Procedings of Technical Sesions, 37 (2021) 01-08
Institute of Physics- SriLanka

Rainfall Forecasting in Sri Lanka Using ANN-Case study for Colombo

T.D.L Gayanga[1], R. D. Hettiarachchi[2]
University of Moratuwa[1], The Open University of Sri Lanka[2]
gayangatdl@gmail.com[1], rdhet@ou.ac.lk[2]

ABSTRACT
This paper presents an Artificial Intelligence approach for rainfall forecasting over Colombo, the commercial capital of Sri Lanka on monthly scale. Feature sets extracted from surface level, large scale climate indices over oceans and continents were used in developing the network. From the available teleconnection indices, a total of ten indices were selected for the present study. This study emphasizes the value of using large-scale climate teleconnections for rainfall forecasting and the significance of Artificial Intelligence approaches like LSTM and ANNs in predicting the uncertain rainfall.

KEYWORDS
rainfall forecasting, large scale climate indices, artificial neural networks (ANNs), Long Short Term– Memory (LSTM)

1 INTRODUCTION
A reliable prediction of Sri Lankan Rainfall (SLR) on a monthly scale is scientifically challenging and important for planning and implementing food production and water management strategies in the country. The continuous changes in global climate and the uneven spatial and temporal distribution of rainfall are causes for severe problems like floods and droughts. Meteorological researchers have used various approaches to study and predict the seasonal and intra-seasonal rainfall. To forecast the rainfall at different spatial and temporal scales, the artificial neural network models used in the past can be broadly classified as empirical and dynamical [1]. This study is concerned with empirical models only. The reasonable success achieved by the empirical approach has motivated persistent exploration of regional/global teleconnections of the monsoon season since Walker's time, which resulted in many predictors as well as a variety of statistical techniques. Because any modeling effort in meteorological forecasts will have to be based on an understanding of the variability of unstructured and noisy climatic past data, because of the variability of weather, ANNs have some special characteristics in this regard to be used. In contrast to conventional modeling approaches, ANNs do not require an in-depth knowledge of driving processes, nor do they require the form of the model to be specified a priori [2]. This is true when modeling various climate variables for forecasting of hydrological variables like rainfall and stream flows. Thus, ANNs are a suitable approach for predicting Sri Lankan rainfall using large scale climate variables as input to the network.
At present, the assessment of the nature and causes of seasonal climate variability is still uncertain. There are still uncertainties associated with local and global climatic variables. For any rainfall prediction model, these are sources of variance in predictability (Kumar et al., 1995). Recently, climatic researchers have studied the influence and the possible relationships between various global climate variables and rainfall. Additionally, they

Procedings of Technical Sesions, 37 (2021) 01-08
Institute of Physics- SriLanka

brought out several regional parameters based on sea-level pressure, temperature, wind fields and sea surface temperature (SST) data from the seas. Although their performance in seasonal forecasting has been encouraging, there is still a large variance in the rainfall unaccounted by the predictors identified so far [3].
Several observational and modeling studies have indicated that the slowly varying surface boundary conditions constitute a major forcing factor on the inter-annual variability of the rainfall. Parameters representing these conditions, global as well as regional, provide a handle for seasonal prediction. No clear pattern or trend has been observed in rainfall in Sri Lanka from previous studies. Some studies observed that the mean rainfall is showing a decreasing trend [4] and some studies identified that the frequency of heavy rainfall events increase in central highlands during the recent period [5]. Analyzing fluctuations in rainfall associated with the four climatic seasons using rainfall data for nearly 130 years (1870-2000) from 15 rainfall stations [6] identified that decrease of rainfall in highlands and increase of rainfall in lowlands in the southwestern sector of Sri Lanka during southwest monsoon season. Rainfall during the Northeast Monsoon season, none of the stations show any significant change. While analyzing rainfall data for more than 100 years (1895-1996) [7] observed that no coherent increase or decrease of rainfall in stations in the wet or dry zones. However, not much research has been reported in literature on long range rainfall forecasting using teleconnection indices in Sri Lanka.

1.1 Teleconnections and Teleconnection patterns

Teleconnections are linkages between weather changes occurring in widely separated regions of the globe.
Repeatedly happening a large-scale pressure and circulation anomalies which extend over vast geographical areas and continuing without any interruption is known as teleconnection patterns. Sometimes they can be appearing for a few sequential years and hence producing significant anomalies in both the interannual and interdecadal changes in the atmospheric circulations. The selected teleconnection indices are; Pacific / North American Pattern (PNA), Western Pacific Oscillations (WP), Eastern Asia/ Western Russia (EA/WR), North Atlantic Oscillations (NAO), North Pacific Pattern (NP), Northern Oscillations (NO), Pacific Decadal Oscillations (PDO), Western Hemispherical Warm Pool (WHWP), Tropical Northern Atlantic Pattern (TNA), Tropical Southern Atlantic Pattern (TSA), Southern Oscillation Index (SOI) and El-Nino Southern Oscillations (Nino).

