in the area, Pluak Daeng station [478004] was selected as representative rainfall in the study area, as shown in Fig. 1. Monthly rainfall data from Pluak Daeng station from 1991 to 2016 was collected from Thai Meteorological Department for analysis in this study. Rainfall in the area is influenced by tropical monsoon, beginning from May and ending in October with some tropical depressions. The average annual rainfall is about 1200 mm.

It has been revealed, from the previous studies [36-39], that seasonal rainfall in different parts of Thailand was influenced by several large atmospheric variables (LAV).

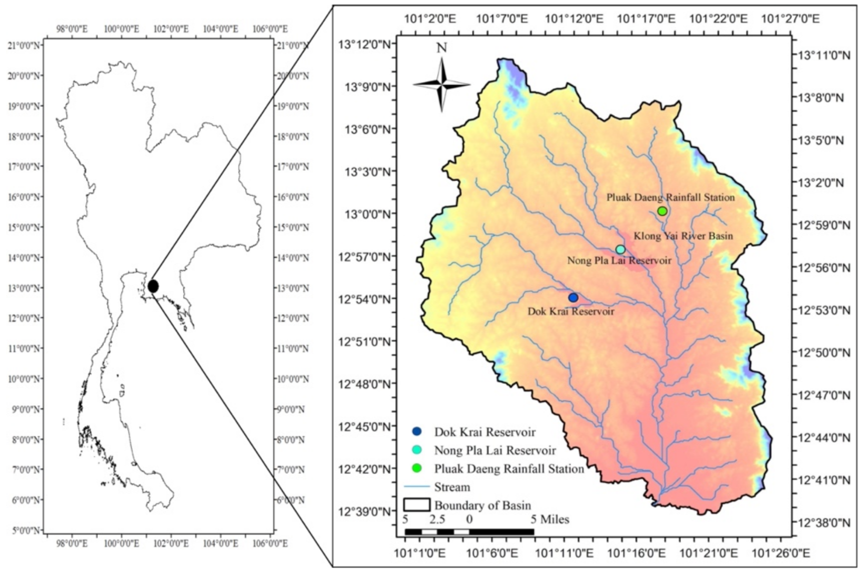


Fig. 1. Study area and location of Pluak Daeng rainfall station.

Table 1. Large Atmospheric Variables (LAV) used in the study.

Variable abbreviation	Variable Name	Height Based on atmospheric pressure (mili bar)	Latitude	Longitude
AT1	Air Temperature	400	25° to 30° N	175° to 180° E
AT2	Air Temperature	400	0° to 5° S	210° to 218° E
AT3	Air Temperature	850	12° 30' to 17° 30' N	157° 30' to 167° 30' E
AT4	Air Temperature	2000	3° S to 3° N	250° to 260° E
AT5	Air Temperature	50	9° to 14° N	110° to 120° E
GH1	Geopotential Height	250	7° to 12° N	180° to 185° E
GH2	Geopotential Height	600	0° to 5° N	135° to 140° E
GH3	Geopotential Height	850	0° to 5° N	240° to 250° E
MW1	Meridional Wind	10	12° to 20° N	180° to 190° E
MW2	Meridional Wind	150	10° to 15° N	207° 30' to 212° 30' E
MW3	Meridional Wind	200	0° to 5° S	155° to 160° E
O1	Omega	500	11° to 16° N	207° 30' to 212° 30' E
O2	Omega	925	5° to 10° N	155° to 160° E
OLR1	Outgoing Longwave Radiation	2000	0° to 5° S	255° to 260° E
P1	Pressure	2000	0° to 5° S	220° to 225° E
PR1	Precipitation Rate	2000	2° S to 3° N	255° to 260° E
PR2	Precipitation Rate	2000	0° to 5° N	202° to 207° E
PW1	Precipitable Water	2000	5° to 10° S	180° to 190° E
PW2	Precipitable Water	2000	7° 30' S to 2° 30' N	105° to 115° E
RH1	Relative Humidity 28up to 300mb only	925	2° to 7° N	135° to 140° E
RH2	Relative Humidity 28up to 300mb only	400	5° to 10° N	202° to 207° E
SH1	Specific Humidity 28up to 300mb only	500	12° 30' to 17° 30' N	205° to 210° E
SH2	Specific Humidity 28up to 300mb only	500	0° to 5° N	180° to 185° E
SH3	Specific Humidity 28up to 300mb only	400	7° to 15° N	120° to 130° E
SH4	Specific Humidity 28up to 300mb only	400	0° to 5° S	110° to 120° E
SH5	Specific Humidity 28up to 300mb only	400	0° to 5° N	110° to 115° E
SH6	Specific Humidity 28up to 300mb only	600	2° 30' S to 2° 30' N	217° 30' to 222° 30' E
SLP1	Sea Level Pressure	2000	2° to 7° S	255° to 260° E
SLP2	Sea Level Pressure	2000	0° to 5° S	207° 30' to 212° 30' E
SST1	Sea Surface Temperature	2000	0° to 5° S	255° to 260° E
ZW1	Zonal Wind	30	10° to 16° N	85° to 95° E
Irf	Monthly Rainfall at Pluak Daeng Station (mm)		13° N	101° 18' E

