MAUSAM, **74**, 3 (July 2023), 699-706

MAUSAM

DOI : https://doi.org/10.54302/mausam.v74i3.4933
Homepage: https://mausamjournal.imd.gov.in/index.php/MAUSAM

UDC No. 551.577 : 551.509.313 : 551.502 (540)

Rainfall forecasting in the Barak river basin, India using a LSTM network based on various climate indices

PARTHA PRATIM SARKAR

Department of Civil Engineering, Faculty of Engineering, Assam Down Town University, Panikhaiti, Guwahati, Assam

(*Received 27 December 2021, Accepted 23 December 2022*)

e mail : partha.sarkar@adtu.in

सार – प्रस्तावित अध्ययन भारत के उत्तरपूर्वी क्षेत्र में बराक नदी द्रोणी में मासिक वर्षा कापूर्वानुमान करने के लिए 4 महीने पहले तक केपूर्वानुमान क्षितिज के लिए एक दीर्घ अल्पकालिक स्मृति (LSTM) तंत्रिका नेट वर्क (NN) को नियोजित करता है। नौ महत्वपूर्ण जलवायु परिवर्तियों में से, समुद्र सतह तापमान (SST), समुद्र स्तर दबाव (SLP), नीनो 3.4 सूचकांक, भारतीय ग्रीष्मकालीन मॉनसून वर्षा (ISMR) विसंगतियों और द्विधुवी मोड सूचकांक (DMI) को सबसे उपयुक्त पूर्वानुमानकर्ता के रूप में पहचाना गया और NN में निवेश के रूप में पेश किए गए। LSTM एक विशेष प्रकार का आवर्तक तंत्रिका नेट वर्क (RNN) है, जो अपनी कोशिका अवस्था में संचयी रूप से सुविधा निष्कर्षण और मेमोरी को स्टोर करने में माहिर है। मॉडल परिणाम संभावित पूर्वसूचक सेट और द्रोणी में वर्षा वितरण के बीच प्रबल सहसंबंध दर्शाते हैं। विभिन्न सांख्यिकीय उपायों के संदर्भ में प्राप्त पूर्वानुमान परिणामों की जाँच की गई और पूर्वानुमानों को वास्तविक समय प्रेक्षणों (0.90 से अधिक सहसंबंध और 85% से अधिक हिट स्कोर) के बराबर पाया गया। मॉडल के परीक्षण चरण में 12.45% से 15.65% की सीमा में रूट माध्य वर्ग त्रुटियाँ उत्पन्न हुई, जो मॉडल के संतोषजनक प्रदर्शन को उजागर करती हैं। विभिन्न जलवायु सूचकांकों को शामिल करने के प्रस्तावित विधि क्षेत्र में वर्षा के पूर्वानुमान के लिए एक नया दृष्टिकोण बनाती है जिससे जल संसाधनों का समय पर और प्रभावी प्रबंधन हो सकता है।

ABSTRACT. The proposed study employs a long short-term memory (LSTM) neural network (NN) to forecast monthly rainfall in the Barak river basin in the northeastern region of India for a prediction horizon up to 4 months in advance. Out of nine significant climate variables, sea surface temperature (SST), sea level pressure (SLP), Nino 3.4 index, the Indian summer monsoon rainfall (ISMR) anomalies and dipole mode index (DMI) were identified to be the best-suited predictors and were introduced as the inputs in the NN. The LSTM is a special kind of recurrent neural network (RNN) which specializes in feature extraction and storing memory in its cell state cumulatively. The model results display strong correlations between the potential predictor sets and the rainfall distribution across the basin. The obtained forecast results were scrutinized in terms of various statistical measures and the predictions were found to be at par with the real time observations (correlations greater than 0.90 and hit score greater than 85%). The testing phase of model produced root mean square errors in the range of 12.45% to 15.65% highlighting satisfactory model performance. The proposed method of incorporating different climate indices form a novel approach to forecast rainfall in the region which may lead to timely and effective management of water resources.

Key words – Climate variables, Dipole Mode index, Forecasting, Indian summer monsoon, Long short-term memory neural network, River basin.

