Modelling Weather Conditions Using Encoder-Decoder and Attention Based on LSTM Deep Regression Model

Khder Alakkari [1], Mostafa Abotaleb [2], Amr Badr[3], Ammar Kadi[4], A. M. Ghazi Al khatib[5], Bayan Mohamad Alshaib[5], El-Sayed M. El-kenawy[6]

[1]Department of Statistics and Programming, Faculty of Economics, University of Tishreen, Tartous P.O. Box 7 2230, Syria
[2] Department of System Programming, South Ural State University, Chelyabinsk 454080, Russia
[3]Faculty of Science, School of Science and Technology University of New England, NSW Armidale, Australia
[4]Department of Food and Biotechnologies, South Ural State University, Chelyabinsk
[5]Department of Banking and Insurance, Faculty of Economics, Damascus University, Damascus, Syrian Arab Republic
[6]Department of Communications and Electronics, Delta Higher Institute of Engineering and Technology, 16 Mansoura, 35111, Egypt

Emails: khderalakkari1990@gmail.com; abotalebmostafa@bk.ru; Amr.Mostafa@live.com; ammarka89@gmail.com; abdullah1991.alkhatib@damascusuniversity.edu.sy; bayan1992.alshaib@damascusuniver-sity.edu.sy; skenawy@ieee.org

Abstract

In the rapidly evolving field of smart cities, the accurate prediction of weather patterns plays a crucial role in various industries such as agriculture, tourism, and socioeconomic development. This study utilizes Artificial Intelligence (AI) and Machine Learning (ML) through advanced machine learning techniques, including Encoder-Decoder LSTM and Attention LSTM models, to analyze daily climatic weather data in the Narmadapuram district. The research investigated the future patterns of key weather parameters, including maximum temperature, minimum temperature, morning relative humidity, evening relative humidity, and bright sunshine hours. The study analyzed daily data collected between November 1, 1977 and April 30, 2022, with 80% used for training and 20% for testing. Results showed that the Encoder-Decoder LSTM model outperformed the Attention LSTM model in forecasting maximum temperature, morning relative humidity, evening relative humidity, and bright sunshine hours, while the Attention LSTM model had better results in predicting minimum temperature. The findings provide valuable insights into climatic patterns and variability and have implications for the development of more precise weather forecasting models. This study demonstrates the potential of AI and ML in addressing the challenges of smart cities and highlights the significance of machine learning techniques in weather forecasting, a critical aspect of urban operations and decision-making.

Keywords: Smart Cities; Weather; Time series; Forecasting; (Encoder-decoder) LSTM; Attention LSTM.

1. Introduction

Long short-term memory models got a significant upward trajectory of use after its pioneering paper [1]. The main goal of its design was to avoid the problems of simple recurrent neural networks and to obtain better results. It is widely preferred over GARCH models have been generalized and applied in wide areas to solve real-world problems at the level of time series that have high frequencies after the pioneering papers of [2] [3], through the great importance implied by these models and their ability to model variable fluctuations Over time and filling the gap in ARIMA models [4], but the problem

Doi: https://doi.org/10.54216/IJAACI.010201
Received: January 10, 2022 Accepted: May 27, 2022

of not taking the random evolution of fluctuations and relying only on deterministic evolution caused the incomplete validity of the predictions of these models

The problem also lies in these data as a result of the validity of their predictions for short-term periods and the inability to obtain all the information from the chain Temporal, and therefore from here comes the need to use deep learning models represented by LSTM model and its developments, which allows obtaining more information from the series, no matter how high the noise, and producing better predictions [5] [6] [7] [8].

The aim of designing and used LSTM networks is to reduce long-term dependency and its negative impact on the learning process. In addition to the four gates that the network depends on for its work, it helps the network to remember the most important information, which greatly improves the quality of the output. One area of use of the LSTM model is to predict stock values and returns in financial markets [9], it has also been used in traffic flow forecasts such as [10], LSTM models have been used in the field of forecasting ability of Solar Irradiance and Photovoltaic Power [11]. Moreover, Water quality analysis and forecasting, like [12].

The use of LSTM models has also developed in the environmental field, as a result of the large climate changes, it refers to long-term shifts in temperatures and weather patterns, Human activities have become the main cause of climate change, mainly due to the burning of fossil fuels, such as coal, oil and gas. Some of the effects of climate change are: higher temperatures - destructive storms - increased drought - loss of species - food shortages. The contribution of renewables to the economy is tested through the structure through the LSTM model [13], they have also been used to capture the precise spatio-temporal relationships of multiple meteorological features for temperature prediction [14]. [15] Use long short-term memory (LSTM) model to capture the dependence between historical climate data. [16] proposed a rainfall-runoff analysis system for the Kratie station on the Mekong River mainstream using the long short-term memory (LSTM) model, forecasting method is proposed based on meteorology data of next day for ideal weather condition, using long short term memory (LSTM) networks [17], [18]formulates the prediction problem as a structured output prediction problem jointly predicting multiple outputs simultaneously. The proposed prediction model is trained by using long short-term memory (LSTM) networks taking into account the dependence between consecutive hours of the same day. As utilized LSTM to obtain a data-driven forecasting model for an application of weather forecasting [19].

Despite the wide use of LSTM models, they contain a set of problems that may in some cases lead to the loss of part of the information and the failure to produce accurate predictions [20]. One of the problems is called the "sequence-to-sequence problem [21] [22] [23], which is called seq2seq for short. This problem can be described as the number of sequence elements at the time of input differs from the number at the time of output, which leads to the loss of important information. Encoder – Decoder LSTM models are widely used because of their superiority in the fields used to solve this problem [24]. Nevertheless, with a long sequence of inputs, as in the case of time series, the ED LSTM model encodes a fixed length input sequence [25]. This imposes limits on the length of input sequences that are in the learning phase and causes worse performance for long input sequences. Attention is used with the aim of freeing the decoder structure from its internal fixed-length representation [26].

Through this research, we aim to examine the future behaviors of climatic weather data like Max Temperature, Min Temperature, Relative Humidity (M), and Relative Humidity (E) And Bright Sun Shine Hours in the Narmadapuram district. Using Encoder-Decoder and Attention based on LSTM Model, This is done by comparing the predictive performance of the two models using a set of performance indicators for both the training and test data. This study attempts to elucidate whether imposing length restrictions on input sequences in the learning phase causes worse performance for long input sequences. The research includes four sections. This study also seeks to provide the National Informatics Center in the district with weather information in light of climate changes and the maximum and minimum of this information. This is in order to preserve the agricultural environment, being a major producer of wheat, and the diversity of livestock, and thus the possibility of planning to face challenges. The first section includes the introduction and previous studies. The second section explains the study methodology and the tools used. The third section includes a discussion of the results, and the fourth and final section contains the results of the research.

Doi: https://doi.org/10.54216/IJAACI.010201
Received: January 10, 2022 Accepted: May 27, 2022

2. Materials and Methods

A. Data and study area

The data was taken from the climate monitoring stations in Narmadapuram. data.csv file contains 6 columns (Date, Max Temperature (°C), Min Temperature (°C), Relative Humidity M, Relative Humidity E, and Bright Sun Shine Hours) and 8155 rows for daily data. Functions for training and testing Datasets were constructed based on the original dataset and the number of previous time steps used as input variables to forecast the future time period (i.e., look back), which are the two major variables. Datasets were constructed with the default setting of creating datasets with the number of observations (X) and look-backs at each point in time (t + look back). During training, we utilized a look-back value of eight days. In order to create the LSTM model, the data has to be transformed. The final format included [samples, time steps, and features]. "Looking back to the previous day's information, the samples were made up of information from that day's data, and the time step was one day (the data was gathered daily). We divided the data into two sets: one for training and the other for testing. 80% of the observations were used for training, while the remaining 20% were used for testing. Afterward, we preserved the test set and randomly selected 80% of the training set as the new training set, while the rest (20%) was the validation set. *You can download datasets and code from GitHub.* (https://github.com/abotalebmostafa11/climate)

The Location of Narmada Puram are 22.75°N and 77.72°E. In central India, the region experiences high temperatures. There is a hot, dry summer with a high of 40–42 °C close to the Tropic of Cancer (April–June). After that, there will be a monsoon with lots of rain. The winters are moderate and dry (November to February). 331 meters (1,086 feet) above sea level is the average elevation, and 134 cm of rain falls there annually (53 inch). Figure. 1 show location map of study area

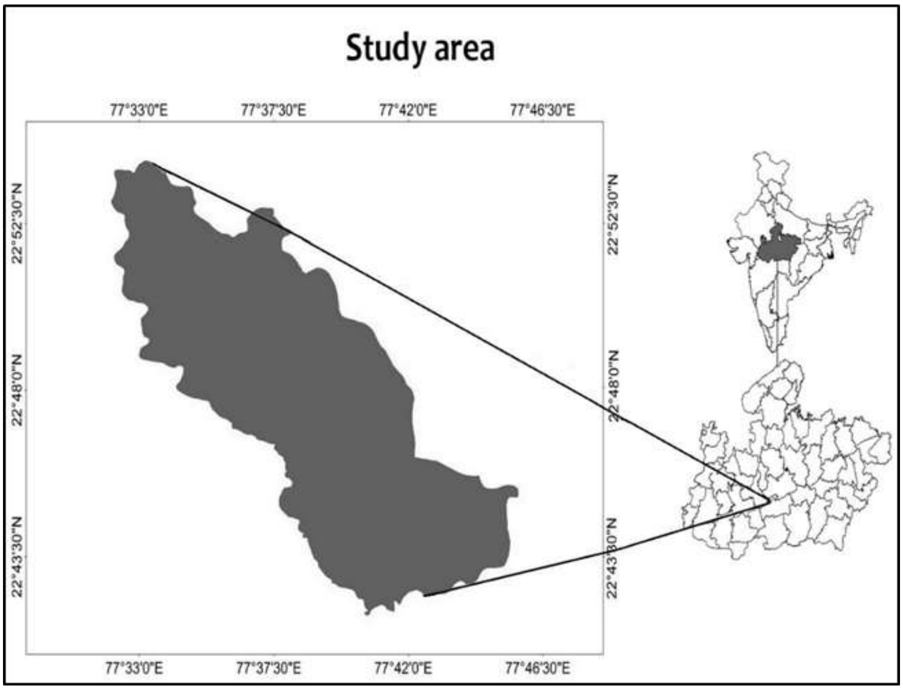

Figure 1: show location map of study area

Doi: https://doi.org/10.54216/IJAACI.010201
Received: January 10, 2022 Accepted: May 27, 2022

3. Methodology

Long Short Term Memory LSTM:

When solving the problem of vanishing gradients, long short-term memory (LSTM) was one of the earliest and most effective methods to be developed [27] [28]. In this context, "long-term" refers to simple recurrent neural networks storing information about their previous decisions as weights. A gradual shift in weights occurs throughout training when new information about the data is retrieved and used to calibrate the model. Short-lived activations hop from one node to another and are therefore referred to here. In the LSTM paradigm, a memory cell serves as intermediate storage. For the first time, multiplex nodes are incorporated into the construction of memory cells, making them a more complicated unit. Three gates (input, output, and forget) make up a generic LSTM unit [29]. With the input gate, LSTM may be programmed to either retain existing data or learn new information. The sigmoid layer and the tanh layer make up this gate's structure. The tanh layer generates a vector of potential new values to be added to LSTM [30], whereas the sigmoid layer specifies which values will be modified. To derive the final result from these layers, we use:

$$\boldsymbol{i_t} = \boldsymbol{\sigma}(\boldsymbol{W^i x_t} + \boldsymbol{U^i h_{t-1}} + \boldsymbol{b^i}) \tag{1}$$

$$\boldsymbol{u_t} = \boldsymbol{tanh}(\boldsymbol{W^u x_t} + \boldsymbol{U^u h_{t-1}} + \boldsymbol{b^u}) \tag{2}$$

where $\boldsymbol{i_t}$ is the updated value; $\boldsymbol{u_t}$ is new candidate values; $\boldsymbol{\sigma}$ is the sigmoid layer (or nonlinear function); $\boldsymbol{x_t}$ is a sequence of length t; $\boldsymbol{b}$ is constant bias; $\boldsymbol{h}$ is RNN memory at time step t; and $\boldsymbol{W}$ and $\boldsymbol{U}$ are weight matrices.