1.2 Forecasting Rainfall using Artificial Neural Networks (ANNs)

The methods for the prediction of data can be basically categorized into two types, classic statistical algorithms, and machine learning algorithms. Statistical methods are applicable for linear applications, whereas machine learning algorithms are applicable for non-linear applications such as prediction of rainfall. In machine learning algorithms, neural network are two types, namely, Feed Forward Neural Networks (FFNNs) and Recurrent Neural Networks (RNNs). Recurrent Neural Network is an extension of feed-forward neural network that has an internal memory. RNN is recurrent in nature as it performs the same function for every input of data while the output of the current input depends on the past one computation. After producing the output, it is copied and send back into the recurrent network. To predict, it considers the output that it has learned from the previous

Procedings of Technical Sesions, 37 (2021) 01-08
Institute of Physics- SriLanka

input. Unlike FFNNs, RNNs can use their internal state to process sequences of input. In RNNs, all the inputs are related to each other

1.3 Long Short – Term Memory (LSTM) Networks

Long Short – Memory (LSTM) Networks are a modified version of recurrent neutral networks, which makes it easier to remember past data in memory. The vanishing gradient problem of RNN is resolved here. LSTM is well-suited to classify, process and predict time series given time lags of unknown duration. It trains the model by using back – propagation.
This study focuses RNN in combination with Long Short- term Memory. The LSTM can hold many previous datasets, thereby mitigating vanishing gradient, which leads to better accuracy in prediction.

2 DATA USED

Monthly mean rainfall data from Colombo surface weather station of the Department of Meteorology, Sri Lanka covering the period 1961- 2010 (50 years) were used. Teleconnection indices were downloaded from the National Centers for Environmental Prediction-National Centers for Atmospheric Research (NCEP-NCAR) reanalysis dataset for the period 1961-2010.

3 METHODOLOGY

3.1 Data Analysis and Pre-processing

	Nino_Index	PDO_Index	NAO_Index	SOI_Index	WP_Index	PNA_Index	EA/WR_Index	TSA_Index	TNA_Index	WHWP_Index	MRf
Nino_Index	1.000000	0.368296	-0.012403	-0.736989	0.112901	0.160224	0.082970	-0.007705	0.192129	0.358215	0.008356
PDO_Index	0.368296	1.000000	0.032879	-0.340638	-0.028997	0.314376	-0.036189	0.027682	0.135703	0.174939	-0.034257
NAO_Index	-0.012403	0.032879	1.000000	-0.012818	0.105469	0.045875	0.056436	-0.130086	-0.251708	-0.133517	0.029125
SOI_Index	-0.736989	-0.340638	-0.012818	1.000000	-0.116022	-0.140773	-0.088735	0.070217	-0.064815	-0.225648	0.039333
WP_Index	0.112901	-0.028997	0.105469	-0.116022	1.000000	-0.007710	0.060732	-0.024688	-0.023945	-0.006435	0.064813
PNA_Index	0.160224	0.314376	0.045875	-0.140773	-0.007710	1.000000	0.020002	0.048375	0.186342	0.218586	0.035670
EA/WR_Index	0.082970	-0.036189	0.056436	-0.088735	0.060732	0.020002	1.000000	-0.129583	-0.153714	-0.104962	-0.015771
TSA_Index	-0.007705	0.027682	-0.130086	0.070217	-0.024688	0.048375	-0.129583	1.000000	0.198520	0.194733	0.025940
TNA_Index	0.192129	0.135703	-0.251708	-0.064815	-0.023945	0.186342	-0.153714	0.198520	1.000000	0.608128	0.058815
WHWP_Index	0.358215	0.174939	-0.133517	-0.225648	-0.006435	0.218586	-0.104962	0.194733	0.608128	1.000000	0.109606
MRf	0.008356	-0.034257	0.029125	0.039333	0.064813	0.035670	-0.015771	0.025940	0.058815	0.109606	1.000000

Figure 3.1 – Correlation matrix values

Analysis was carried out to determine the correlation of each factor to the Mean Rainfall of Colombo Sri Lanka. As the first step data acquired is indexed based on the time. Monthly mean rainfall data is indexed based on the assumption of average rainfall for the entire month. Indexes are also calculated based on the monthly frequency. To carry out further analysis, correlation for matrix was calculated to find the relation between the monthly mean rainfall and the indexes. According to the correlation matrix we can see that there is no direct correlation to any of the indexes. Even though the indexes are not

Procedings of Technical Sesions, 37 (2021) 01-08
Institute of Physics- SriLanka

related directly according to the timestamp of the given month, it can be observed that a pattern exists in the indexes but in different time steps. Because of that it cannot be further analyzed by the statistical methods but need to use a Neural Network approach.