1. Introduction

For an agricultural country like India, precise prediction of rainfall is of utmost importance as irregular forecasts can lead to damage to crops, farms and well as property. Proper rainfall prediction also helps in flood mitigation, management as well as resource utilization. However, forecasting rainfall is a challenging task and depends on various climatic factors. Traditional methods (numerical and statistical models) and coupled climate model approaches have been long used in forecasting rainfall. With the advent of artificial intelligence, data driven and machine learning has come to the fore in climate change/modeling studies. The suitability of the mentioned methods lies in the fact that they are able to generalize nonlinear processes due to their intrinsic architecture. Rainfall prediction is also a highly complex and non-linear process. Deep learning neural networks (NN), an ancillary of data driven techniques, have been widely used in such studies as they have the

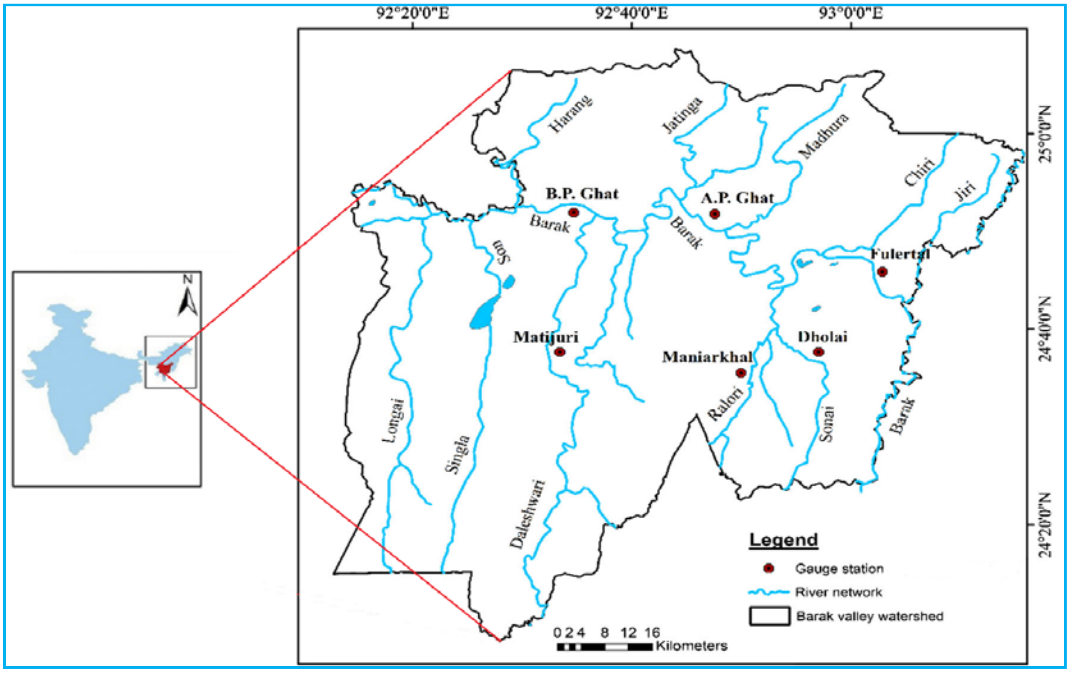

Fig. 1. Location of the study area and the meteorological stations

capability to identify and mimic natural climatic (non-linear) processes by mapping potential predictors and the predictand without necessarily having the knowledge of the working of the natural phenomena.

Researchers have formulated a considerable number of rainfall prediction models by employing artificial neural networks (ANNs). Wu and Chau (2010) used ANNs to predict monthly rainfall for meteorological stations in China. Yuan *et al*. (2016) utilized climate indices to forecast summer rainfall in the yellow river. Htike and Khalifa (2010) formulated a focused time delay NN model to forecast rainfall across many stations in India one, five and ten years in advance by only using archival summer rainfall data. Chattopadhyay and Chattopadhyay (2008) predicted rainfall across several stations in India using max. and min. temperatures as the climate variables to their NN model. Kumar *et al*. (2012) forecasted rainfall using NNs at Udupi, Karnataka, India utilizing average humidity and wind speeds. Abbot and Marohasy (2012) predicted rainfall across the Murray Darling Basin, Australia by considering monthly rainfall, climate indices, air temperature and solar data as the input to their NN model. Chakraverty and Gupta (2008) forecasted the southwest monsoon across India using ANN models. Dabhi and Chaudhary (2014) forecasted precipitation in Anand, India by using a wavelet ANN by considering temperatures, humidity and the previous day's rainfall. The above-mentioned studies showcase the capability of ANNs to forecast rainfall at longer time intervals. Badr *et al*. (2013) and Rasel *et al*. (2015) in their studies revealed the superior forecasting ability of data driven approaches as compared to their statistical and coupled climate model counterparts. However, with constantly improving and evolving artificial intelligence techniques, the need of the hour is to build a model that can provide precise and accurate results. In the above-mentioned studies, ANN models have been successfully applied to many regions all across India except the North East (NE) part of the country. The NE region of India, located between 22° N - 29° N boasts unique topographic features. This region has a prolonged winter lasting from early November until mid-march with many mountainous ranges and valleys. The NE region has a sub-tropical climate and receives the highest average rainfall as compared to other regions of the country. This rainiest region of the world, Cherrapunji, a place in Meghalaya that receives an average annual rainfall of 11,420 mm, lies in close proximity to the Barak river basin in NE India. Except for a few places, most of the region receives an average rainfall of around 2000 mm yearly. The number of rainy days compared to other parts of the country are comparatively much higher, varying from 90-120 days. The temperatures in this region are also very contrasting in nature. The temperatures in the Barak valley are around 30° C to 31° C during the summer months while the mean temperatures in winter are around 12° C - 13° C. All the above-mentioned factors and features add to the complexity of our problem to forecast rainfall in the Barak river basin.