Forget gates, whose sigmoid functions are used to choose data for deletion from LSTM, are discussed in detail in [31]. The values of h and $\boldsymbol{x_t}$ are used heavily in making this determination. This gate has an output f that may take on the values 0 and 1, where 0 signifies full erasure of the acquired value and 1 represents complete preservation of the value. This result is derived by:

$$\boldsymbol{f_t} = \boldsymbol{\sigma}(\boldsymbol{W^f x_t} + \boldsymbol{U^f h_{t-1}} + \boldsymbol{b^f}) \tag{3}$$

where $\boldsymbol{f_t}$ is updated value; $\boldsymbol{\sigma}$ is the sigmoid layer (or nonlinear function); $\boldsymbol{x_t}$ represents a sequence of length t; $\boldsymbol{b}$ is constant bias; $\boldsymbol{h}$ represents RNN memory at time step t; and $\boldsymbol{W}$ and $\boldsymbol{U}$ are weight matrices.

The input gate uses a sigmoid layer to determine which sub-tree of the LSTM is responsible for the output. A nonlinear tanh function is then used to assign values between -1 and 1 after that. The output from the sigmoid layer is then multiplied by the final product. Following are

some formulae that are used to determine output:

$$\boldsymbol{o_t} = \boldsymbol{\sigma}(\boldsymbol{W^o x_t} + \boldsymbol{U^o h_{t-1}} + \boldsymbol{b^o}), \tag{4}$$

$$\boldsymbol{h_t} = \boldsymbol{o_t tanh_t c_{t-1}} \tag{5}$$

where $\boldsymbol{o_t}$ is an output gate and $\boldsymbol{h_t}$ is a value between [1, -1].

The LSTM is kept current by combining these two layers. The forget gate layer works by first doubling the previous value, $\boldsymbol{c_{t-1}}$, and then adding the candidate value, $\boldsymbol{i_t u_t}$, to forget the current value. Specifically, this process requires the following equation:

$$\boldsymbol{c_t} = \boldsymbol{i_t u_t} + \boldsymbol{f_t c_{t-1}}, \tag{6}$$

Doi: https://doi.org/10.54216/IJAACI.010201
Received: January 10, 2022 Accepted: May 27, 2022

where $\boldsymbol{c_t}$ represents a memory cell and $\boldsymbol{f_t}$ represents a value between 0 and 1 produced by the forget gate. Specifically, a value of 0 denotes that the value is nullified, whereas a value of 1 indicates that it is retained [32]. Figure 8 depicts a potential configuration including these components.

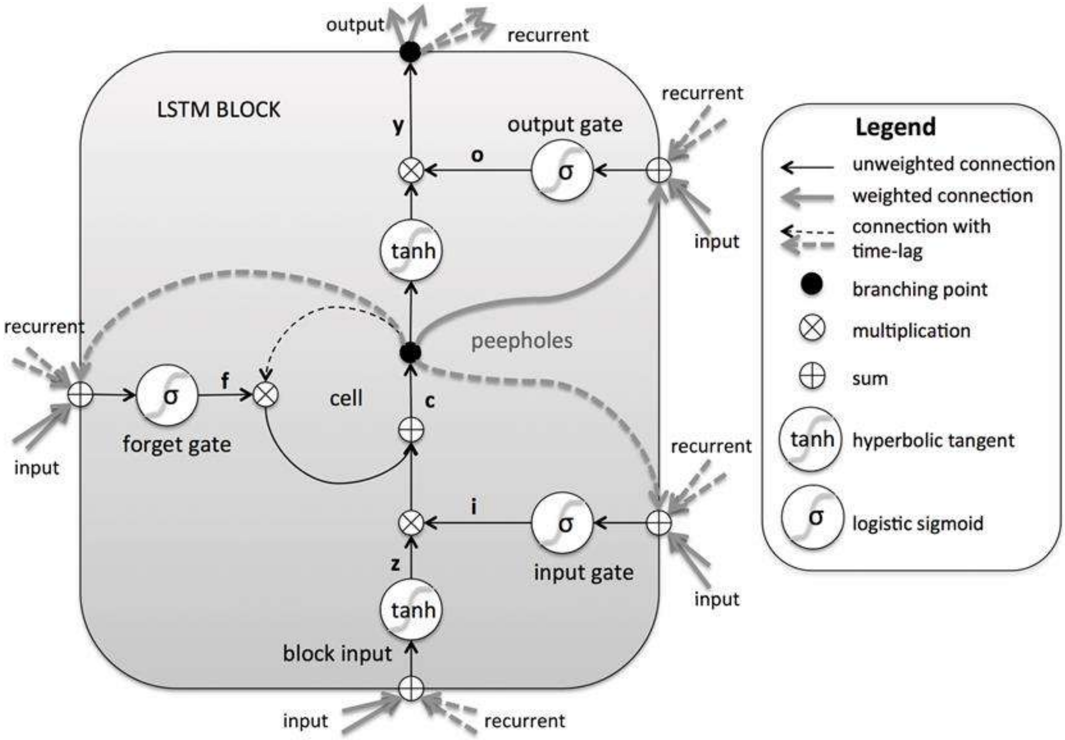

Figure 2: The Long Short-Term Memory (LSTM) model.

Figure 2 explains us the mechanism work of LSTM model, the input information is passed to the forget layer, at which point the model decides to: (a) keep the information in the past and use it for prediction, or (b) forget the information and rely on the instantaneous state, then send this information to a tanh function to normalize the information and extract features and patterns and remove noise from them [33]

Encoder-Decoder LSTM:

Encoder – Decoder LSTM model were primarily designed to address the sequence-to-sequence problem, which is called seq2seq for short. This problem can be described as the number of sequence elements at the time of input differs from the number at the time of output, which leads to the loss of important information [34]. The modeling problem in this case is that the length of the input sequence may differ from the length of the output sequence due to the multiple lengths of the input and output steps. Accordingly, Encoder – Decoder LSTM is used, which is one of the methods that have proven effective to avoid the problem of seq2seq [35].

This architecture is comprised of two models: one for reading the input sequence and encoding it into a fixed-length vector, and a second for decoding the fixed-length vector and outputting the predicted sequence [36] [37]. Which can be merged of Encoder-Decoder LSTM designed specifically for seq2seq problems.

The main objective of the coding phase is to extract more features and information from the input time series data. The data of an asymmetric sequence of length $X = \{x_1, x_2, .. x_t\}$ is used as input and the encoder encodes the sequence into a fixed-length state vector$\boldsymbol{c}$, which is used as input to the decoder [38]. In the decoder stage, the decoder decodes the state vector $\boldsymbol{c}$ and predicts the next time sequence $\boldsymbol{Y}$ by integrating the input data for the current time.

Doi: https://doi.org/10.54216/IJAACI.010201
Received: January 10, 2022 Accepted: May 27, 2022

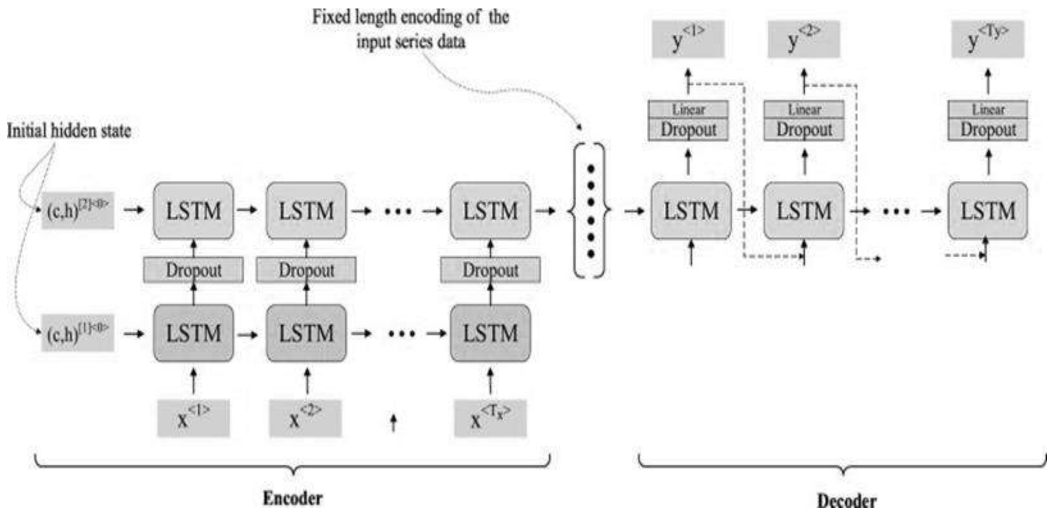

Figure 3: The mechanism of work of the Encoder-Decoder LSTM model.

Figure 3 show us the mechanism work of Encoder – Decoder LSTM model, the hidden layer state h is evolved each time the input data is read. When reading the end of the sequence X, the hidden layer variable ***c***, can be thought of as a summary of the input sequence. Which means that the features and information in the sequence have been extracted and mapped in ***c***.

Attention LSTM:

Encoder – Decoder LSTM models are widely used because of their superiority in the fields used. However, with a long sequence of inputs, as in the case of time series, the ED LSTM model encodes a fixed length input sequence [25]. This imposes limits on the length of input sequences that are in the learning phase and causes worse performance for long input sequences [39] [40].

Attention is used with the aim of freeing the decoder structure from its internal fixed-length representation. The attention mechanism allows obtaining different information of first-order and lower-order importance and not just the first-order important information. It is described as mapping a query and a set of key-value pairs to an output, where the query, keys, values, and output are all vectors [41]. The output is computed as a weighted sum of the values, where the weight assigned to each value is computed by the query's compatibility function with the corresponding key [42]. There are also the same technique using in machine learning models .

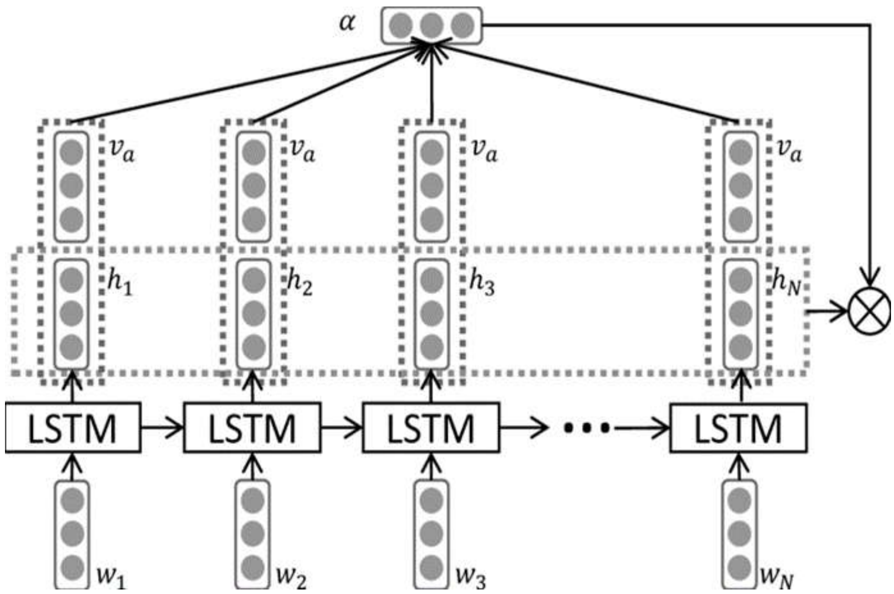

Figure 4: The architecture of attention – based LSTM.

Doi: https://doi.org/10.54216/IJAACI.010201
Received: January 10, 2022 Accepted: May 27, 2022

Figure 4 show us how attention LSTM work, it include The first step, we map x_t to h_t:

$$\boldsymbol{h_t = f_t(h_{t-1}, x_t)} \qquad (7)$$

Where f is nonlinear activation function, $h_t \in R_t^s$ is hidden state at time t, s is size of hidden state, in the second step, an attention mechanism is built through a stochastic attention model. For a particular feature sequence$x^k = (x_1^k, \dots, x_m^k)$. From the previous hidden state h_{t-1} and cell state $\boldsymbol{c_{t-1}}$in the LSTM unit, it is determined [43]:

$$\boldsymbol{a_t^k = V^T \tanh\left(W_1.[h_{t-1}, c_{t-1}] + W_2 x^k\right)} \qquad (8)$$

$$\boldsymbol{\beta_t^k = softmax(a_t^k) = \frac{\exp(a_t^k)}{\sum_{i=1}^{n} \exp(a_t^k)}} \qquad (9)$$

where$V$: is vector, W_1 and W_2 are matrices and both learnable parameters by model. a_t^k: is vector has length m and its $h_t - i$ measures the importance of $h_k - i$ input features sequence at time t. and normalized by softmax. β_t^k: is an attention weight, which contains a score of how much attention should be put on $h_k - i$ feature sequences.