These LAV were used as predictors in modified K-nearest neighbor for forecasting seasonal rainfall in northeastern region of Thailand [42]. The LAV with different layers were also tested as predictors in modified K-nearest neighbor Model for seasonal rainfall forecast for cropping pattern planning in Thailand [43]. Recently, variety of LAV in different layers were examined to be predictors of monthly rainfall in this study area using DNN model.

In this study LAV in different atmospheric layers, as shown in Table 1, were used as input of the model. LAV data were retrieved from Earth System Research Laboratory (ESRL) which provides monthly LAV data with a grid cell of 2.5° latitude x 2.5° longitude covering the area between longitude 65° E to 160° E and latitude 20° E to 20° W. Each LAV computes various atmosphere layers with different height from the ground, as shown in Table 1.

3. Rainfall Forecast Model Coupling with Input Selection Technique

3.1 Formulation of monthly rainfall forecast model

In this study, DNN was adopted as computational tool in monthly rainfall forecast model. In general, neural network consists of three layers: input layer, hidden layer and output layer. Input layer are feed from external data and output layer feed data to external destination. The difference between Artificial Neural Network and Deep Neural Network is that ANN has single hidden layer whereas DNN had multiple (much deeper) hidden layers. In the study, five hidden layers are proposed for the implementation of DNN model. Two major calculation behind Neural Network were used to develop DNN model which are firstly Feed-Forward Propagation and secondly Back Propagation as follows:

Feed-Forward Propagation

At first, multiplication of each input node with random weight takes place. The sum of input node and random weight with bias is passed through the activation function as presented in Eq. (3. 1). In general, input layer receives the input, followed by preprocessing of input is done in hidden layer, and finally output is obtained in output layer. Then an activation function in the form of sigmoid function is applied on the hidden layer, represented in Eq. (3.2). Finally, error is calculated after an activation function is applied, as illustrated in Eq. (3.3).

$$h_j = \sigma\left(z_{net_j}\right) = \sigma\left(v_j + \sum_{i=1}^{N} X_{ij} + W_{ij}\right), \quad (3.1)$$

where σ refers an activation function, z_{net_j} indicates network input node of sigmoid function on hidden node j, v_j is bias on hidden node j, X_{ij} is input of i value on hidden node j, W_{ij} denotes weight of i value on hidden node j and h_j is output on hidden node j.

The output of sigmoid function on hidden node j is given as in Eq. (3.2).

$$Z_j = \sigma\left(Z_{net_j}\right) = \frac{1}{1+\left(e^{-Z_{net\,j}}\right)}, \quad (3.2)$$

Calculation of error is given by Eq. (3.3).

$$\mathrm{E} = \frac{1}{2}\left(O-P\right)^2, \quad (3.3)$$

where O is observed output and P is target output.

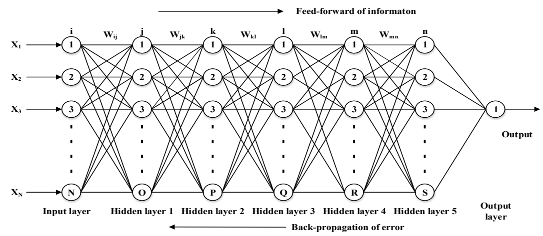


Fig. 2. Schematic representation of DNN with five hidden layers [35].

Backward Propagation

In back propagation, once error is calculated between observed and predicted output, error is back propagated to change the weight in hidden layer until and unless error becomes minimal as illustrated in Eq. (3. 4). After that weight is updated in the hidden layer, as shown in Eq. (3. 5), and again updated weight is passed in the output layer. This process is repeated until the error becomes minimal to target value in output layer. The schematic diagram of DNN used for the study is shown in Fig. 2. The back-propagation method is represented in the following equation.

$$\begin{aligned}\delta_{W_{ij}} &= \left(O-P\right)h_j^{'} \\ &= \left(O-P\right)Z_{net_j}\left(1-Z_{net_j}\right)Z_{net_j}^{'}, \end{aligned} \quad (3.4)$$

where $\delta_{W_{ij}}$ represents weight change of i value at hidden node j, $h_j^{'}$ is called derivative of activation function at hidden node j and $z_{net_j}^{'}$ is derivative of network at node j.