TABLE 1

Statistics of the monthly rainfall data across the meteorological station

Station name	Discharge Max. (m^3/s)	Discharge Min. (m^3/s)	Mean Discharge (m^3/s)	Standard Deviation (m^3/s)
Badarpur Ghat	4811.23	380.03	2120.23	890.20
Annapurna Ghat	3401.27	128.20	1281.23	580.21
Matijuri	5705.45	1059.91	1440.34	879.45
Fulertal	1498.48	118.56	351.71	289/12
Dholai	161.20	20.01	39.21	29.21
Maniarkhal	5530.25	250.81	2201.23	940.12

In the proposed study, we have used a deep learning long short-term memory (LSTM) NN to forecast rainfall in the Barak river basin by resorting to various lagged climate indices. The advantage of the LSTMs to learn long-term temporal dependencies and extracting features from datasets were utilized in this present work. The features extracted using the LSTMs were then used to calibrate the lagged correlations between the climate indices and rainfall in the region. The potential and best-suited climate indices were then used as predictors in the model to forecast rainfall in the basin. The obtained results were measured in terms of various statistical measures to validate the model predictions. The proposed work is structured as follows. The next section introduces the study area and the different climate variables followed by two sections dedicated to methodology and model results respectively. The last section provides the conclusion.

TABLE 2

Climate indices along with their highest lagged correlation coefficient values

Climate Index	Lag time (in months)	Correlation coefficient
SLP	5	0.48
SST	6	-0.45
AT	8	0.25
uWND	6	0.27
vWND	6	0.31
200HGT	8	0.30
Nino 3.4	4	0.45
ISMR anomalies	5	0.39
DMI	5	-0.51

2. Study area and description of data sets

The Barak river basin lies in the NE region of the India with an area about 41,157 km^2 and lies between east longitudes 90°10′ to 95°7′ and north latitudes 21°58′ to 26°24′ shown in Fig. 1 below. The bed gradient of the Barak River is very flat and varies from 1:10,000 in the upper reach to 1:20,000 in the lower reach. The surface and ground water potential of the basin are 48.3 km^3 and 1.8 km^3 respectively. The basin is surrounded by terrain on three sides and reaches Bangladesh at the downstream end, finally entering the Bay of Bengal. More than 70 % of the basin is covered with vegetation and the river accounts for only 1.5 % of the total basin area. Interestingly, this region is the one of the only two regions (Western Ghats) across the whole country that receives very heavy rainfall after the withdrawal of the southwest monsoon, *i.e.*, retreating monsoon.

The basin boundary is delineated from the 30 m resolution SRTM DEM obtained from the United States Geological Survey (USGS) with the help of ArcSwat tool of ArcGIS software in GIS environment. The monthly rainfall data across the basin for different meteorological stations were acquired from the data archival section of the Indian Institute of tropical meteorology (IITM), Pune for the period 1950-2017. The study aims at forecasting rainfall during the summer monsoon period from May to August (MJJA). The statistics of the monthly rainfall data is shown below in Table 1.