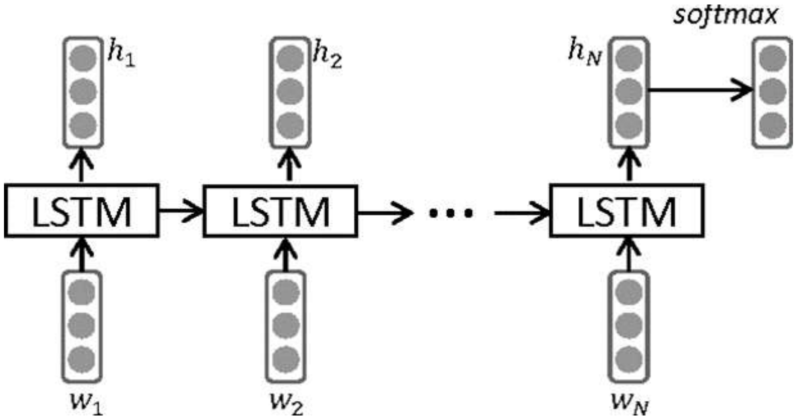

Figure 5: The architecture of attention – based LSTM Normalized by softmax.

From Figure 5, the different information in the sequence length is sent to the Softmax function to calibrate to uniform weights. Accordingly, the output of the attention model at time t of weighted input feature $u_t$is as follows:

$$u_t = (\beta_t^1 x_t^1, \beta_t^2 x_t^2, \dots, \beta_t^n x_t^n)^T \qquad (10)$$

Thus, the x in equation 1 is replaced by the weights $u_t$in the current equation to develop the attention model. It is possible to obtain attention-based time series with better features than input sequence elements.

Performance indicators:

We use indicators to evaluate the performance of the models used to determine their ability to explain the features and information contained in the data. This is done by examining the extent to which the estimated values using the model correspond to the actual values, taking into account the avoidance of an under fitting problem that may appear from the training data, and an over fitting problem that appears through the test data. The following performance indicators include:

Indicator	Formula
Mean Absolute Error (MAE)	$\boldsymbol{\frac{1}{n}\sum_{t=1}^{n}\left\lvert y_t - \hat{y}_t\right\rvert}$ (11)
R-Squared	$\frac{Cov(y_t, \hat{y}_t)}{\sqrt{V(y_t)\,V(\hat{y}_t)}}$ (12)

Doi: https://doi.org/10.54216/IJAACI.010201
Received: January 10, 2022 Accepted: May 27, 2022

Root Mean Square Error (RMSE)	$\sqrt{\frac{\sum_{t=1}^{n}(\hat{y}_t - y_t)^2}{n}}$ (13)
Relative Root Mean Square Error (RRMSE)	$\sqrt{\frac{\frac{1}{n}\sum_{t=1}^{n}(\hat{y}_t - y_t)^2}{\sum_{t=1}^{n}(\hat{y}_t)^2}}$ (14)
Mean Square Error (MSE)	$\frac{\sum_{t=1}^{n}(\hat{y}_t - y_t)^2}{n}$ (15)

Where $\hat{y}_t$the forecast is value; y_t is the actual value; and n is the number of fitted observed. The smaller the values of these indicators, the better the performance of the model.

4. Results

Daily data collected for Narmadapuram district between November 1, 1977 and April 30, 2022 was divided into two parts: training (80%) and testing (20%).

Table 1: Descriptive statistics for MAX Temperature, MIN Temperature, Relative Humidity (M), Relative Humidity (E) and Bright Sun Shine Hours between November 1, 1977 and April 30, 2022

	MAX Temperature	MIN Temperature	Relative Humidity (M)	Relative Humidity (E)	Bright Sun Shine Hours
Length	8155	8155	8155	8155	8155
Mean	26.19	9.53	85.47	41.47	7.23
STD	6.58	5.45	13.70	18.23	2.93
Mini	8	-3.5	11	4	0
25%	21.45	5.30	81	29	5.90
50%	25.4	9	90	38	8.10
75%	30.4	13.2	95	51	9.20
max	45.6	30.50	100	100	74.80
kurtosis	-0.26	-0.30	2.44	0.37	34.75
skew	0.37	0.45	-1.58	0.83	0.53
Std. error of the mean	0.073	0.060	0.15	0.202	0.032
Median	25.4	9	90	38	8.1

Doi: https://doi.org/10.54216/IJAACI.010201
Received: January 10, 2022 Accepted: May 27, 2022

95% Conf	26.049	9.42	85.17	41.08	7.17
Interv al	26.33	9.66	85.77	41.87	7.30

From November 1, 1977 to April 30, 2022, the MAX Temperature in Narmadapuram district has increased during the period from (8 °C) to (45.6 °C). Average MAX Temperature in Narmadapuram district was (26.19 °C). Kurtosis value was (-0.26) and the value of skewness was (0.37) which is close to 0.5 indicating the distributions is mesokurtic. The MIN Temperature in Narmadapuram district has increased during the period from (-3.5 °C) to (30.5 °C). Average MIN Temperature in Narmadapuram district was (9.53 °C). Kurtosis value was (-0.30) and the value of skewness was (0.45) which is close to 0.5 indicating the distributions is mesokurtic. The Relative Humidity (M) in Narmadapuram district has increased during the period from (11%) to (100%). Average Relative Humidity (M) in Narmadapuram district was (85.47%). Kurtosis value was (2.47) indicating that the distribution is flat and has thin tails (Platykurtic distributions) and the value of skewness was (-1.58), the negative skew refers to a longer or fatter tail on the left side of the distribution. The Relative Humidity (E) in Narmadapuram district has increased during the period from (4%) to (100%). Average Relative Humidity (E) in Narmadapuram district was (41.47%). Kurtosis value was (0.37) indicating the distributions is mesokurtic. the value of skewness was (0.83), indicating there is an opportunity of rising in Relative Humidity (E) in Narmadapuram district. The Bright Sun Shine Hours in Narmadapuram district has increased during the period from (0 h) to (74.80 h). Average Bright Sun Shine Hours in Narmadapuram district was (7.23 h). Kurtosis value was (34.75) indicating the data follows a Leptokurtic distribution which shows heavy tails on either side, which means there are outliers in the data. the value of skewness was (0.53), indicating there is an opportunity of rising in Bright Sun Shine Hours in Narmadapuram district. (Table, 1).

Table 2: the comparison between (Encoder-decoder) LSTM and Attention LSTM on training dataset

	(Encoder-decoder) LSTM	Attention LSTM
MAX Temperature		
MAE	1.57	1.67
R-Squared	0.91	0.90
RMSE	2.13	2.18
RRMSE	1.46	1.48
MSE	4.54	4.75
MIN Temperature		
MAE	1.86	1.86
R-Squared	0.78	0.79
RMSE	2.48	2.42
RRMSE	1.58	1.56
MSE	6.15	5.87
Relative Humidity (M)		
MAE	5.94	5.73

Doi: https://doi.org/10.54216/IJAACI.010201
Received: January 10, 2022 Accepted: May 27, 2022

R-Squared	0.62	0.60
RMSE	8.18	8.39
RRMSE	2.86	2.90
MSE	66.93	70.33
Relative Humidity (E)		
MAE	8.25	8.48
R-Squared	0.62	0.59
RMSE	11.43	12.00
RRMSE	3.38	3.47
MSE	130.60	144.12
Bright Sun Shine Hours		
MAE	1.73	1.68
R-Squared	0.42	0.37
RMSE	2.26	2.36
RRMSE	1.50	1.54
MSE	5.09	5.54

The comparison between best-fitted models (between (Encoder-decoder) LSTM and Attention LSTM on training dataset) based on, lowest values of MAE, RMSE, RRMSE, MSE and the highest value of R-Squared are shown in (Table, 2). It can be seen from (Table, 2) that (Encoder-decoder) LSTM is outperforming the Attention LSTM in most time series forecasting in training dataset in this study; the (Encoder-decoder) LSTM is the best model for forecasting MAX Temperature, Relative Humidity (M), Relative Humidity (E) and Bright Sun Shine Hours. The encoder-decoder network is useful to deal with the challenging sequence-to sequence prediction problems lately [44]. Time series forecasting tasks usually include the framework: a sequence of one or multiple input time steps transformed to a sequence of one output time step. The encoder LSTM encodes the input sequence into a fixed length vector, and the decoder LSTM decodes the fixed length vector and outputs the predicted. Since the encoder-decoder network fails to work well in predicting long sequences, AM is introduced to generate a vector based on a weighted sum of all the encoded information [45]. The only exception is in case of MIN Temperature forecasting, whereas the Attention LSTM is the best model. attention LSTM not only considers past and future data information but also highlights the effective temporal features in time series. The principal idea of Attention LSTM is to allocate corresponding attention according to the importance of information, which significantly improves the reception sensitivity and processing speed of information in the concentration area. AM is commonly used to optimize the sequence processing model by allocating correlative attention weight to the features extracted from the input sequence [46].

The performance comparison of actual and predicted values of MAX Temperature, MIN Temperature, Relative Humidity (M), Relative Humidity (E) and Bright Sun Shine Hours using the Encoder-Decoder LSTM and Attention LSTM models is shown in Figures 6 to 25. These figures have the same x-axis range (0-8155) and represent the number of values from the time series included in the plot. It should be noted that the figures provided give a visual representation of the performance of the Encoder-Decoder LSTM and Attention LSTM models in forecasting different weather parameters.

Doi: https://doi.org/10.54216/IJAACI.010201
Received: January 10, 2022 Accepted: May 27, 2022

The upper graph in each figure includes all data points in the dataset, while the lower graph includes only the past observations in the testing dataset (the past 1631 observations in the testing dataset). The blue lines represent the actual values, while the orange lines indicate the forecasted values. By comparing the forecasted and actual values, we can evaluate the performance of the two forecasting methods and gain insights on which one is more accurate. Overall, it can be seen that the forecast trends of the two models are similar to the real trend. However, for a few outliers, the forecast accuracy of each model is small. For some peak values of actual MAX Temperature, Relative Humidity (M), Relative Humidity (E) and Bright Sun Shine Hours, it can be seen that the prediction value of the Encoder-Decoder LSTM model is the closest to the actual values in the testing dataset. Encoder-Decoder LSTM is particularly useful in this study as it is able to store information for long periods of time, and can automatically feed the previous target value as an input to predict the next step in the LSTM decoder part. (Encoder-decoder) LSTM is also one type of (Encoder-decoder) RNN that is useful in dealing with challenging sequence-to-sequence prediction problems. It works by encoding the input sequence into a fixed length vector, which is then decoded by the LSTM decoder to produce the predicted output [47]. Attention Mechanism (AM) is introduced to overcome the limitation of Encoder-Decoder LSTM in predicting long sequences. AM generates a vector based on a weighted sum of all the encoded information.

On the other hand, Attention LSTM not only considers past and future data information, but it also highlights the effective temporal features in time series. It allocates corresponding attention weight to the features extracted from the input sequence, which improves the sensitivity and processing speed of information. Additionally, the Attention LSTM model can underline the important time series characteristics and reduce the influence of some outliers on the prediction effect. This is evident in the case of MIN Temperature forecasting, where the Attention LSTM model (Figures 12 and 13) performs better than the Encoder-Decoder LSTM model (Figures 10 and 11). In conclusion, the Encoder-Decoder LSTM model was found to be the best model for forecasting MAX Temperature, Relative Humidity (M), Relative Humidity (E) and Bright Sun Shine Hours, while the Attention LSTM model was the best for forecasting MIN Temperature. Both models have their own strengths and can be used in time series forecasting tasks to improve performance. This study highlights the significance of using advanced deep learning models in time series forecasting tasks, particularly in the field of weather prediction. Accurate weather forecasting is crucial in various industries, such as agriculture, transportation, and energy, as well as for public safety and disaster management. With the use of machine learning models like Encoder-Decoder LSTM and Attention LSTM, more accurate weather predictions can be made, leading to better decision-making and improved outcomes in the relevant industries and fields.