In the next step, the weight is updated as in Eq. (3.5).

$$W_{newij} = W_{oldij} + \delta_{W_{ij}}, \qquad (3.5)$$

where W_{newij} is new weight of i value at hidden node j and W_{oldij} is old weight of i value at hidden node j.

Deciding the number of hidden nodes and hidden layers is very important part of DNN model architecture. Rule of thumb was used to determine the suitable number of hidden nodes as following: The number of hidden neurons should be between the size of size of input layer and output layer, 2/3 the size of input layer plus size of output layer and less than twice the size of input layer [44]. Similarly, capabilities of several hidden layer were summarized. One hidden layer can approximate function that contains a continuous mapping. More than two hidden layers can learn complex representations [45].

3.2 Development of input selection technique

There are 31 LAV including one monthly rainfall as 32 variables input to DNN monthly rainfall forecast model. This study proposed input selection procedure allowing defining only most predictable variables yielding more accurate results of forecast. Stochastic efficiency is used to evaluate efficiency of model in forecasting monthly forecast, which is defined in this study as follows in Eq. (3.6).

$$\text{Stochastic efficiency} = \frac{P}{N} * 100\%, \qquad (3.6)$$

where P is number of forecasted monthly rainfalls being in between mean monthly rainfall plus and minus standard deviation of monthly rainfall, which can be represented in Eq. (3.7).

$$RO_{i,j} - \sigma_i \leq RP_{i,j} \leq RO_{i,j} + \sigma_i, \qquad (3.7)$$

where $RP_{i,j}$ is forecasted monthly rainfall of month i in year j, $RO_{i,j}$ is observed monthly rainfall of month i in year j, σ_i is standard deviation of monthly rainfall in month i (i = 1…12) and N is total number of forecasted monthly rainfalls during validation period.

In this study, due to availability of data, monthly rainfall during 1991 to 2010 was used for training process and monthly rainfall during 2011-2016 was used in validation process. More simulation in other river basin should be conducted further to explain behaviour of model clearly for all situations. Simple input selection procedure was conducted firstly to rank most predictive variables. Each variable, amongst 32 LAV as shown in Table 1, was tested as input to the DNN model one by one, using rainfall during 1991-2010 in training process. Then rainfall during 2011-2016 was used for validation process. Forecasted monthly rainfall from 2011-2016 were evaluated by the accuracy of forecast using stochastic efficiency as shown in Eq. (3.6).

Table 2 represents result of simulation showing influence of each LAV on rainfall forecasting. The result reveal that only half of LAV has sufficiency influence on predicting monthly rainfall, i.e., only 16 has stochastic efficiency more than 50%. The predictive LAV were chosen based on their performance when their capability of prediction was correct more than 50%. Then only LAV that provide more than 55% of stochastic efficiency of forecast were selected as criteria to be input of DNN model in the proposed study. The combination of these selected LAV are used as input to DNN model as shown in Table 3. Finally, 8 models are configured and are tested based on stochastic efficiency of forecasted monthly rainfall to define best combination of input variables.

This study proposed input selection procedure allowing defining only most predictable variables yielding more accurate results of forecast.

Stochastic efficiency is used to evaluate efficiency of model in forecasting monthly forecast, which is defined in this study as follows:

$$\text{Stochastic efficiency} = \frac{P}{N} \times 100\%, \quad (3.6)$$

where P is number of forecasted monthly rainfalls being in between mean monthly rainfall plus and minus standard deviation of monthly rainfall, which can be represented in the following equation:

$$RO_{i,j} - \sigma_i \le RP_{i,j} \le RO_{i,j} + \sigma_i, \quad (3.7)$$

where $RP_{i,j}$ is forecasted monthly rainfall of month i in year j, $RO_{i,j}$ is observed monthly rainfall of month i in year j, σ_i is standard deviation of monthly rainfall in month i (i = 1…12), and N is total number of forecasted monthly rainfalls during validation period.

In this study, due to availability of data, monthly rainfall during 1991 to 2010 was used for training process and monthly rainfall during 2011-2016 was used in validation process. More simulation in other river basin should be conducted further to explain behavior of model clearly for all situations. Simple input selection procedure was conducted firstly to rank most predictive variables. Each variable, amongst 32 LAV as shown in Table 1, was tested as input to the DNN model one by one, using rainfall during 1991-2010 in training process. Then rainfall during 2011-2016 was used for validation process. Forecasted monthly rainfall from 2011-2016 were evaluated by the accuracy of forecast using stochastic efficiency as shown in Eq. (3.6).