Nine climate indices were used as predictors in the proposed LSTM model. Climate variables, namely, Sea level pressure (SLP), air temperature (AT), u wind (u WND) component, v wind (v WND) component and Geo potential height anomalies at 200 hPa (200 HGT) were obtained from the National centers of environmental prediction (NCEP) for the period 1980-2017 gridded at 2.5° × 2.5° for the entire basin region. Sea surface temperature (SST), Dipole Mode Index (DMI) and Nino 3.4 index gridded at 2° × 2° were utilized from the data archival section of the National Oceanic and Atmospheric

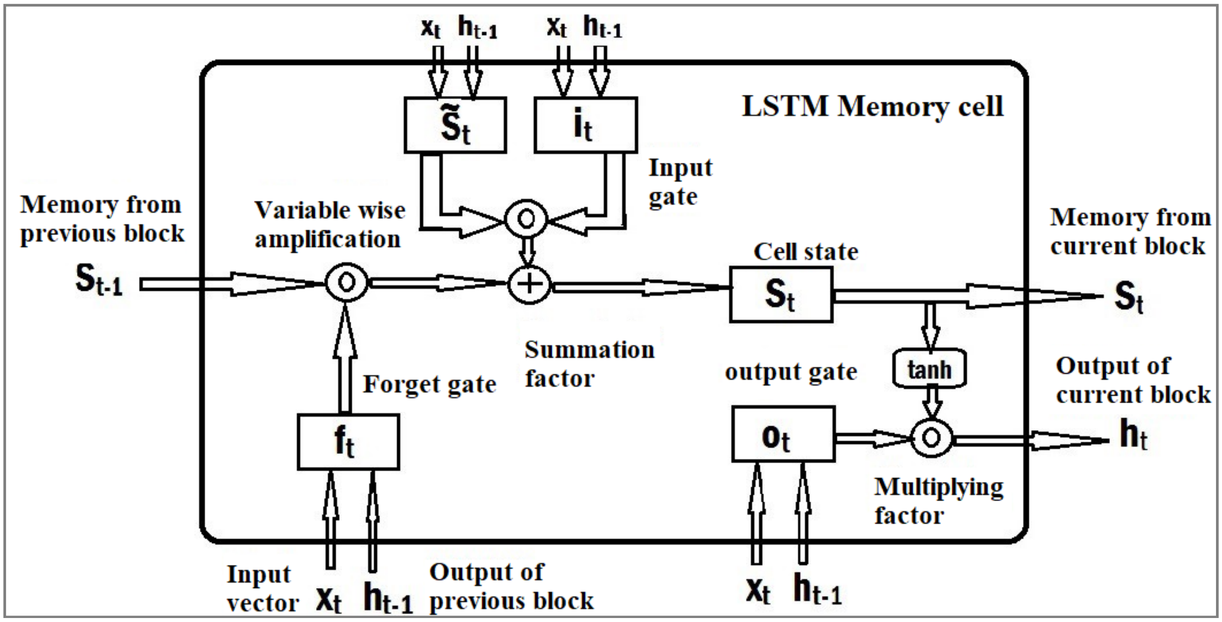

Fig. 2. A typical long short-term memory (LSTM) neural network architecture

Administration (NOAA) for the same period. The Indian summer monsoon anomalies were obtained from the Indian institute of Tropical Meteorology (IITM), Pune for 1980-2017 for all meteorological stations in the river basin. Numerous past studies have substantiated the use of lagged climate variables as potential predictors for data driven neural network models (Abbot and Marohasy, 2017; Peres *et al*. 2015). The lagged correlations (CC) between the selected climate variables and the precipitation during the months MJJA are presented in Table 2.

Out of nine overall indices, SLP, SST, Nino 3.4 and ISMR anomalies correlated highly with the predictand while DMI negatively correlated the most with the monthly rainfall for the period MJJA. The SST and DMI that are negatively correlated indicate inverse relationship with the predictand (forecasted rainfall). The positive correlations indicate direct influence over the forecasted predictand.

3. Methodology

The LSTMs are a special kind of recurrent neural networks (RNNs) that have the ability to learn long-term dependencies and extract significant features from the available data sets. Certain gates achieve this, namely forget, input and output gates (Gers and Schmidhuber, 2001). The forget gate analyzes the data and utilizes only the information that is deemed necessary. The input gate preferences the part of current input vector. The output gate determines the outcome of the hidden and reserved state. This modulation is achieved by activation functions, namely, hyperbolic tangent across the input gate and sigmoid function across the forget and the output gate. The cell state store the information cumulatively (Schmidhuber, 2015). The working of a typical LSTM cell state is shown in Fig. 2.