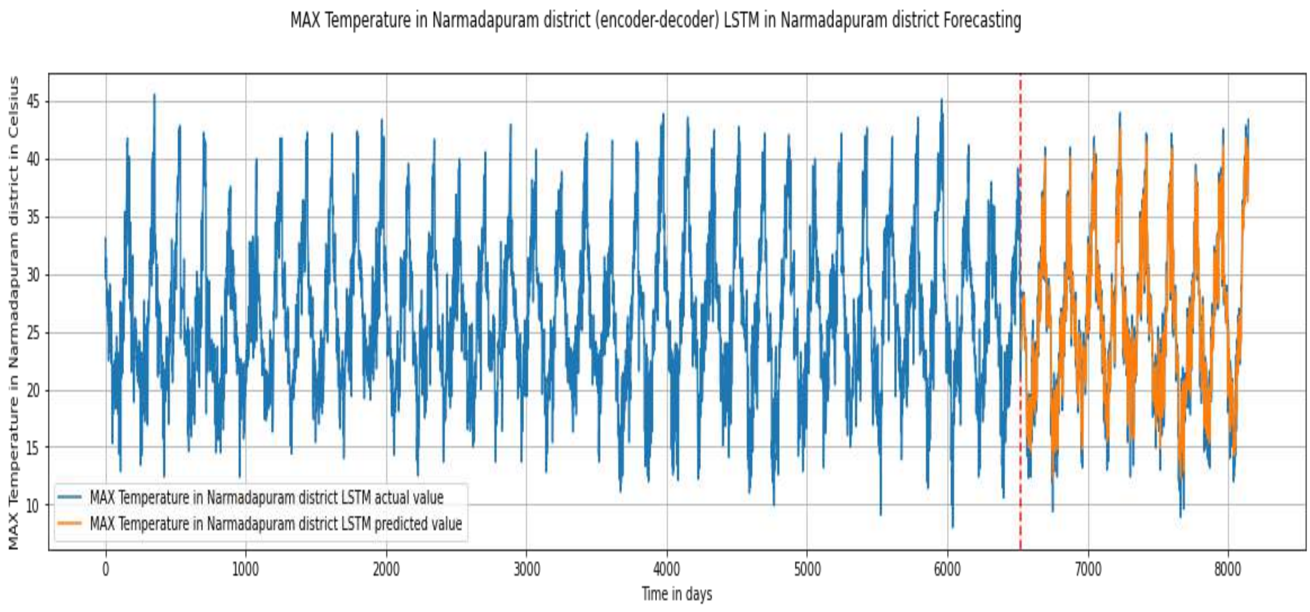

Figure 6: Forecasted vs. Actual Maximum Temperature using Encoder-Decoder LSTM Forecasting in whole dataset

Doi: https://doi.org/10.54216/IJAACI.010201
Received: January 10, 2022 Accepted: May 27, 2022

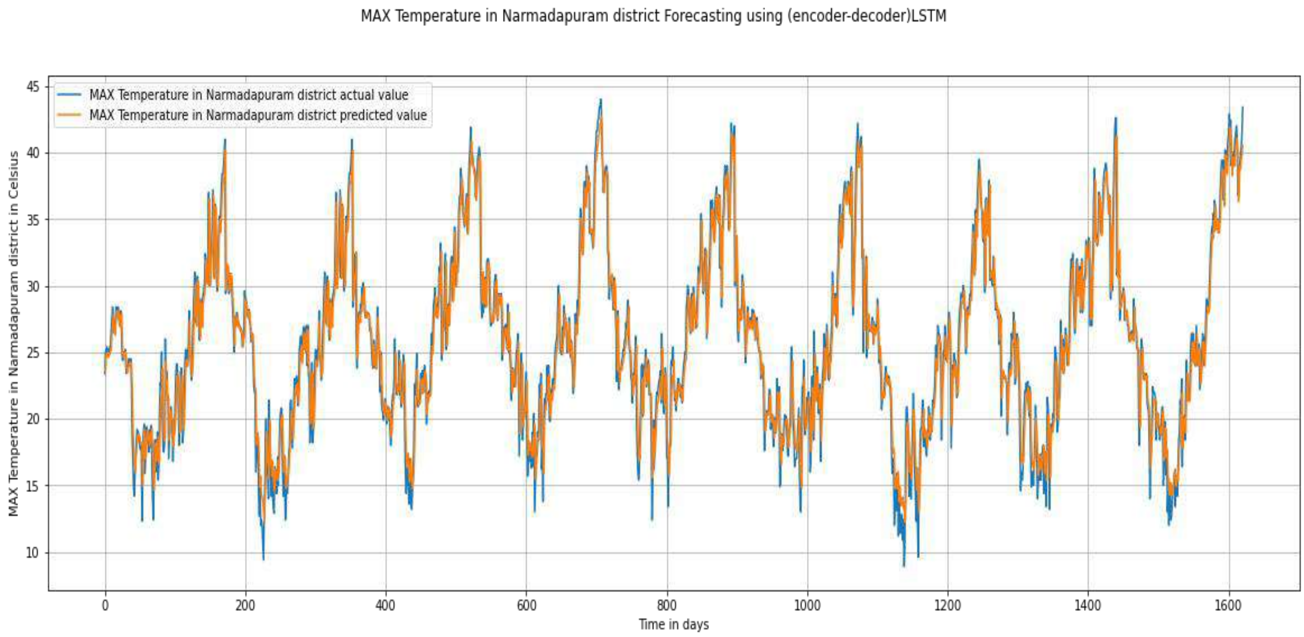

Figure 7: Forecasted vs. Actual Maximum Temperature using Encoder-Decoder LSTM Forecasting in testing dataset

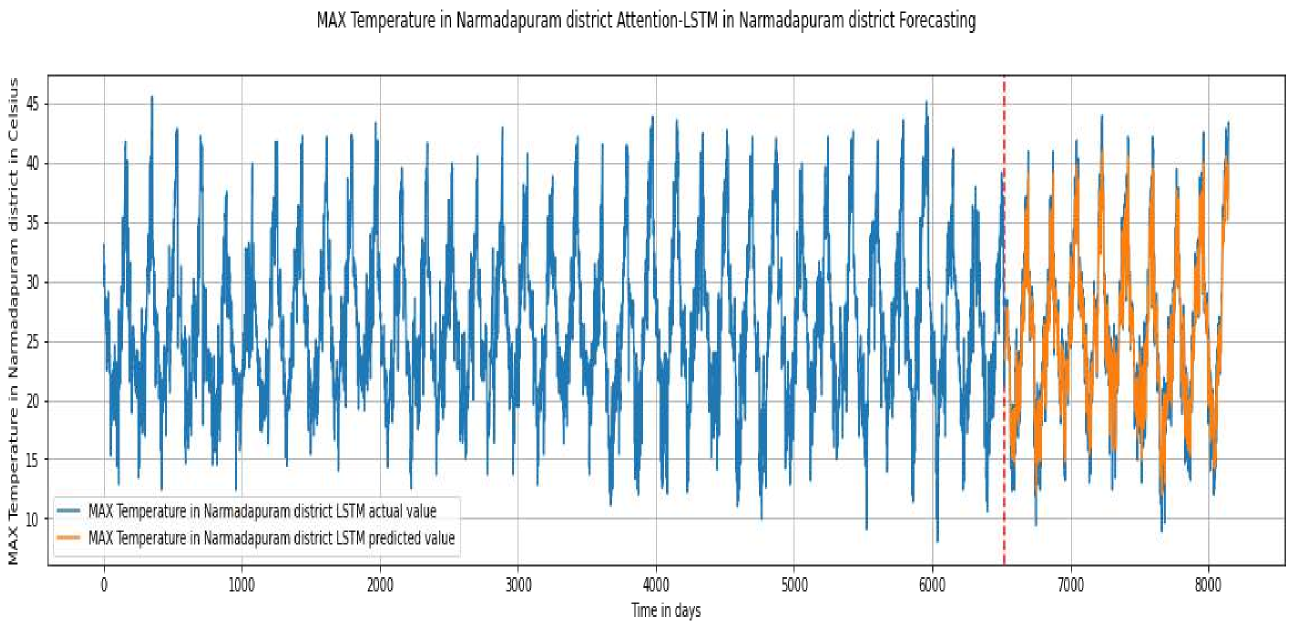

Figure 8: Forecasted vs. Actual Maximum Temperature using Attention LSTM Forecasting in whole dataset

Doi: https://doi.org/10.54216/IJAACI.010201
Received: January 10, 2022 Accepted: May 27, 2022

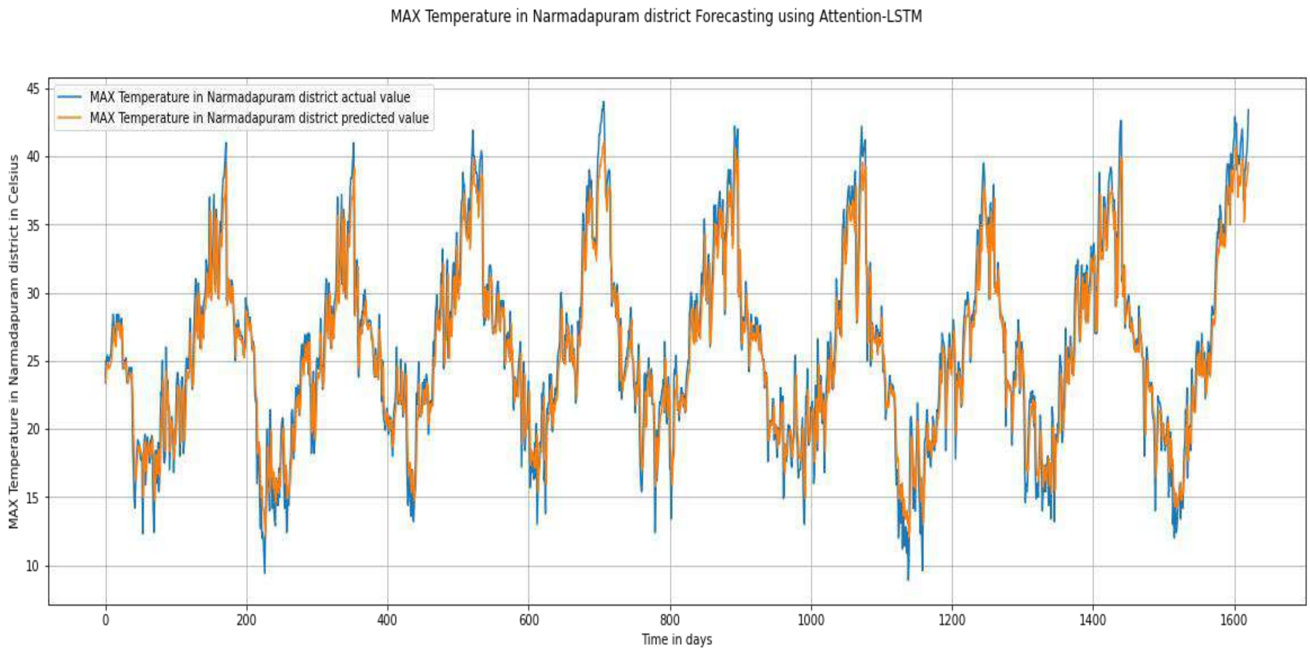

Figure 9: Forecasted vs. Actual Maximum Temperature using Attention LSTM Forecasting in testing dataset