Table 2 represents result of simulation showing influence of each LAV on rainfall forecasting. The result reveal that only half of LAV has sufficiency influence on predicting monthly rainfall, i.e. only 16 has stochastic efficiency more than 50% . The predictive LAV were chosen based on their performance when their capability of prediction was correct more than 50%. Then only LAV that provide more than 55% of stochastic efficiency of forecast were selected as criteria to be input of DNN model in the proposed study. The combination of these selected LAV are used as input to DNN model as shown in Table 3. Finally, 8 models are configured and are tested based on stochastic efficiency of forecasted monthly rainfall to define best combination of input variables.

Table 2. Stochastic efficiency of DNN model using one by one LAV as input for forecasting monthly rainfall with 1-month leading time.

LAV	Stochastic Efficiency (%)
AT1	62.50
PW1	59.72
OLR1	59.72
AT4	58.33
PW2	56.94
SH3	56.94
SH4	55.56
SST1	55.56
MW3	54.17
PR1	54.17
MW2	52.78
AT3	52.78
GH3	52.78
ZW1	52.78
Irf	51.39
RH1	51.39
O1	50.00
RH2	50.00
SH6	50.00
SH2	45.83
MW1	43.06
SH1	41.67
O2	41.67
AT2	40.28
SH5	40.28
GH1	38.89
AT5	36.11
GH2	36.11
P1	36.11
PR2	36.11
SLP1	36.11
SLP2	36.11

4. Monthly Rainfall Forecasting Performance by DNN Model

To design DNN model architecture, the number of hidden nodes and hidden layers (as shown in Fig. 2) has to be determined. In this study, simulation of DNN model with variety set of hidden nodes and hidden layers was conducted using available data set from 1991 to 2016. The best test of simulation, based on stochastic efficiency criteria, was selected as best DNN model architecture. Table 3 presents selected DNN model architecture for 8 models. The selected 8 models were tested their performance in forecasting monthly rainfall with different leading time of forecast, i.e., 1-month, 3-months, 6-months and 12-months ahead, using data sets from 2011-2016. Stochastic efficiency was used as performance criteria in selecting the best model.

Table 4 summarizes model performance in forecasting monthly rainfall, which indicates that model 8 is the best one in forecasting monthly rainfall with 1-month and 3-months leading time while model 3 is best in forecasting monthly rainfall with 6-months and 12-months leading time. The result of the simulation reveals that for each type model, accuracy of forecast decreases when leading time of forecast increases, and that models with more LAV input perform better than less LAV input. In this study, simulation of model using combinations of LAV having more than 55% stochastic efficiency provides good results when using 8 LAV.

Comparison between forecasted monthly rainfall by model 8 with 1-month and 3-months leading time and observed monthly rainfall during 2011-2016 were shown in Fig. 3(a) and Fig. 3(b). Correlation coefficients between forecasted and observed rainfall are quite good with value about 0.75. However, it reveals that better accuracy obtains when rainfall is in normal situation (mean monthly rainfall, i.e., around 200 mm).

Table 3. Model developed for Deep Neural Network.

Model	Parameters	Architecture of DNN
Model 1	AT1	[12, 4, 4, 4, 4, 4, 1]
Model 2	AT1, PW1	[24, 8, 8, 8, 8, 8, 1]
Model 3	AT1, PW1, OLR1	[36, 12, 12, 12, 12, 12, 1]
Model 4	AT1, PW1, OLR1, AT4	[48, 16, 16, 16, 16, 16, 1]
Model 5	AT1, PW1, OLR1, AT4, PW2	[60, 20, 20, 20, 20, 20, 1]
Model 6	AT1, PW1, OLR1, AT4, PW2, SH3	[72, 24, 24, 24, 24, 24, 1]
Model 7	AT1, PW1, OLR1, AT4, PW2, SH3, SH4	[84, 28, 28, 28, 28, 28, 1]
Model 8	AT1, PW1, OLR1, AT4, PW2, SH3, SH4, SST1	[96, 32, 32, 32, 32, 32, 1]

For dry season, when rainfall is less than 200 mm, forecasted rainfall is quite scattered. While for wet season, when monthly rainfall is greater than 200 mm, forecasted rainfall is always underestimated to keep less error of prediction of all the range of rainfall. As a result, further development of model is required to adjust the prediction of rainfall over the mean values. Comparison between observed and forecasted monthly rainfall by DNN model 7 with 6-months and 12-months leading time were presented in Fig. 3(c) and Fig. 3(d).