The climate variables and the monthly rainfall relationship for the period MJJA is calibrated using Spearman's correlation as shown in Table 1. The LSTM sequence-to-sequence prediction is basically a multiple input multiple output (MIMO) forecast. For a given time series, T = $(\varphi^1, \varphi^2 \ldots \varphi^N)$, MIMO statuettes T to establish a relationship between the different time series data, X = $(\varphi^{t-n+1}, \varphi^{t-1} \ldots \varphi^t)$ and subsequent observation, Y = $(\varphi^{t+1}, \varphi^{t+2} \ldots \varphi^{t+H})$,

Where H > 1 is the prediction horizon. The proposed LSTM model aims to estimate and evaluate the conditional probability: P $(y_1, y_{2\ldots}\, y_T{}^1 / x_1, x_2 \ldots x_T)$, where $(x_1, x_2 \ldots x_T)$ is the input arrangement of predictors and $(y_1, y_{2\ldots}\, y_T{}^1)$ is the eventual target output, where time T^1 varies from T. The overall number of predictor neurons are equal to the number of the input climate indices. The monthly rainfall datasets were further divided into three segments, above, near and below average rainfall conditions for MJJA. The categorization of rainfall data was done under the assumption that it follows a normal distribution with a mean μ and standard deviation σ based on μ + 0.43σ. The associative strength of each input climate index on the final output (MJJA monthly rainfall) was achieved by determining variable's importance/ contribution to final

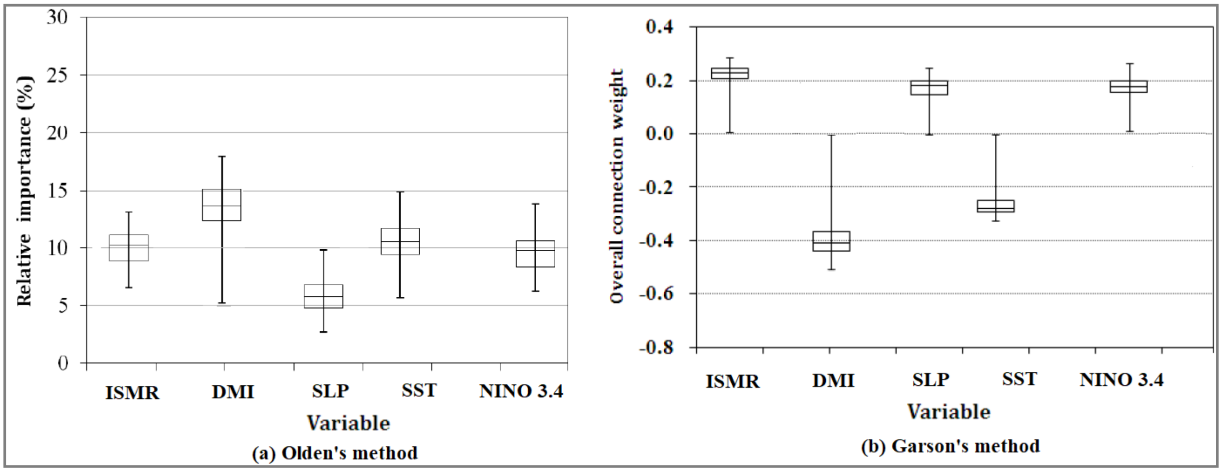

Fig. 3. Overall relative importance and connection weights of the five selected climate variables using Olden's and Garson's methods respectively

output by analyzing their connection weights. We used Garson's and Olden's methods to actuate the allied significance of each index on the final output (Garson, 1991). Although, Olden's method is a more flexible approach to access individual variable's importance as the proposed appraoch employs the connected weight algorithm (Olden and Jackson, 2002). This step is necessary as it requires less time and we do not have to rely on the arduous hit or miss approach for obtaining the optimal number of hidden nodes and associated weights. The associative influence of the input variables obtained from Garson's and Olden's method is shown in Fig. 3 in the form of a box plot.

Results obtained through Garson's method show that the difference between max. and min. values were about 8% - 14% for the input climate variables. The greatest deviation was observed in the case of DMI and the lowest in the case of HGT 200. The average values for the predictor variables ranged from 3% to 20%, with DMI and SST showing the strongest connection and AT and HGT 200 showing the weakest connection. The top five significant predictor variables identified were DMI, SST, Nino 3.4, ISMR anomalies and SLP (in order of decreasing connection strength). However, Garson's method has certain drawbacks, as the technique does not consider the direction of the input- output relationship. In this method, there can be a counteraction of the response of a variable when the positive response of one hidden node meets the negative response of another hidden node, finally nullifying the overall response. Therefore, Olden's method was additionally applied to preserve such response information of a climate variable. DMI and SST showed a greater negative response of –0.55 and –0.47 respectively, as compared to the results achieved using Garson's method. ISMR anomalies showed a greater response to MJJA rainfall than Nino 3.4 and SLP. However, both the methods predicted the same climate variables as the inputs in the LSTM model. The response of the variables in decreasing order of association are as follows: DMI, SST, ISMR, Nino 3.4 and SLP.