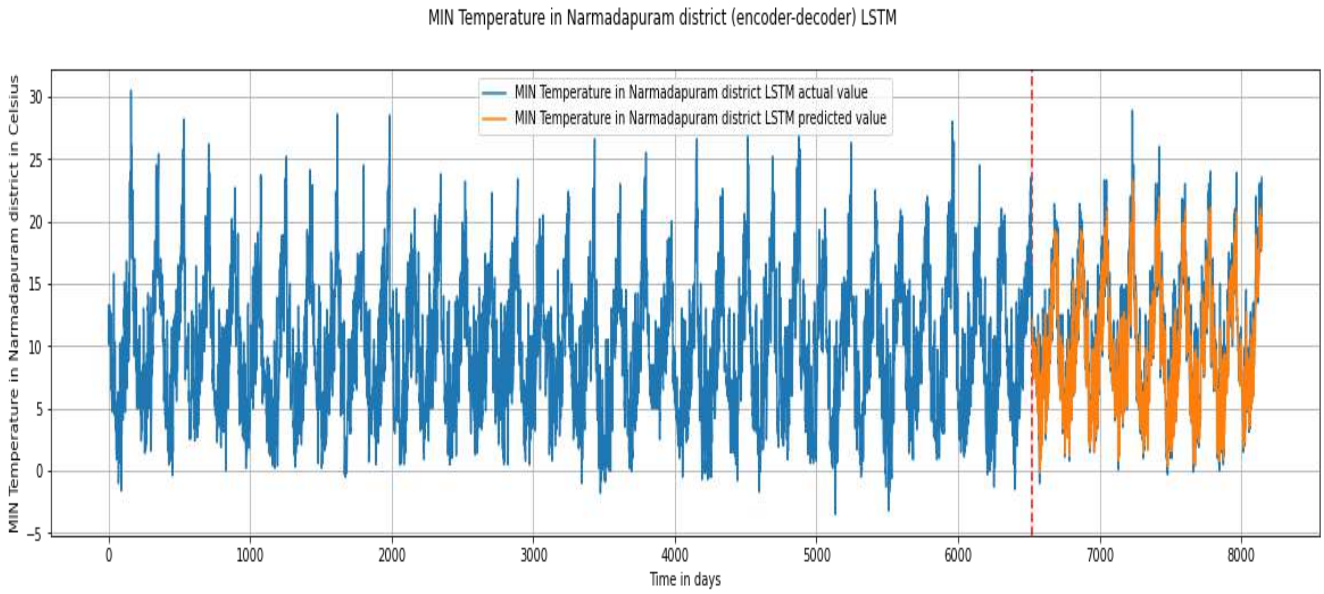

Figure 10: Forecasted vs. Actual Minimum Temperature using Encoder-Decoder LSTM Forecasting in whole dataset.

Doi: https://doi.org/10.54216/IJAACI.010201
Received: January 10, 2022 Accepted: May 27, 2022

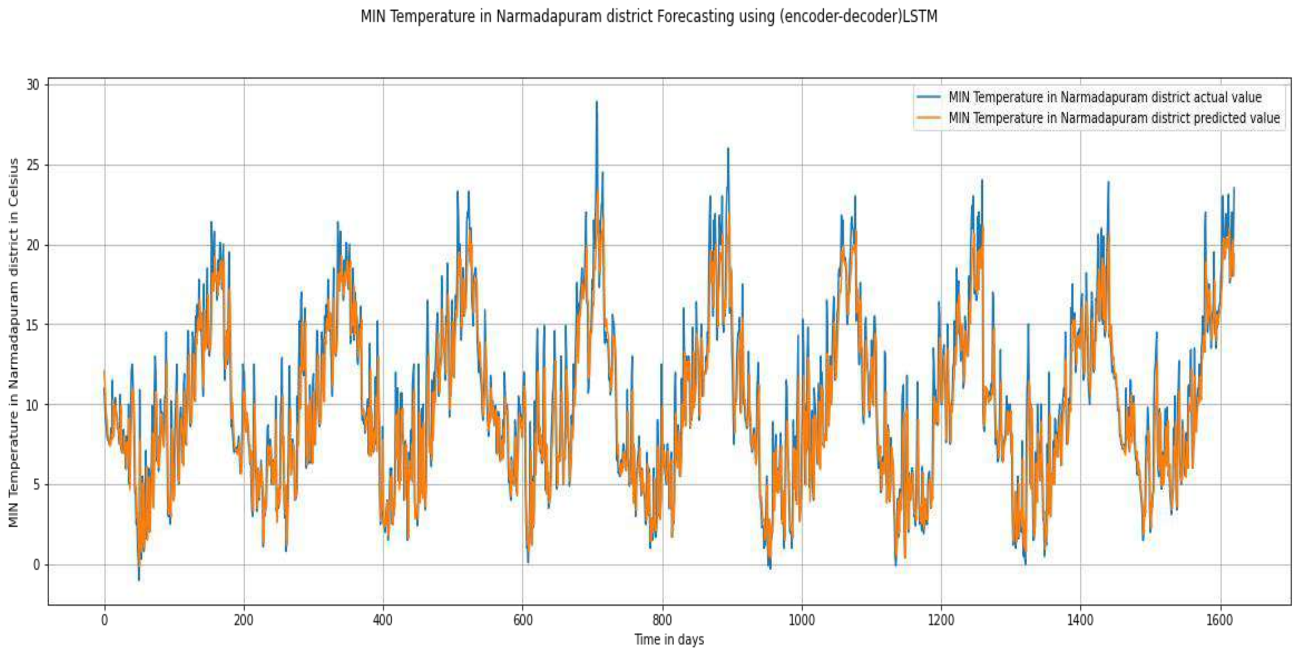

Figure 11: Forecasted vs. Actual Minimum Temperature using Encoder-Decoder LSTM Forecasting in testing dataset

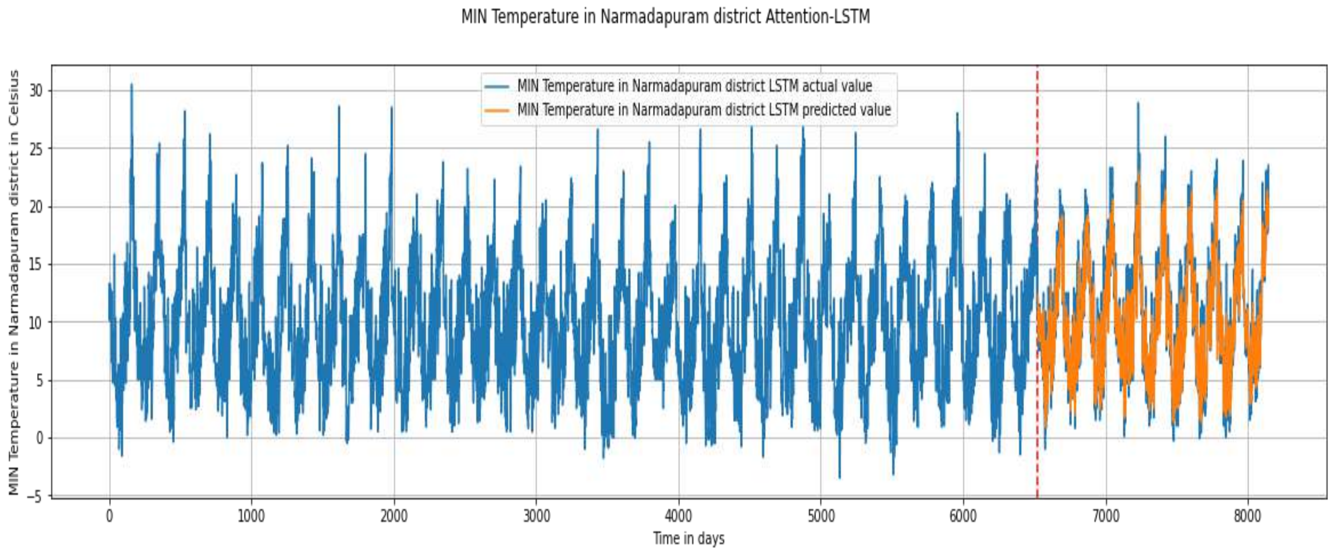

Figure 12: Forecasted vs. Actual Minimum Temperature using Attention LSTM Forecasting in whole dataset

Doi: https://doi.org/10.54216/IJAACI.010201
Received: January 10, 2022 Accepted: May 27, 2022

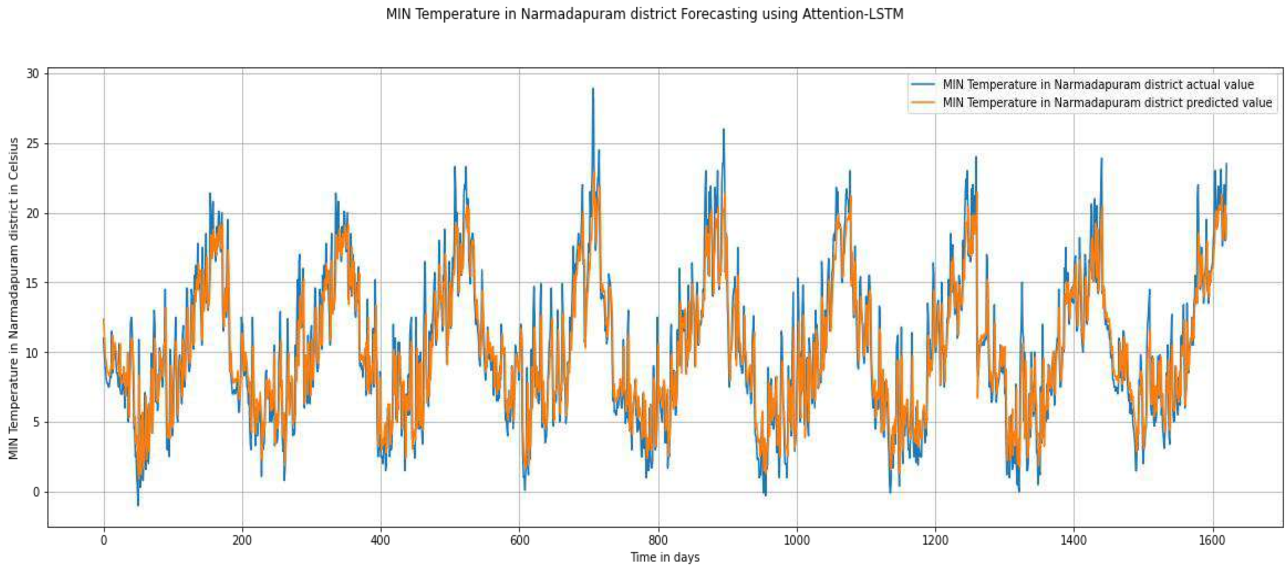

Figure 13: Forecasted vs. Actual Minimum Temperature using Attention LSTM Forecasting in testing dataset

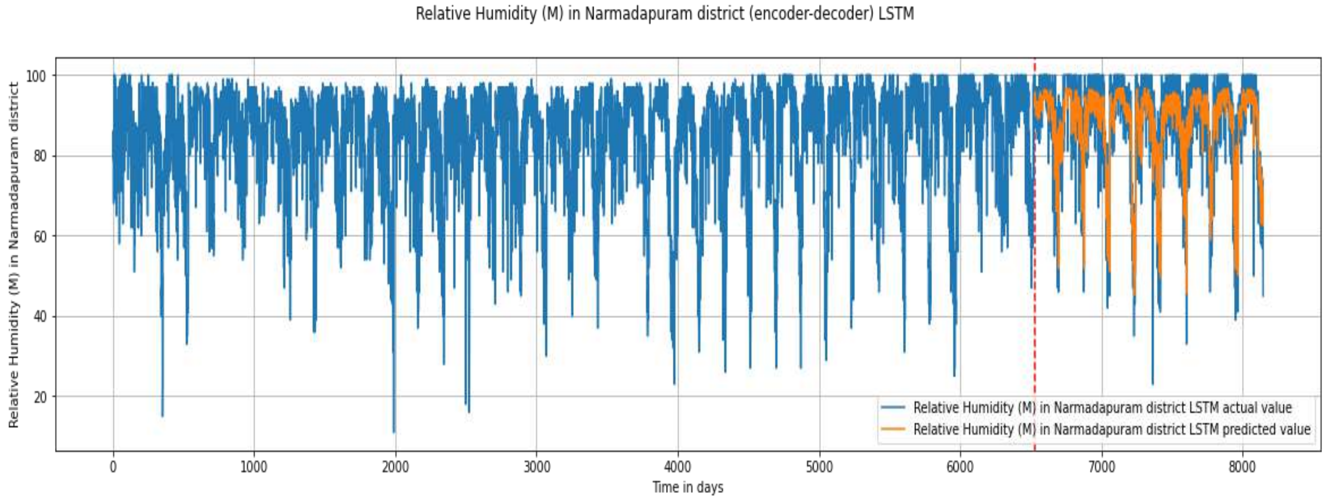

Figure 14. Forecasted vs. Actual Relative Humidity (Morning) using Encoder-Decoder LSTM Forecasting in whole dataset