Table 4. Evaluation of DNN model performance in forecasting monthly rainfall with different leading time of forecast.

Model	Stochastic efficiency (%)			
	1-month	3-months	6-months	12-months
1	61.11	59.72	58.33	56.94
2	63.88	58.33	58.33	59.72
3	68.05	69.44	65.28	63.88
4	70.83	69.44	58.33	62.50
5	70.83	69.44	59.72	59.72
6	70.83	66.67	61.11	58.33
7	70.83	69.44	63.89	59.72
8	72.22	69.44	62.50	59.72

Correlation coefficients between forecasted and observed monthly rainfall are 0.74 and 0.73 for 6-months and 12-months leading time of forecast respectively. It has been noticed that model perform well when monthly rainfall is in mean value range and that less performance occurs for both dry and wet season. For dry season, forecasted rainfall is quite scattered while forecasted rainfall is always underestimated for wet season.

Fig. 4 presents comparison between observed and forecasted monthly rainfall by DNN model with 1-month, 3-months, 6-months and 12-months leading time. The outcomes of the simulation reveal that forecasted monthly rainfall often within the mean values plus and minus their standard deviation when it is in wet season.

However, performance of DNN decreases in dry season. In order to verify whether the DNN model with input selection technique proposed in the study perform better than the DNN model without input selection technique or not, result of model simulation in previous study [35] were selected as benchmark.

Table 5. Classification of annual rainfall level during 1991-2010.

Percentile	Amount of annual rainfall (mm)	Level of rainfall
1	311	VERY DRY
10	1008	
11	1012	DRY
30	1051	
31	1057	NORMAL
70	1270	
71	1282	WET
90	1496	
91	1506	VERY WET
100	1608	

Table 6. Comparison of model performance between DNN model with and without input selection technique for 1-month leading time case.

Year	Annual rainfall (mm)	Level of rainfall	Stochastic efficiency (%)	
			Without input selection technique	With input selection technique
2011	1642	VERY WET	75	83
2012	1394	WET	67	67
2013	1350	WET	67	67
2014	1449	WET	67	75
2015	1435	WET	67	75
2016	1087	NORMAL	50	67
Average			65	72

Based on classification of level of annual rainfall conducted [35], annual rainfall was categorized into 5 class by percentile method as shown in Table 5. Annual rainfalls during validation period (2011-2016) were classified as very wet, wet and normal as shown in Table 6.

Performance of DNN model with input selection technique, proposed in the study, was evaluated each year using stochastic efficiency as criteria.

Comparison between performances of DNN model proposed in this study and DNN model without input selection technique was conducted and shown in Table 6. The simulation forecasting of monthly rainfall of 1-month ahead was selected as a case study due to its best performance.

It has been revealed that input selection technique assists the DNN model in improving accuracy of forecast in every case, particularly for very wet season with stochastic efficiency of 83% instead of 75%.

However, further development of the DNN model proposed in the study needs to be conducted particularly for dry season and for longer period of forecast. The input selection technique outperforms because the influential predictors are studied in depth and only most influential parameters are selected for the study. Based on the outcomes, it shows the superior result in forecasting rainfall. Thus, it can be concluded that input selection technique using DNN model can be used to improve the monthly rainfall forecast.