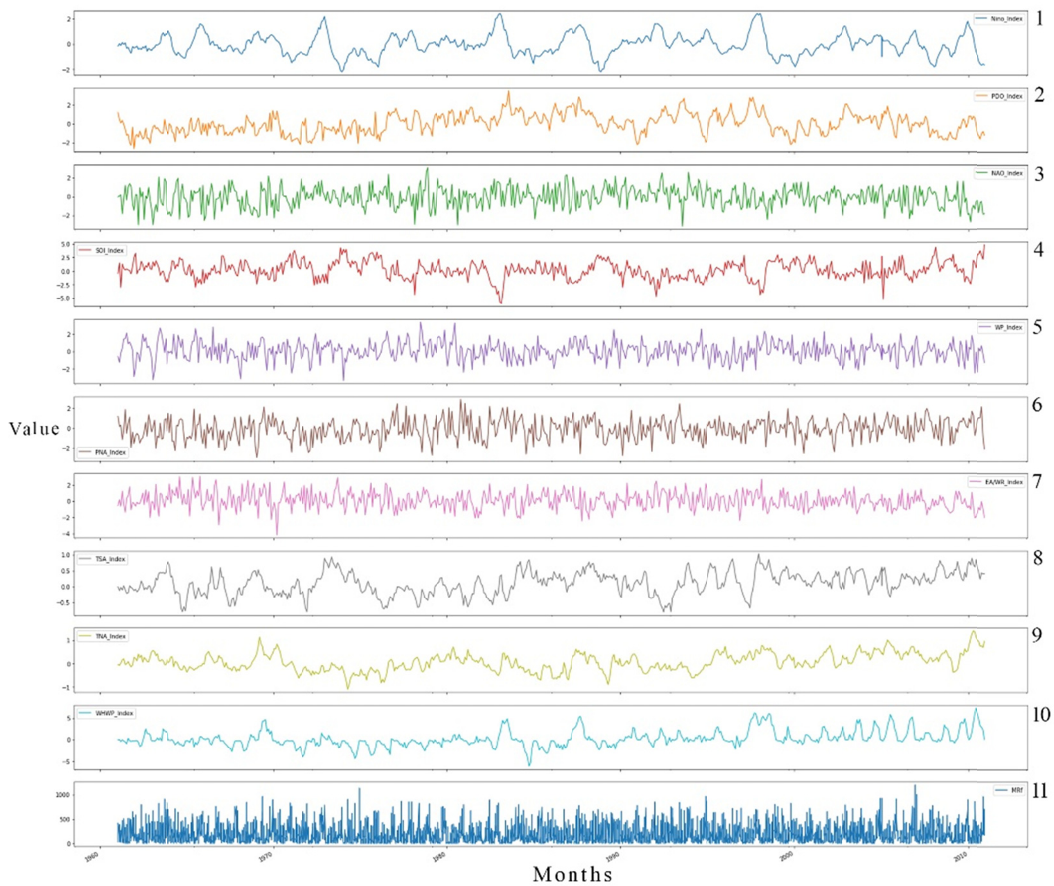

Figure 3.2 - Monthly Mean Rainfall (MRf[11]) against Nino[1], PDO[2], NAO[3], SOI[4], WP[5], PNA[6], EA/WR[7], TSA[8], TNA[9] and WHWP[10] indexes

3.2 Neural Network

In the process of developing a neural network there are many methods that can be used to develop the neural network. When analyzing the predicting variable, it can be observed that the past rainfall (last month) also an affecting factor. Rainfall alone itself also indexed on the time. Hence the predicting model need to comprise of the time series. Considering all those characteristics best neural network to be used for this prediction is the Recurrent Neural Network. In the Recurrent Neural Networks there are different types of nodes that can be used such as Gated Recurrent Units and Long Short-Term Memory Architecture. Since LSTM gives us the most control out of the others LSTM's are used.

Procedings of Technical Sesions, 37 (2021) 01-08
Institute of Physics- SriLanka

Univariate LSTM Models. LSTMs can be used to model univariate time series forecasting problems. These are problems comprised of a single series of observations and a model is required to learn from the series of past observations to predict the next value in the sequence.