Doi: https://doi.org/10.54216/IJAACI.010201
Received: January 10, 2022 Accepted: May 27, 2022

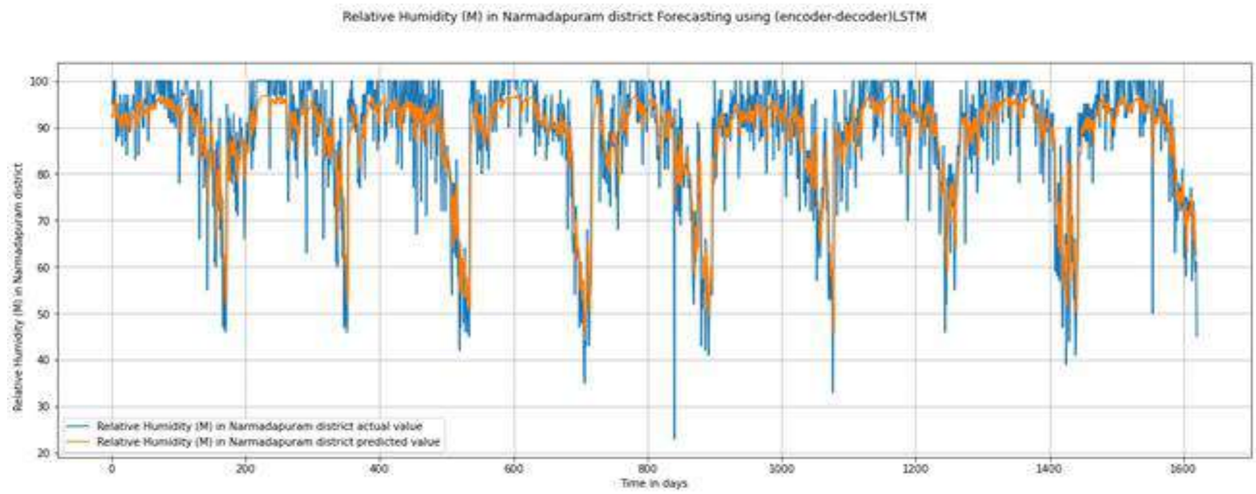

Figure 15: Forecasted vs. Actual Relative Humidity (Morning) using Encoder-Decoder LSTM Forecasting in testing dataset

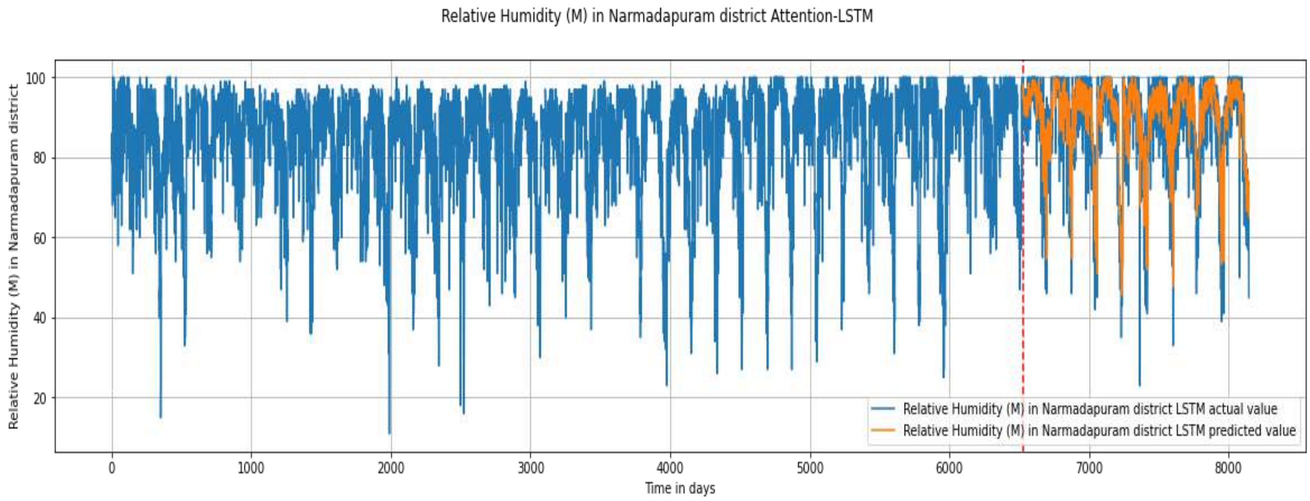

Figure 16: Forecasted vs. Actual Relative Humidity (Morning) using Attention LSTM Forecasting in whole dataset

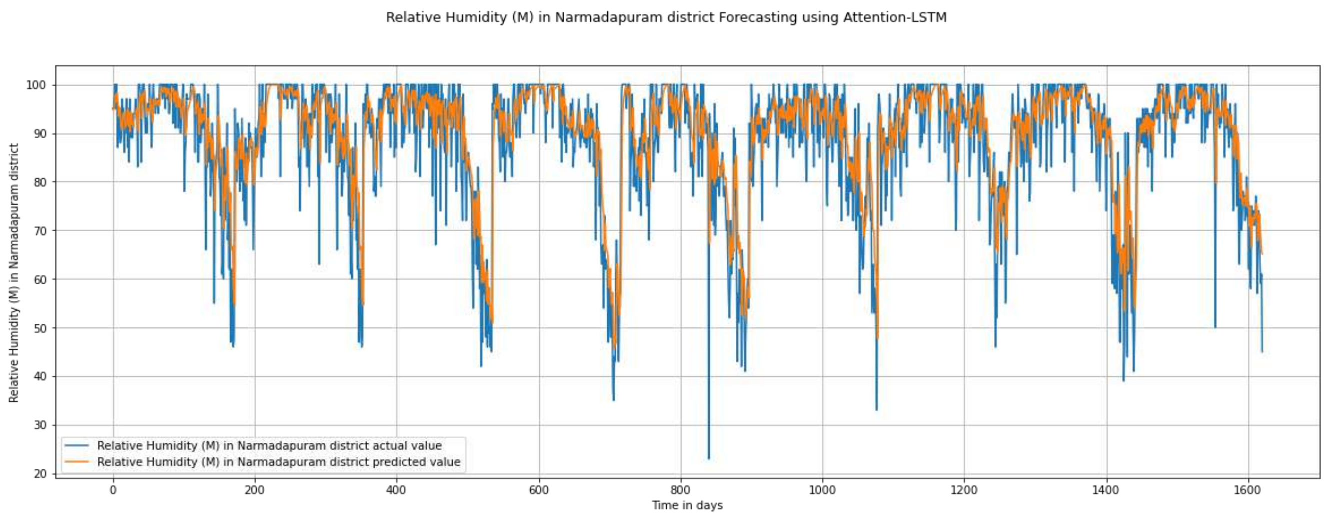

Figure 17: Forecasted vs. Actual Relative Humidity (Morning) using Attention LSTM Forecasting in testing dataset

Doi: https://doi.org/10.54216/IJAACI.010201
Received: January 10, 2022 Accepted: May 27, 2022

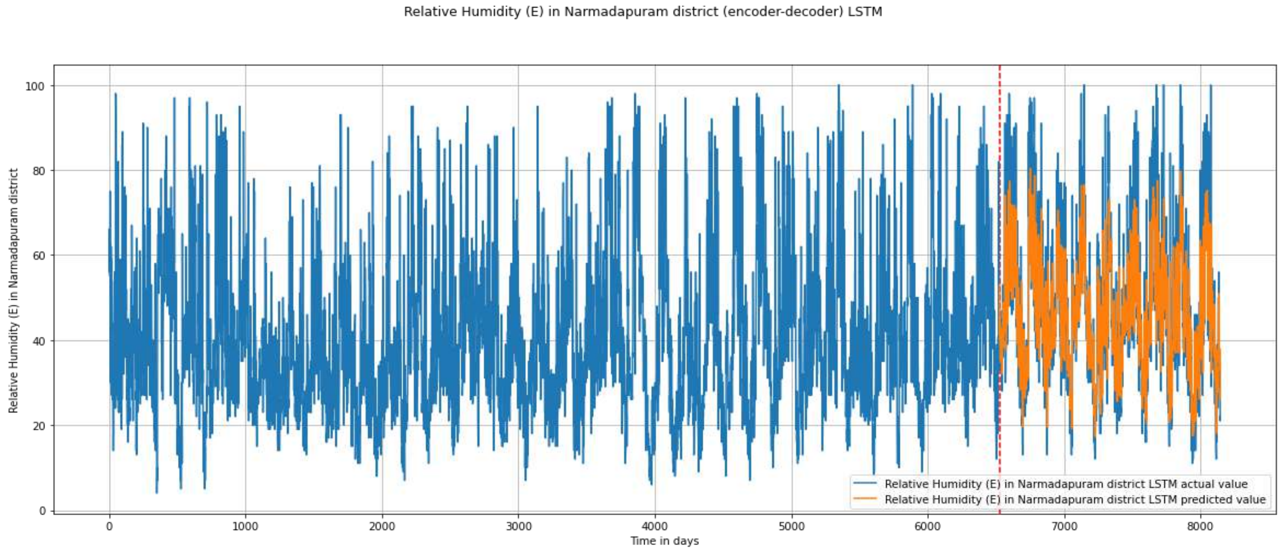

Figure 18: Forecasted vs. Actual Relative Humidity (Evening) using Encoder-Decoder LSTM Forecasting in whole dataset

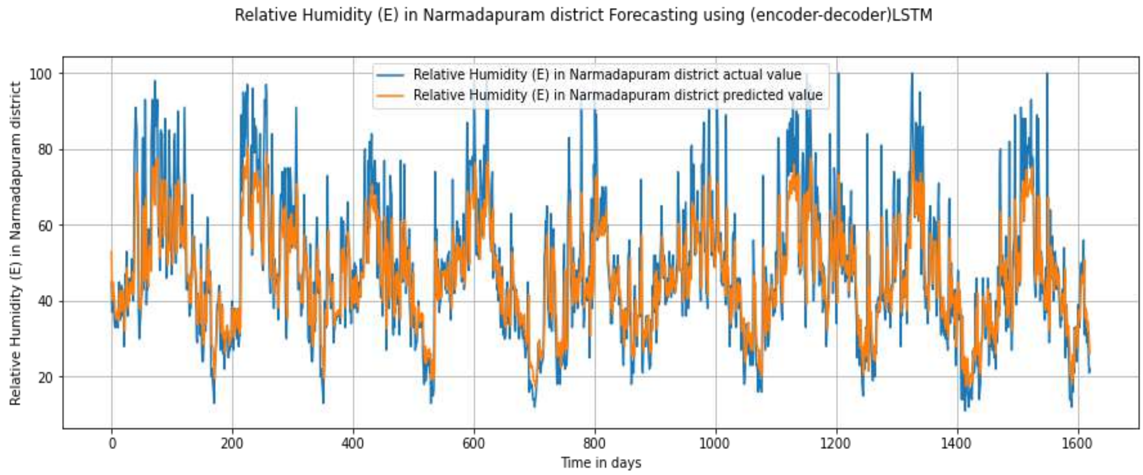

Figure 19 : Forecasted vs. Actual Relative Humidity (Evening) using Encoder-Decoder LSTM Forecasting in testing dataset