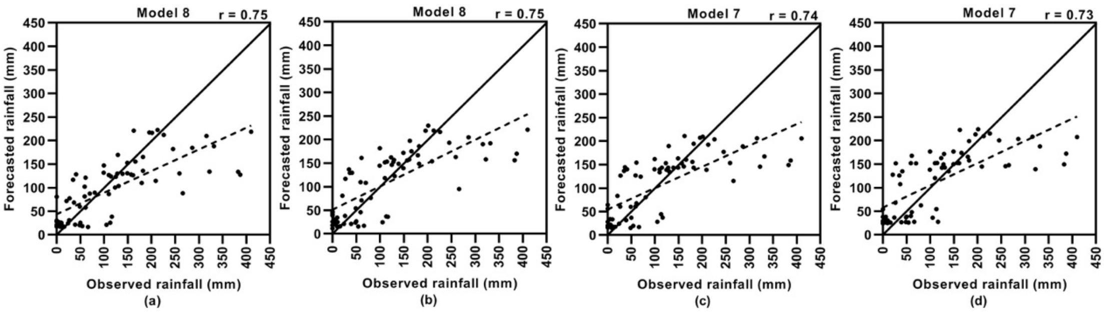
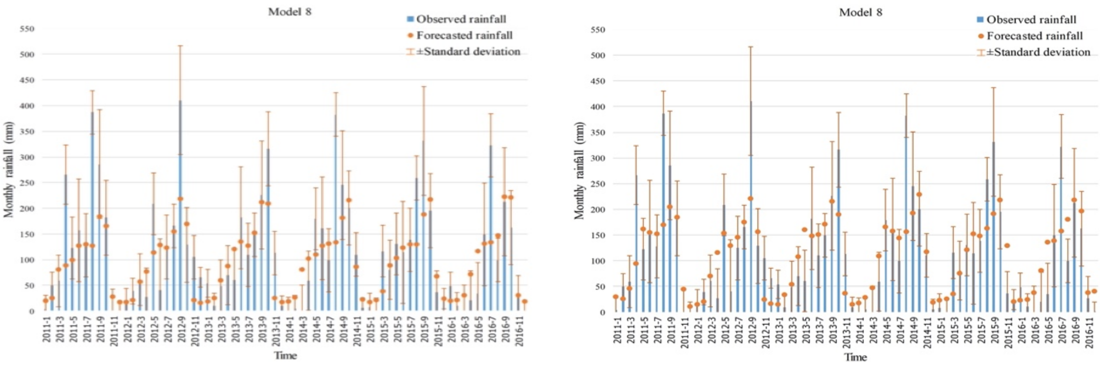
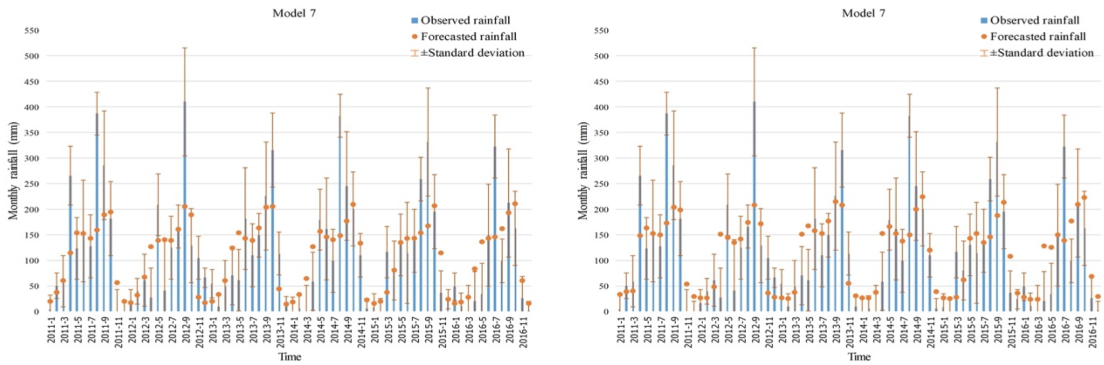


Fig. 3. Comparison of correlation coefficients between observed and forecasted monthly rainfall by DNN model during 2011-2016 (a) one-month lead time, (b) three-month lead time, (c) six-month lead time and (d) one-year lead time.



(a)

(b)



(c) (d)

Fig. 4. Comparison between observed and forecasted monthly rainfall by DNN model during 2011-2016 (a) one-month lead time, (b) three-month lead time, (c) six-month lead time and (d) one-year lead time.

5. Conclusion

An attempt to improve accuracy of forecast monthly rainfall by DNN model using input selection technique was conducted in the study. A river basin in the eastern region of Thailand was selected as study area. Monthly rainfall at Pluak Daeng station from 1991 to 2010 was used for training process in the DNN model, while monthly rainfall from 2011 to 2016 was used for model validation. Stochastic efficiency was used as criteria to evaluate accuracy of forecast. By using input selection technique, large atmospheric variables (LAV) as predictors in the model are reduced from 32 variables to only 8 variables, which are Air Temperature, Sea surface Temperature, Precipitable Water, Specific Humidity, and Outgoing Longwave Radiation, providing more accuracy of forecast.

The best performance of forecast was found during rainy season, particularly for monthly rainfall with average values. More development of the model to improve accuracy of forecast during dry season and very wet season has to be further studied. The DNN proposed in the study is appropriate for practical operation up to 3 months leading time of forecast with accuracy of forecast about 70%. Further study to improve accuracy of forecast for longer period leading time of forecast has to be conducted.

Acknowledgements

This research has been supported by the Faculty of Engineering, Thammasat University. Appreciation is also extended to Royal Irrigation Department and Meteorological department for providing hydrological data used in the study. Acknowledgement is also extended to Google for proving open source machine learning library such as Tensorflow-Regressor Application Programming Ineterface (API).