Since the set of observations have multiple variables univariate time series is not applicable hence, we need to use multivariate time series analysis. Monthly mean rainfall needs to be predicted based on the different indexes.

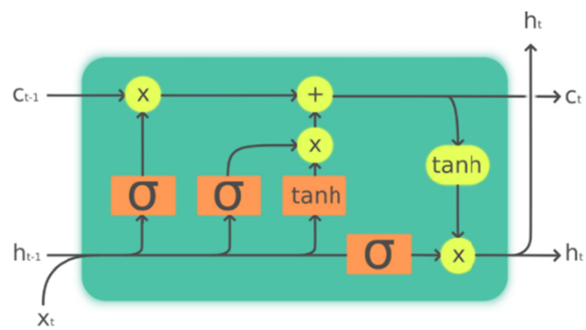

Figure 3.1 - Long Short Term Memory Node

3.2.1 Data preparation

Initial dataset consists of year, monthC, monthN, monthly mean rainfall, Nino Index, PDO Index, NAO Index, SOI Index, WP Index, PNA Index, EA/WR Index, TSA Index, TNA index and WHWP Index. To make this a time series prediction year, monthC and monthN was taken together to form the index of the dataset. The data of the month was assumed to be the 15th of each month since the rainfall of the entire month is taken as the mean rainfall of the entire month. From the dataset it can be observed that the data ranges for each column is quite different and hence this might deprecate the training of the neural network and the accuracy. Because of that the dataset is normalized using standard MinMaxScaler with the range between 0 and 1. This will make sure all the values of the columns are between 0 and 1 and does not have a bias on the prediction.Next step of the process is to make the sequence of the data to feed to the recurrent neural network. For the sequence, two pass month data are used and predicting the future month mean rainfall. For each layer sigmoid activation function is used and the “Adam” optimizer is used to compile the model. For the measurement of the model accuracy Mean Absolute Error is measured. Training hours for the model used as 10 years. For the training of the model 200 epochs were used and results trained model obtained in the end.

3.2.2 Model

Model consists of two LSTM layers each containing 100 neurons each and one dense layer.

Procedings of Technical Sesions, 37 (2021) 01-08
Institute of Physics- SriLanka

```
Model: "sequential"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
lstm (LSTM)                  (None, 1, 100)            43200
_________________________________________________________________
lstm_1 (LSTM)                (None, 100)               80400
_________________________________________________________________
dense (Dense)                (None, 2)                 202
=================================================================
Total params: 123,802
Trainable params: 123,802
Non-trainable params: 0
_________________________________________________________________
```

Figure 3.4 - Neural Network Architecture

4 EVALUATION

After training the model the results were compared against the test set of the data. Figure 4.1 shows the mean square error of the predictions. It can be observed that the test sets mean absolute error is lower than that of the train mean absolute error. This implies that the model might be over trained and would be showing biased results for the trained dataset. Reason for this biasness may appear due to the effective time lag of the indexes effect for the monthly mean rainfall of the Colombo. The reason for this time delay is the distance between the indexes(ocean) and the Colombo City. Since this effect cannot be mitigated, it needed to be addressed.

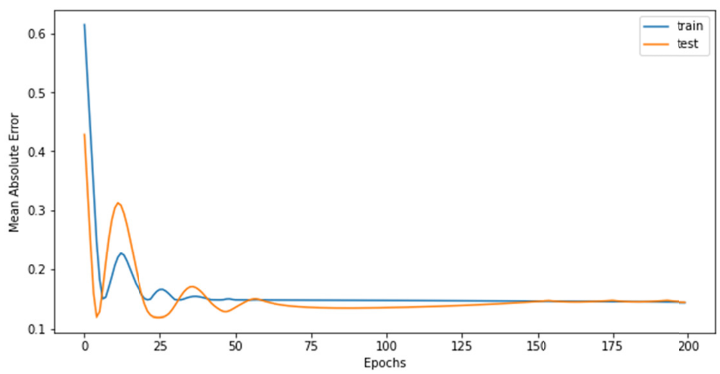

Figure 4.1 - Mean Absolute Error (MAE) against the Epochs

To address the lag time steps and the overtraining, lagged time step was introduced to the variables(indexes). In the model developed time unit is taken as the month and two month time delay was introduced as the time lag to the variables. After training the model the following results can be obtained. Observing the above graph, we can see that the mean absolute error of the test is higher than that of the train and hence the model is not overfitting to the training data. This result proves that the index values that are affecting the mean rainfall of Colombo are lagged by two months. The conclusion that can be observed from this is that Mean Rainfall of City Colombo are related to Nino Index, PDO

Procedings of Technical Sesions, 37 (2021) 01-08
Institute of Physics- SriLanka

Index, NAO Index, SOI Index, WP Index, PNA Index, EA/WR Index, TSA Index, TNA index and WHWP Index with a time lag of two months and can be predicted accurately using those indexes.