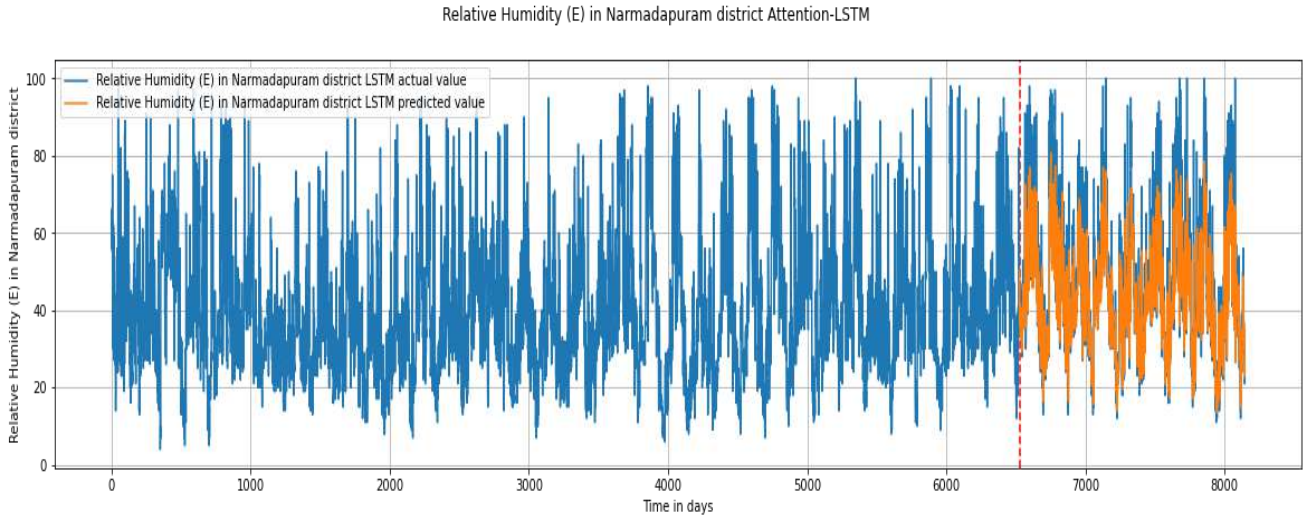

Figure 20: Forecasted vs. Actual Relative Humidity (Evening) using Attention LSTM Forecasting in whole dataset

Doi: https://doi.org/10.54216/IJAACI.010201
Received: January 10, 2022 Accepted: May 27, 2022

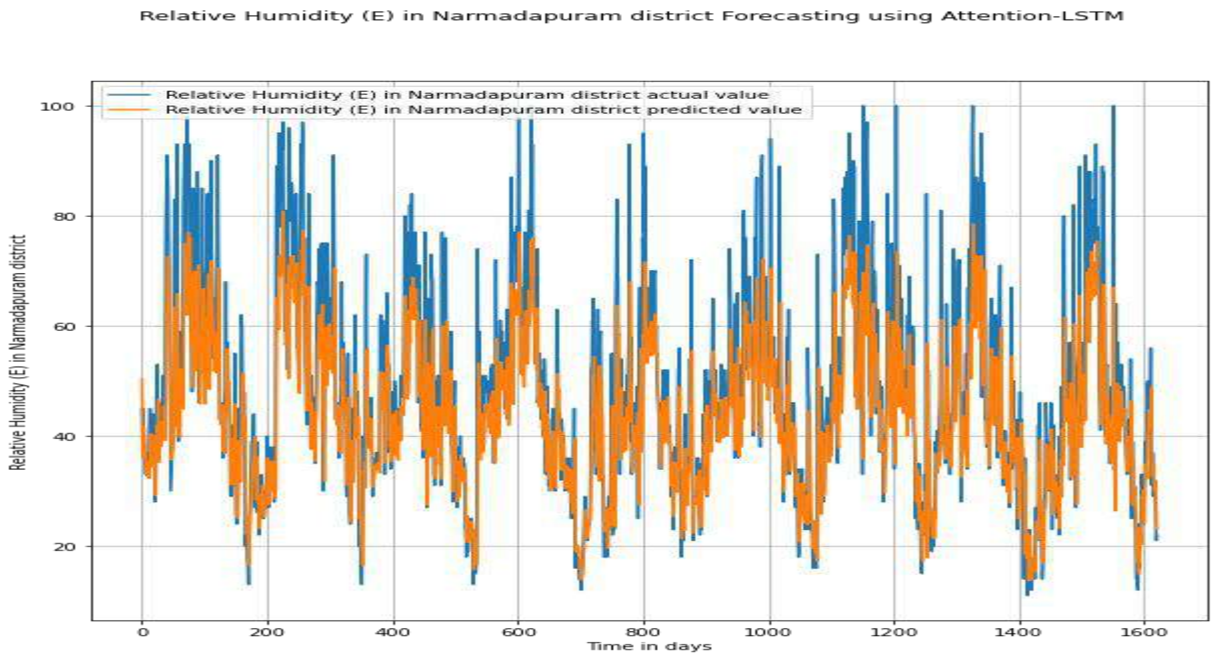

Figure 21: Forecasted vs. Actual Relative Humidity (Evening) using Attention LSTM Forecasting in testing dataset

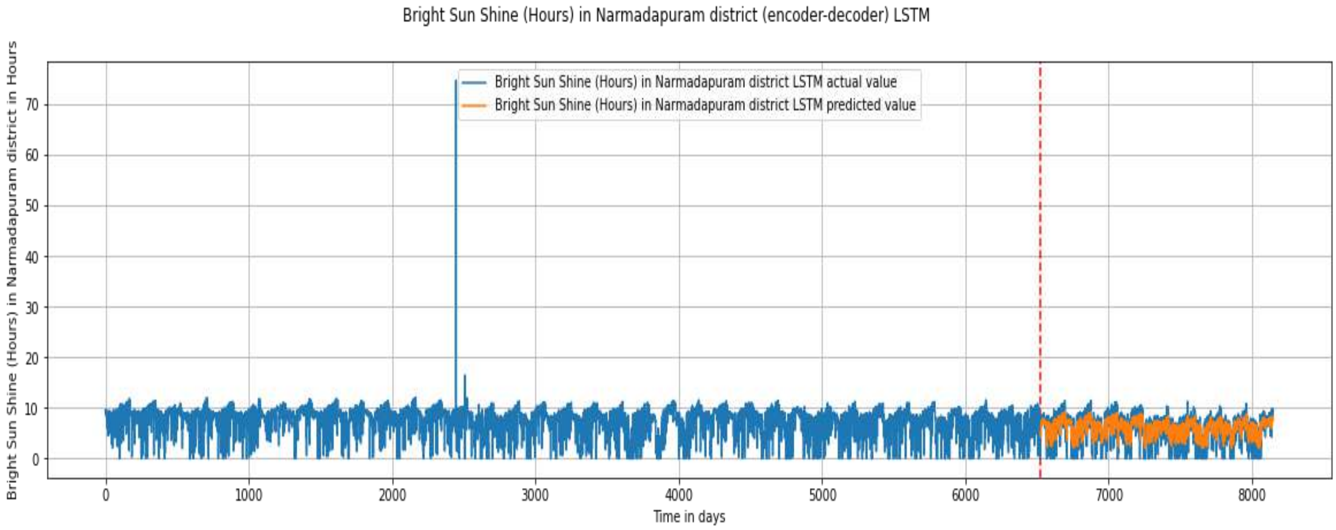

Figure 22: Forecasted vs. Actual Bright Sun Shine Hours using Encoder-Decoder LSTM Forecasting in whole dataset

Doi: https://doi.org/10.54216/IJAACI.010201
Received: January 10, 2022 Accepted: May 27, 2022

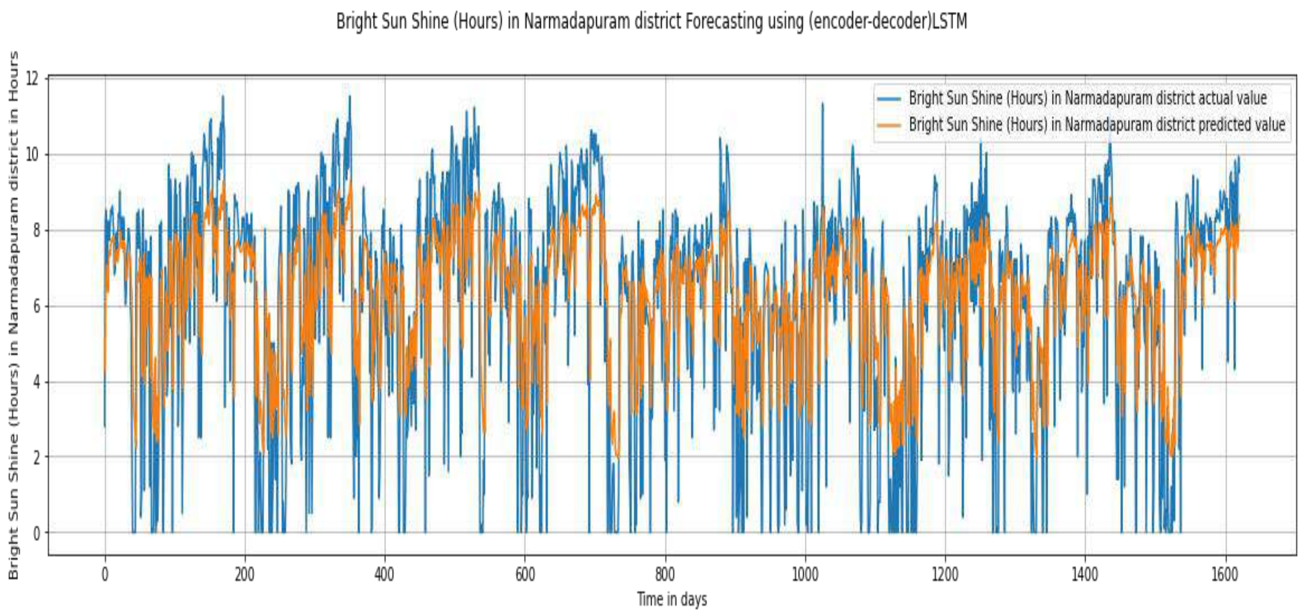

Figure 23: Forecasted vs. Actual Bright Sun Shine Hours using Encoder-Decoder LSTM Forecasting in testing dataset.

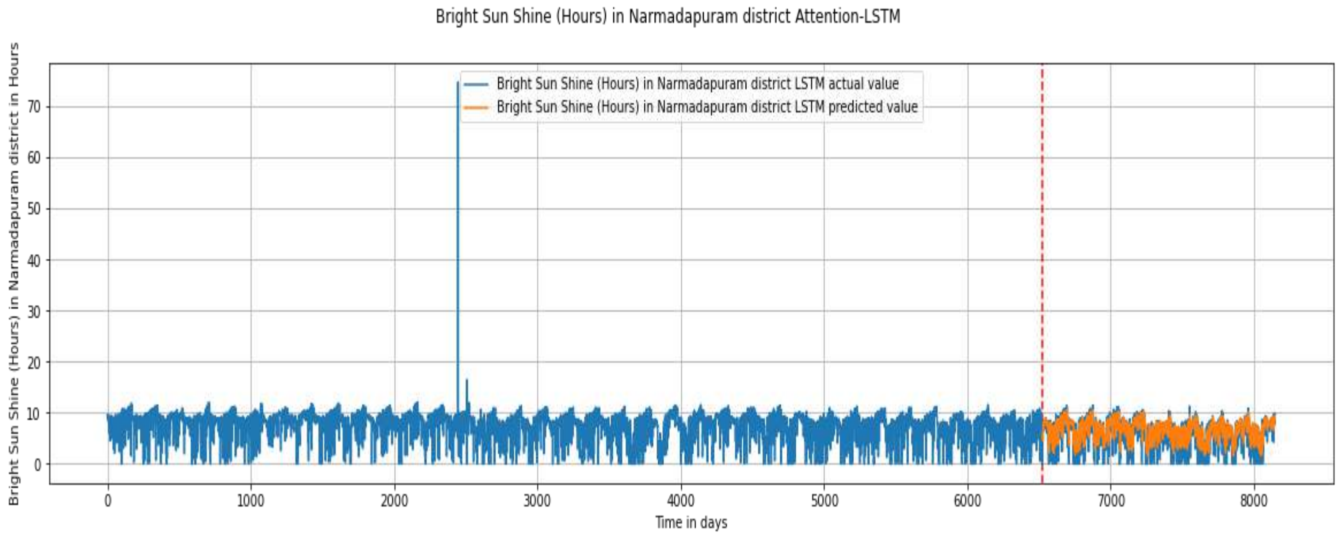

Figure 24: Forecasted vs. Actual Bright Sun Shine Hours using Attention LSTM Forecasting in whole dataset

Doi: https://doi.org/10.54216/IJAACI.010201
Received: January 10, 2022 Accepted: May 27, 2022