3.1. *Optimal LSTM network for forecasting*

The potential climate variables (predictors) are reconstructed into a three-dimensional (S, M, F) shape, where S is the window size of the training progression, M is the number of iteration steps and F is the number of extracted features, all of which are normalized to a range of -1 to 1. The predictand (MJJA monthly rainfall) is also transformed into two dimensions (1, M).The data sets involving all the potential climate indices (DMI, SST, ISMR, Nino 3.4 and SLP) were divided into 3 parts, namely, training (70%), testing (15%) and validation (15%) sets. The initial work was to find the optimal LSTM metrics that can forecast MJJA over longer time steps ahead in the future. The initial learning rate and momentum rate were kept at 0.005 and 0.0. The number of neurons per layer were set at 50. There were no tuning of hyper parameters during this initial stage of training. Hyperparameters were tuned after exhaustive training including all the predictors with the set of epochs initially set at 200 to obtain the optimal architecture of the LSTM network. The model performance is validated using K-fold cross validation (Stone, 1974; Rodriguez *et al.*, 2010). This is necessary to eradicate the problem of over fitting of the model. The model forecast results are then statistically evaluated in terms of root mean square error, coefficient of correlation, mean square error and hit score rate (rate expressed as percentage for correct prediction of near, below and above average rainfall conditions for MJJA).

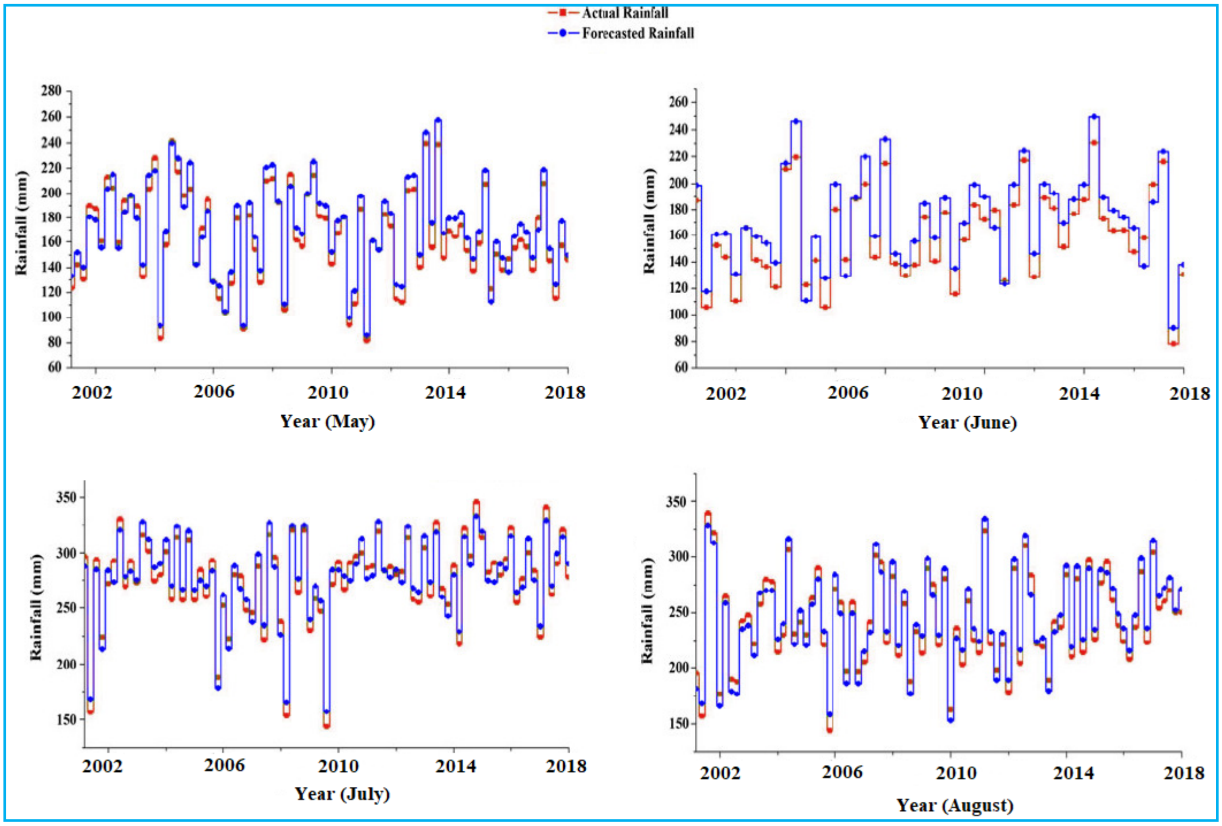

Fig. 4. Observed and predicted monthly rainfall for the months of May, June, July and August (MJJA) respectively for the testing period 2002-2018