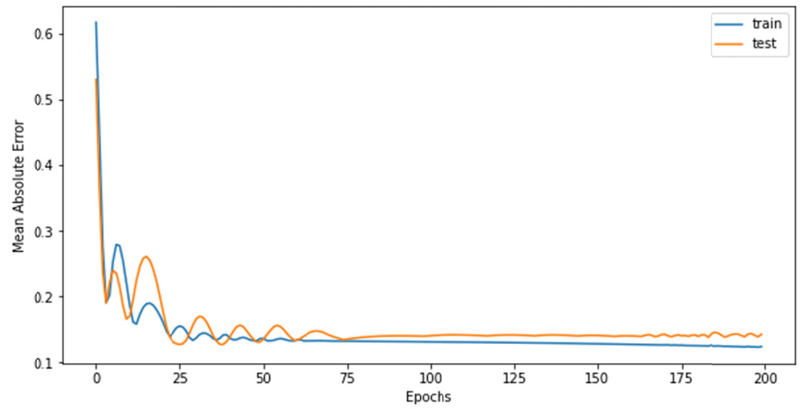

Figure 4.2 - Mean Absolute Error (MAE) against Epochs with lagged variable time-steps

Observing the above graph, we can see that the mean absolute error of the test is higher than that of the train and hence the model is not overfitting to the training data. This result proves that the index values that are affecting the mean rainfall of Colombo are lagged by two months. The conclusion that can be observed from this is that Mean Rainfall of City Colombo are related to Nino Index, PDO Index, NAO Index, SOI Index, WP Index, PNA Index, EA/WR Index, TSA Index, TNA index and WHWP Index with a time lag of two months and can be predicted accurately using those indexes.

5 DISCUSSION AND FUTURE WORK

The teleconnection indices, namely Nino Index, PDO Index, NAO Index, SOI Index, WP Index, PNA Index, EA/WR Index, TSA Index, TNA index and WHWP Index have been used as predictor variables to predict the monthly rainfall in city of Colombo Sri Lanka.In this study only 50 years data (600 datapoints) were used. A large datapoints might provide a good forecasting tool. At the same time too few or too many input parameters can affect either learning or prediction capability of the network.Selection of input parameters of ANNs depends on the area, topography, amount, and intensity of rainfall. Spatiotemporally different teleconnections may have different time lags and therefore with the consideration of them more accurate predictions may be possible.

6 REFERANCES

[1] A. K. Sahai, M. K. Soman, and V. Satyan, “All India summer monsoon rainfall prediction using an artificial neural network,” *Clim. Dyn.*, vol. 16, no. 4, pp. 291–302, 2000, doi: 10.1007/s003820050328.

Procedings of Technical Sesions, 37 (2021) 01-08
Institute of Physics- SriLanka

[2] "Seasonal forecasting of Indian summer monsoon rainfall: A review - Publications of the IAS Fellows." http://repository.ias.ac.in/90481/ (accessed Dec. 26, 2020).

[3] D. N. Kumar, M. J. Reddy, and R. Maity, "Regional rainfall forecasting using large scale climate teleconnections and artificial intelligence techniques," *J. Intell. Syst.*, vol. 16, no. 4, pp. 307–322, 2007, doi: 10.1515/jisys.2007.16.4.307.

[4] Basnayake, B.R.S.B., Fernando, T.K. and Vithanage, J.C. (2002) *Variation of Air Temperature and Rainfall during Yala and Maha Agricultural Seasons,* Proceedings of the 58th Annual Session of Sri Lanka Association for the Advancement of Science (SLASS), Section E1, p. 212.

[5] Premaalal, K.H.M.S. (2009) *Weather and Climate Trends, Climate Controls & Risks in Sri Lanka,* Presentation made at the Sri Lanka Monsoon Forum, April 2009. Department of Meteorology, Sri Lanka.

[6] Malmgren, B.A., Hulugalla, R., Hayashi, Y. and Mikami, T. (2003*) Precipitation Trends in Sri Lanka since the 1870s and Relationships to El Niño—Southern Oscillation,* International Journal of Climatology, 23, 1235-1252.

[7] Jayawardene, H.K.W.I., Sonnadara, D.U.J. and Jayewardene, D.R. (2005) *,* Sri Lankan Journal of Physics, 6, 7-17.