References

[1] French MN, Krajewski WF, Cuykendall RR. Rainfall forecasting in space and time using a neural network. Journal of hydrology 1992;137(1-4):1-31.

[2] Luk KC, Ball JE, Sharma A. An application of artificial neural networks for rainfall forecasting. Mathematical and Computer modelling 2001;33(6-7):683-93.

[3] Delleur JW, Kavvas ML. Stochastic models for monthly rainfall forecasting and synthetic generation. Journal of

Applied Meteorology 1978;17(10):1528-36.

[4] Weesakul U, Lowanichchai S. Rainfall forecast for agricultural water allocation planning in Thailand. Science & Technology Asia 2005;10(3):18-27.

[5] Salas JD. Applied modeling of hydrologic time series. Water Resources Publication; 1980.

[6] Toth E, Brath A, Montanari A. Comparison of short-term rainfall prediction models for real-time flood forecasting. Journal of hydrology 2000;239(1-4):132-47.

[7] Chantasut N, Charoenjit C, Tanprasert C. Predictive mining of rainfall predictions using artificial neural networks for Chao Phraya River. In: 4th International Conference of the Asian Federation of Information Technology in Agriculture and the 2nd World Congress on Computers in Agriculture and Natural Resources. Bangkok, Thailand; 2004.

[8] Ramirez MC, de Campos Velho HF, Ferreira NJ. Artificial neural network technique for rainfall forecasting applied to the Sao Paulo region. Journal of hydrology 2005;301(1-4):146-62.

[9] Terzi Ö, Çevik E. Rainfall estimation using artificial neural network method. SDU International Journal of Technological Science 2012;4(1).

[10] Mekanik F, Imteaz MA, Gato-Trinidad S, Elmahdi A. Multiple regression and Artificial Neural Network for long-term rainfall forecasting using large scale climate modes. Journal of Hydrology 2013;503:11-21.

[11] Nastos PT, Moustris KP, Larissi IK, Paliatsos AG. Rain intensity forecast using artificial neural networks in Athens, Greece. Atmospheric Research 2013;119:153-60.

[12] Elsafi SH. Artificial neural networks (ANNs) for flood forecasting at Dongola Station in the River Nile, Sudan. Alexandria Engineering Journal 2014;53(3):655-62.

[13] Abbot J, Marohasy J. Application of artificial neural networks to rainfall forecasting in Queensland, Australia. Advances in Atmospheric Sciences 2012;29(4):717-30.

[14] Abbot J, Marohasy J. Input selection and optimisation for monthly rainfall forecasting in Queensland, Australia, using artificial neural networks. Atmospheric Research 2014;138:166-78.

[15] Abbot J, Marohasy J. Improving monthly rainfall forecasts using artificial neural networks and single-month optimisation: a case study of the Brisbane catchment, Queensland, Australia. WIT Transactions on Ecology and the Environment; 2015. 196: p. 3-13.

[16] Rangsiwanichpong P, Ekkawatpanit C, Kositgittiwong D. Seasonal rainfall forecasting in the Chao Phraya River Basin using artificial neural networks. In: Proceeding of the 20th National Convention on Civil Engineering. Chonburi, Thailand; 2015.

[17] Hardwinarto S, Aipassa M. Rainfall monthly prediction based on artificial neural network: A case study in Tenggarong Station, East Kalimantan-Indonesia. Procedia Computer Science 2015;59:142-51.

[18] Thammakul P, Kaewprapha P. Rainfall forecasting using artificial neural networks in eastern river basin of Thailand. In: The 40th Electrical Engineering Conference. Chonburi, Thailand; 2017.

[19] Mikolov T, Karafiát M, Burget L, Černocký J, Khudanpur S. Recurrent neural network based language model. In: Eleventh annual conference of the

international speech communication association. Chiba, Japan; 2010.

[20] Hochreiter S, Schmidhuber J. Long short-term memory. Neural computation 1997;9(8):1735-80.

[21] Cho K, Van Merriënboer B, Gulcehre C, Bahdanau D, Bougares F, Schwenk H, Bengio Y. Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv preprint arXiv:1406.1078; 2014.

[22] Donahue J, Anne Hendricks L, Guadarrama S, Rohrbach M, Venugopalan S, Saenko K, Darrell T. Long-term recurrent convolutional networks for visual recognition and description. In: Proceedings of the IEEE conference on computer vision and pattern recognition. Boston, MA, USA; 2015. p. 2625-2634.