TABLE 3

LSTM model performance for training, validation and test sets in terms of RMSE and correlation coefficient

Hidden neurons	RMSE (in %)			Coefficient of correlation		
	Training	Validation	Testing	Training	Validation	Testing
50	14.32	15.87	17.49	0.82	0.77	0.85
100	14.08	16.89	16.84	0.87	0.83	0.87
150	12.45	14.02	15.65	0.92	0.88	0.91
200	15.34	17.32	17.61	0.86	0.82	0.87
250	16.37	18.87	19.32	0.81	0.79	0.83
300	18.01	20.24	22.40	0.77	0.76	0.77

4. Model results

The training, validation and test results of the LSTM model is shown in Table 3. It is observed from Table 3 below that with the increase in hidden neurons, the RMSE decreases during both testing and validation to a certain extent and then increases again. A Similar trend is observed with correlation coefficient as well. At 150 hidden neurons, the LSTM model reaches the most optimal architecture for forecasting MJJA rainfall. RMSE during validation and testing ranged from 12.45% to 15.65%. Correlation coefficient (CE) ranged from 88.54% to 91.25% during the validation and test phases. Mean square error (MAE) ranges from 5.89% to 8.35%, which

TABLE 4

Hit and fail percentages for the different rainfall groups

Group	Below Normal	Above Normal	Near Normal	Total
Hit (correctly predicted)	18	16	17	51
Fail (wrongly predicted)	3	3	2	8
Total outcomes	21	19	19	59
Hit percentage	85.71	84.21	89.47	86.44

is decent and acceptable. The obtained model results are fair and satisfactory in terms of CE, RMSE and MAE. LSTM model with 150 hidden neurons, threshold gradient set at 1.0, learning rate set at 0.002, learn rate drop period and learn rate drop factor set at 125 and 0.2 respectively produced the best forecast results. Maximum iterations were set at 400.

The forecasted and observed MJJA rainfall for the period 2002 to 2018, considered as the testing of the model, is shown in Fig. 4.

The forecasted and observed monthly rainfall values during MJJA for the test period 2002-2018 is shown in Fig 4. Hit rate ratio is the number of times that a correct prediction was made in the ratio of total predictions made. The hit rate score for correctly forecasting above, below and more than mean monthly rainfall was ascertained and it was observed to be more than 84% for all the three scenarios shown in Table 4.

The present study deliberates the use of a data driven technique to predict rainfall in the Barak river basin of NE India for several months in the future, a first attempt of such a kind for the region. The model results are at par with the existing models, to forecast monthly rainfall across an entire region. Higher correlation and hit score values depict a clear picture of the model learning the intricacies and complexity of rainfall modeling across the basin. Spearman's correlation for the observed and the forecasted values were ascertained at 0.85 showing satisfactory performance in terms of a nonlinear statically parameter as well.

5. Conclusion

The constructed deep learning model to forecast monthly rainfall was intended basically to shorten the time-consuming work of developing a data driven neural network by choosing appropriate and important parameters (climate indices). The outcome of monthly rainfall over a basin depends on several climate variable-atmosphere interaction. The trial and error approach is not always suitable for such an extensive task. Therefore, correlation analysis was employed to determine potential and possible climate lagged indices. Out of many indices used in the proposed study to forecast MJJA rainfall, we selected DMI, SST, Nino 3.4, ISMR anomalies and SLP based on Olden's and Garson's weight correlation. The most optimal LSTM architecture was then determined and results were evaluated using several statistical metrics namely, RMSE, CE and MAE. The forecast results were also compared with real time data for the period 2002-2018, which showed adequate and satisfactory performance of the model. This approach of assessment of the input variables based on their relative importance to the final output, monthly rainfall is significant as the proposed approach enhances the predictive capability of the model by removal of parameters that are less crucial or decisive.

Successful prediction of lower rainfall values (below average) gives an extra edge to the modeling capability of this data driven approach. Accurate prediction for the forecast horizon of four months from May to August, which is mainly the Indian summer rainfall period, could lead to a greater formable and effective management of water resources across the entire region. The proposed technique could provide adequate time to develop plans against potential flooding or drought conditions in the Barak river basin as the entire NE India receives considerably more rainfall than the rest of the country. The model results that are at par with other traditional models, also convinces us to employ such AI based techniques, as an effective alternative.