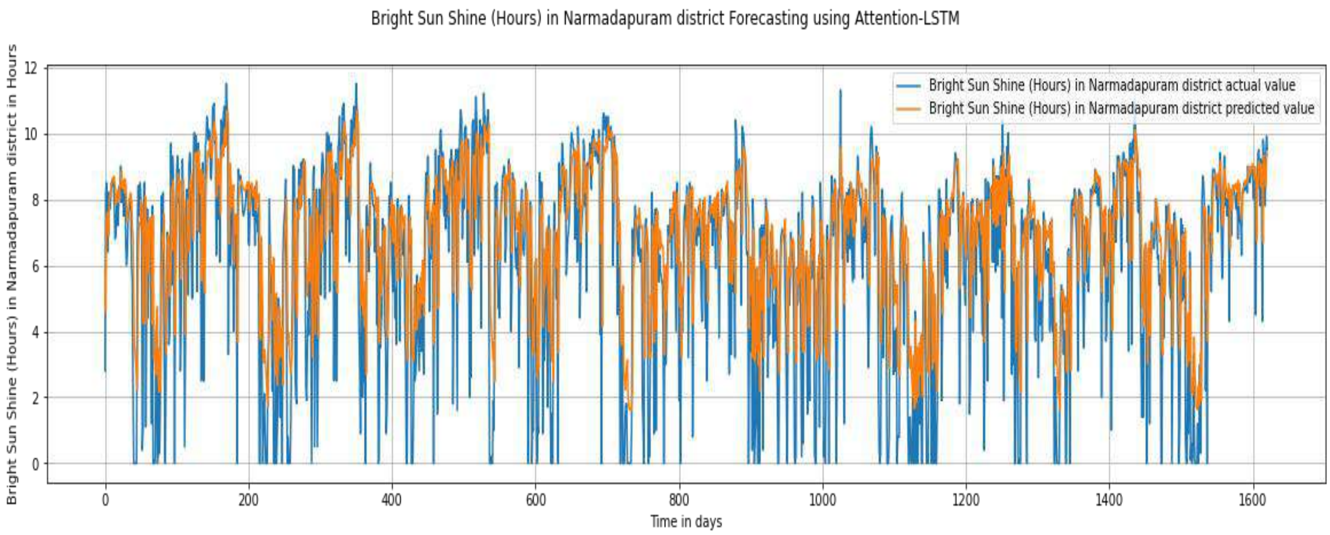

Figure 25: Forecasted vs. Actual Bright Sun Shine Hours using Attention LSTM Forecasting in testing dataset

5. Conclusions

The study highlights the effectiveness of using AI and ML techniques, including Encoder-Decoder LSTM and Attention LSTM models, in analyzing and predicting weather patterns in the Narmadapuram district. The results demonstrate the superiority of Encoder-Decoder LSTM in forecasting maximum temperature, morning relative humidity, evening relative humidity, and bright sunshine hours, while Attention LSTM outperforms in forecasting minimum temperature. These findings validate the potential of machine learning in addressing the challenges of urban development in smart cities, specifically in the area of weather forecasting. The results highlight the importance of identifying and utilizing the right machine learning model for a specific forecasting task, and can inform the development of more precise and accurate weather forecasting models in the future.

Funding: “This research received no external funding”

Conflicts of Interest: “The authors declare no conflict of interest.”

References

[1] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural computation, vol. 9, no. 8, pp. 1735-1780, 1997.

[2] R. F. Engle, "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation," Econometrica: Journal of the econometric society, pp. 987-1007, 1982.

[3] T. Bollerslev, "Generalized autoregressive conditional heteroskedasticity," Journal of econometrics, vol. 31, no. 3, pp. 307-327, 1986.

[4] T. H. Naylor, T. G. Seaks and D. W. Wichern, "Box-Jenkins methods: An alternative to econometric models," International Statistical Review/Revue Internationale de Statistique, pp. 123-137, 1972.

[5] Z. Shen, Q. Wan and D. J. Leatham, "Bitcoin Return Volatility Forecasting: A Comparative Study between GARCH and RNN," Journal of Risk and Financial Management, vol. 14, no. 7, p. 337, 2021.

[6] H. Y. Kim and C. H. Won, "Forecasting the volatility of stock price index: A hybrid model integrating LSTM with multiple GARCH-type models," Expert Systems with Applications, vol. 103, pp. 25-37, 2018.

[7] R. Zhang, Z. Guo, Y. Meng, S. Wang, S. Li, R. Niu and Y. Li, "Comparison of ARIMA and LSTM in Forecasting the Incidence of HFMD Combined and Uncombined with Exogenous Meteorological Variables in Ningbo, China," International journal of environmental research and public health, vol. 16, no. 11, p. 6174, 2021.

Doi: https://doi.org/10.54216/IJAACI.010201
Received: January 10, 2022 Accepted: May 27, 2022

[8] A. Kochhar, H. Singh, S. Sahoo, P. K. Litoria and B. Pateriya, "Prediction and forecast of pre-monsoon and post-monsoon groundwater level: using deep learning and statistical modelling," Modeling Earth Systems and Environment, vol. 8, no. 2, pp. 2317-2329, 2022.

[9] A. Moghar and M. Hamiche, "Stock market prediction using LSTM recurrent neural network," Procedia Computer Science, vol. 170, pp. 1168-1173, 2020.

[10]B. Yang, S. Sun, J. Li, X. Lin and Y. Tian, "Traffic flow prediction using LSTM with feature enhancement," Neurocomputing, vol. 332, pp. 320-327, 2019.

[11]R. A. Rajagukguk, R. A. Ramadhan and H. J. Lee, "A review on deep learning models for forecasting time series data of solar irradiance and photovoltaic power," Energies, vol. 13, no. 24, p. 6623, 2020.

[12]P. Liu, J. Wang, A. K. Sangaiah, Y. Xie and X. Yin, "Analysis and prediction of water quality using LSTM deep neural networks in IoT environment," Sustainability, vol. 11, no. 7, p. 2058, 2019.

[13]M. Mele, A. R. Gurrieri, G. Morelli and C. Magazzino, "Nature and climate change effects on economic growth: an LSTM experiment on renewable energy resources," Environmental Science and Pollution Research, vol. 28, pp. 41127-41134, 2021.

[14]M. A. R. Suleman and S. Shridevi, "Short-Term Weather Forecasting Using Spatial Feature Attention Based LSTM Model," IEEE Access, vol. 10, pp. 82456-82468, 2022.

[15]Y. Liu, D. Li, S. Wan, F. Wang, W. Dou, X. Xu and L. Qi, "A long short-term memory-based model for greenhouse climate prediction," International Journal of Intelligent Systems, vol. 37, no. 1, pp. 135-151, 2022.

[16]D. Lee, G. Lee, S. Kim and S. Jung, "Future runoff analysis in the Mekong river basin under a climate change scenario using deep learning," Water, vol. 12, no. 6, p. 1556, 2020.

[17]M. Gao, J. Li, F. Hong and D. Long, "Day-ahead power forecasting in a large-scale photovoltaic plant based on weather classification using LSTM," Energy, vol. 187, p. 115838, 2019.

[18]X. Qing and Y. Niu, "Hourly day-ahead solar irradiance prediction using weather forecasts by LSTM," Energy, vol. 148, pp. 461-468, 2018.

[19]Z. Karevan and J. A. Suykens, "Transductive LSTM for time-series prediction: An application to weather forecasting," Neural Networks, vol. 125, pp. 1-9, 2020.

[20]Y. Wu, Y. Ding, Y. Zhu, J. Feng and S. Wang, "Complexity to forecast flood: Problem definition and spatiotemporal attention LSTM solution," Complexity, 2020.

[21]M. A. Zaytar and C. El Amrani, "Sequence to sequence weather forecasting with long short-term memory recurrent neural networks," International Journal of Computer Applications, vol. 143, no. 11, pp. 7-11, 2016.

[22]I. Sutskever, O. Vinyals and Q. V. Le, "Sequence to sequence learning with neural networks," Advances in neural information processing systems, p. 27, 2014.

[23]Z. Xiang, J. Yan and I. Demir, "A rainfall-runoff model with LSTM-based sequence-to-sequence learning," Water resources research, vol. 56, no. 1, p. e2019WR025326, 2020.

[24]J. H. Bappy, C. Simons, L. Nataraj, B. S. Manjunath and . A. K. Roy-Chowdhury, "Hybrid lstm and encoder–decoder architecture for detection of image forgeries," IEEE Transactions on Image Processing, vol. 28, no. 7, pp. 3286-3300, 2019.

[25]Y. Wang, M. Huang, X. Zhu and L. Zhao, "Attention-based LSTM for aspect-level sentiment classification," In Proceedings of the 2016 conference on empirical methods in natural language processing, pp. 606-615, 2016.

[26]J. R. Jiang, J. E. Lee and Y. M. Zeng, "Time series multiple channel convolutional neural network with attention-based long short-term memory for predicting bearing remaining useful life," Sensors, vol. 20, no. 1, p. 166, 2019.

[27]S. Hochreiter and J. Hochreiter, "Long short-term memory," Neural computation, vol. 8, no. 9, pp. 1735-1780, 1977.

[28]F. A. Gers, D. Eck and J. Schmidhuber, "Applying LSTM to time series predictable through time-window approaches," Neural Nets WIRN Vietri-01, pp. 193-200, 2002.

Doi: https://doi.org/10.54216/IJAACI.010201
Received: January 10, 2022 Accepted: May 27, 2022

[29]H. Huynh, L. M. Dang and . D. Duong, "A new model for stock price movements prediction using deep neural network," In Proceedings of the Eighth International Symposium on Information and Communication Technology, pp. 57-62, 2017.

[30]X. Zhang, X. Liang, A. Zhiyuli, S. Zhang, R. Xu, Z. Cheng and et al, "AT-LSTM: An attention-based LSTM model for financial time series prediction," In IOP Conference Series: Materials Science and Engineering, vol. 569, no. 5, p. 052037, july 2019.

[31]X. Song, Y. Liu, L. Xue, J. Wang, J. Zhang, J. Wang, Z. Cheng and et al, "Time-series well performance prediction based on Long Short-Term Memory (LSTM) neural network model," Journal of Petroleum Science and Engineering, vol. 186, 2020.

[32]G. Van Houdt, C. Mosquera, G. Nápoles and et al, "A review on the long short-term memory model," Artif Intell Rev, vol. 53, p. 5929–5955, 2020.

[33]D. S. Reddy and P. Prasad, "Prediction of vegetation dynamics using NDVI time series data and LSTM," Modeling Earth Systems and Environment, vol. 4, no. 1, pp. 409-419, 2018.

[34]N. Zeydalinejad, "Artificial neural networks vis-à-vis MODFLOW in the simulation of groundwater: A review.," Modeling Earth Systems and Environment, pp. 1-22, 2022.

[35]S. Wang, L. Mu and D. Liu, "A hybrid approach for El Niño prediction based on Empirical Mode Decomposition and convolutional LSTM Encoder-Decoder," Computers & Geosciences, vol. 149, p. 104695, 2021.

[36]S. H. Park, B. Kim, C. M. Kang, C. C. Chung and J. W. Choi, "Sequence-to-sequence prediction of vehicle trajectory via LSTM encoder-decoder architecture," IEEE Intelligent Vehicles Symposium, pp. 1672-1678, 2018.

[37]M. J. Ellis and V. Chinde, "An encoder–decoder LSTM-based EMPC framework applied to a building HVAC system," Chemical Engineering Research and Design, vol. 160, pp. 508-520, 2020.

[38]B. Zhang, G. Zou, D. Qin, Y. Lu, Y. Jin and H. Wang, "A novel Encoder-Decoder model based on read-first LSTM for air pollutant prediction," Science of The Total Environment, vol. 765, p. 144507, 2021.

[39]S. Kim and M. Kang, "Financial series prediction using Attention LSTM," arXiv preprint arXiv, pp. 1902-10877, 2019.

[40]G. Liu and J. Guo, "Bidirectional LSTM with attention mechanism and convolutional layer for text classification," Neurocomputing, pp. 325-338, 2019.

[41]S. Biswas and M. Sinha, "Performances of deep learning models for Indian Ocean wind speed prediction," Modeling Earth Systems and Environment, vol. 7, no. 2, pp. 809-831, 2021.

[42]Y. Li, Z. Zhu, D. Kong, H. Han and Y. Zhao, "EA-LSTM: Evolutionary attention-based LSTM for time series prediction," Knowledge-Based Systems, vol. 181, p. 104785, 2019.

Doi: https://doi.org/10.54216/IJAACI.010201
Received: January 10, 2022 Accepted: May 27, 2022