[23] Sutskever I, Vinyals O, Le QV. Sequence to sequence learning with neural networks. In: Advances in neural information processing systems. Montreal, Canada; 2014. p. 3104-3112.

[24] LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. Proceedings of the IEEE 1998; 86(11):2278-324.

[25] Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems. Lake Tahoe, Nevada, USA; 2012. p. 1097-1105.

[26] Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems. Lake Tahoe, Nevada, USA; 2012. p. 1097-1105.

[27] Szegedy C, Toshev A, Erhan D. Deep neural networks for object detection. In: Advances in neural information processing systems. Lake Tahoe, Nevada, USA; 2013. p. 2553-2561.

[28] Conneau A, Schwenk H, Barrault L, Lecun Y. Very deep convolutional networks for natural language processing. arXiv preprint arXiv:1606.01781; 2016.

[29] Graves A, Mohamed AR, Hinton G. Speech recognition with deep recurrent neural networks. In: 2013 IEEE international conference on acoustics, speech and signal processing. Vancouver, BC, Canada; 2013. p. 6645-6649.

[30] Seltzer ML, Yu D, Wang Y. An investigation of deep neural networks for noise robust speech recognition. In: 2013 IEEE international conference on acoustics, speech and signal processing. Vancouver, BC, Canada; 2013. p. 7398-7402.

[31] Tajbakhsh N, Shin JY, Gurudu SR, Hurst RT, Kendall CB, Gotway MB, Liang J. Convolutional neural networks for medical image analysis: Full training or fine tuning? IEEE transactions on medical imaging 2016;35(5):1299-312.

[32] Min S, Lee B, Yoon S. Deep learning in bioinformatics. Briefings in bioinformatics 2017;18(5):851-69.

[33] Saha M, Mitra P. Recurrent neural network based prediction of Indian summer monsoon using global climatic predictors. In: 2016 International Joint Conference on Neural Networks (IJCNN). Vancouver, BC, Canada; 2016. p. 1523-1529.

[34] Dash Y, Mishra SK, Panigrahi BK. Rainfall prediction for the Kerala state of India using artificial intelligence approaches. Computers & Electrical Engineering 2018;70:66-73.

[35] Weesakul U, Kaewprapha P, Boonyuen K, Mark O. Deep learning neural

network: A machine learning approach for monthly rainfall forecast, case study in eastern region of Thailand. Engineering and Applied Science Research 2018;45(3):203-11.

[36] Weesakul U, Singhrattna N, Viwat H. Identification of large-scale atmospheric predictors for rainfall forecasting of Nan River Basin. In: 5th National Water Resources Engineering Symposium. Chiengrai, Thailand; 2013.

[37] Weesakul U, Singhratta N, Luangdilok N. Identification of large-scale atmospheric predictors for rainfall forecasting of Chi River Basin (Thailand). In: 5th National Convention on Water Resources Engineering. Chiengrai, Thailand; 2013.

[38] Weesakul U, Singhratta N, Yodpongpiput P. Statistical relationships between large-scale atmospheric variables and rainfall in Mun River Basin (Thailand). In: 5th National Convention on Water Resources Engineering. Chiengrai, Thailand; 2013.

[39] Weesakul U, Singhrattna N, Khetped N. Large-scale atmospheric variables influencing on rainfall in Tapee River Basin. In: Proceedings of the 5th National Water Resources Engineering Symposium. Chiengrai, Thailand; 2013.

[40] Weesakul U, Yodpongpipu P. Interannual variability of monsoon rainfall in Thailand: an inconsistent effect of climate change over south East Asian region. American Transactions on Engineering and Applied Sciences 2015;4(4):251-60.

[41] Weesakul U, Singhrattna N, Luangdilok N. Effects of climate change on Thailand rainfall variability. In: The 3rd International Engineering Symposium on Engineering, Energy and Environments. Bangkok, Thailand; 2013.

[42] Weesakul U, Singhratta N, Luangdilok N. Rainfall forecast in northeast of thailand using modified k-nearest neighbor. Engineering and Applied Science Research 2014;41(2):253-61.

[43] Weesakul U, Singhratta N. Seasonal rainfall forecast for cropping pattern planning using a modified k-nearest neighbor model. Engineering and Applied Science Research 2016;43(3):156-61.

[44] Panchal G, Ganatra A, Kosta YP, Panchal D. Behaviour analysis of multilayer perceptrons with multiple hidden neurons and hidden layers. International Journal of Computer Theory and Engineering 2011;3(2):332-7.

[45] Hinton GE, Osindero S, Teh YW. A fast learning algorithm for deep belief nets. Neural computation 2006;18(7):1527-54.