Declarations

Funding : Not Applicable

Conflicts of interest/Competing interests : The authors hereby declare no conflicts of interest or competing interests.

Availability of data and material : This study employs climate variable datasets from the National Oceanic and Atmospheric Administration (NOAA), the National centers of environmental prediction (NCEP) and the Indian institute of Tropical Meteorology (TropMet), Pune. The modeling and analysis of the above-mentioned approaches are performed in Matlab R2018b, specifically the Neural Network toolbox and the Deep Learning toolbox.

Code availability : Not applicable

Disclaimer : The contents and views expressed in this study are the views of the authors and do not necessarily reflect the views of the organizations they belong to.

References

Abbot, J. and Marohasy J., 2012, "Application of artificial neural networks to rainfall forecasting in Queensland, Australia", *Adv. Atmos. Sci.*, **29**, 717-730.

Abbot, J. and Marohasy, J., 2017, "Application of artificial neural networks to forecasting monthly rainfall one year in advance for locations within the Murray Darling basin, Australia", *Int. J. Sustain. Dev. Plan*., **12**, 1282-1298.

Badr, H. S., Zaitchik, B. F. and Guikema, S. D., 2013, "Application of statistical models to the prediction of seasonal rainfall anomalies over the Sahel", *J. Appl. Meteorol. Climatol*., **53**, 614-636.

Chakraverty, S. and Gupta, P., 2008, "Comparison of neural network configurations in the long-range forecast of southwest monsoon rainfall over India", *Neural Comput. Appl*., **17**, 187-192.

Chattopadhyay, S. and Chattopadhyay, G., 2008, "Comparative study among different neural net learning algorithms applied to rainfall time series", *Meteorological applicat.*,**15**, 2, 273-280.

Dabhi, V. K. and Chaudhary, S., 2014, "Hybrid Wavelet-Postfix-GP model for rainfall prediction of the Anand region of India", *Advances in Artificial Intell*., 1-11.

Garson, G. D., 1991, "Interpreting neural-network connection weights", *AI Expert*., **6**, 47-51.

Gers, F.A. and Schmidhuber, E., 2001, "LSTM recurrent networks learn simple context-free and context-sensitive languages". *IEEE Transactions on Neural Networks*, **12**, 6, 1333-1340.

Htike, K. K. and Khalifa, O. O., 2010, "Rainfall forecasting models using Focused Time-Delay Neural Networks", *Comput. And Commun. Eng. (ICCCE). Int. Conf. at IEEE.*

Kumar, A., Kumar, A., Ranjan, R. and Kumar, S., 2012, "A rainfall prediction model using artificial neural network", *Control and Syst. Graduate Research Colloq. (ICSGRC*), 82-87.

Olden, J. D. and Jackson, D. A., 2002, "Illuminating the "black box": A randomization approach for understanding variable contributions in artificial neural networks", *Ecol. Model*., **154**, 135-150.

Peres, D. J., Iuppa, C., Cavallaro, L., Cancelliere, A. and Foti, E., 2015, "Significant wave height record extension by neural networks and reanalysis wind data", *Ocean Model*., **94**, 128-140.

Rasel, H. M., Imteaz, M. A., Hossain, I. and Mekanki, F., 2015, "Comparative study between linear and non-linear modelling techniques in rainfall forecasting for South Australia", *In Proceedings of the International Congress on Modelling and Simulation, Gold Coast, Australia*, 29 November-4 December 2015, 2012-2018.

Rodriguez, J. D., Perez, A. and Lozano, J. A., 2010, "Sensitivity analysis of k-fold cross validation in prediction error estimation", *IEEE Trans. Pattern Anal. Mach. Intell*., **32**, 569-575.

Schmidhuber, J., 2015, "Deep learning in neural networks: An overview", *Neural Network*, **61**, 85-117.

Stone, M., 1974, "Cross-validatory choice and assessment of statistical predictions", *J. R. Stat. Soc. Ser. B.*, **36**, 111-147.

Wu, C. L., Chau, K. W. and Fan, C., 2010, "Prediction of rainfall time series using Modular Artificial Neural Networks coupled with data preprocessing techniques", *J. of hydrology*, **389**, 1, 146-167.

Yuan, F., Berndtsson, R., Uvo, C. B., Zhang, L. and Jiang, P., 2016, "Summer precipitation prediction in the source region of the Yellow River using climate indices", *Hydrol. Res.*, **47**, 847-856.