2022 4th International Conference on Sustainable Technologies for Industry 4.0 (STI), 17-18 December, Dhaka

# Implementation of Encoder-Decoder based Long Short-Term Memory Network for Short-Term Electrical Load Forecasting

Amit Deb Roy*, Ashfak Yeafi*
*Department of Electrical and Electronic Engineering, Khulna University of Engineering and Technology, Khulna 9203, Bangladesh
Email: roy1603088@stud.kuet.ac.bd, yeafiashfak@gmail.com

***Abstract* - Electrical load forecasting has extensive importance in power system designing and optimization. Keeping pace with the growing economy and moving towards a digitalized network, Bangladesh Power System (BPS) has to be renovated in terms of efficient load forecasting strategies. To meet the issue, this research proposes an effective short-term load forecasting (STLF) technique based on a unique form of recurrent neural network (RNN) known as the Long Short-Term Memory (LSTM) network. A new dataset gathered from a public owned power company of Bangladesh, is applied into the model which will help to get a better and recent understanding of BPS load. Employing this viable method, one hour ahead electrical load of BPS can be forecasted easily with a minimized error rate (MAPE 2.29%). In addition, Random Forest (RF) regression model is used to compare with the forecasting results of proposed LSTM. Again, the proposed LSTM model is used to get predictions for the different months of a given year to visualize how much load fluctuation of BPS load takes place in different seasons. In case of STLF, the prediction outcomes on the Bangladesh's electricity load demand, indicate that the proposed model can work flawlessly in grid optimization with greater precision.**

***Index Terms - Short-Term Load Forecasting, Long-Short-Term Memory Network, Bangladesh Power System, Evaluation Metrics***

## I. INTRODUCTION

To ensure a sustainable operational strategy and control system, effective short and long-term load demand prediction is necessary for all the power concerned companies.

Here, short-term load forecasting is considered a pivotal experimental area for ensuring cost effective dispatch of power plants. Notable advantages ensured by electricity dispatch are optimal usage of electricity that is generated utilizing modern facilities, power generation's cost reduction, and decreasing the possibilities of load-shedding [1]. Various models are used for years to execute short-term load forecasting (STLF) like in time series based methods, typical statistical models are widely considered in use. These include the Box-Jenkins model [2], autoregressive integrated moving average model (ARIMA) [3], autoregressive moving average model (ARMA) [4], linear regression [5], exponential smoothing [6], etc. algorithms. But the prediction performance of those models does not show satisfactory results in STLF [7].

To mitigate these disadvantages, later on artificial intelligence based forecasting model is evolved which has been enormously used in STLF [8]. Clustering methods, fuzzy logic control [9], support vector machine (SVM), artificial neural network (ANN), and random forest (RF) are the notable strategies of superficial machine learning algorithms [10]. Noteworthy, recurrent neural network (RNN) is introduced to operate on sequence based information and permit algorithm to train the sequentially connected data for flawless memorization [11]. However, the prime drawbacks of RNN like vanishing and exploding gradient problem causing reduced accuracy, gave motivation to develop the LSTM network later.

### *A. Related Works*

Addressing time series issues, the LSTM architecture is introduced in STLF [12]. In [13] LSTM based framework is proposed for individual residual load forecasting where the single-meter power prediction problem is solved after testing several benchmarks. A LSTM based RNN scheme is implemented in [14] to forecast the non-linear, univariate, time series electricity load data utilizing the long duration dependencies in the time series. In [15] authors proposed a LSTM based forecasting framework in which appliance consumption learning helped to overcome the challenges of residents' volatile load nature. Again, a novel forecasting scheme is proposed to implement the policy of replacing coal by electricity with the help of LSTM model which takes temperature and wind force as inputs to avoid the shortcomings of exploding gradient [16]. Future peak load demand of Indonesia has been predicted by an exclusive approach of the LSTM network using the historical load data collected from the area sub-station [17]. This forecasting scheme showed better performance for weekdays load calculation after processing the univariate load time series. In [18] a day ahead short-term power prediction is done for bivariate time series using LSTM employed in Peshawar Electric Supply Company (PESCO) where model's weights are determined through three supervised learning algorithms. In Bangladesh, most of the load forecasting researches have been prioritized on region wise loads using traditional methods [19]-[20]. Though some of the neural network based architectures are used in BPS load prediction, LSTM based researches are quite rare focusing on daily, weekly, and monthly load forecasting in BPS network [21]. In this paper, the proposed LSTM architecture may contribute to BPS significantly as it shows a better prediction accuracy using novel model features.

### *B. Paper Contribution*

The following are the key contributions of this research in brief:

- The prime contribution of the paper lies in its introduction of the application of customized encoder-decoder based LSTM model in electrical load forecasting to leverage its sequence learning capability.

- Very few researches considered seasonal load fluctuation of BPS. We experimented different months load predictions to visualize the changing load nature and the proposed model has illustrated the seasonal load behaviors of Bangladesh with least MAPE values.
- The developed methodology shows a high accuracy of 97.71% in STLF. So, it can extensively be used in future grid planning and optimization of Bangladesh as the used load data till 2021 can ensure the better understandings of recent load trends.

The next section will describe the proposed methodology of this paper in details.

## II. METHODOLOGY

### A. *LSTM Architecture*

LSTM network is a special sort of neural network responsible for remembering data for a longer period of time in its memory unit.

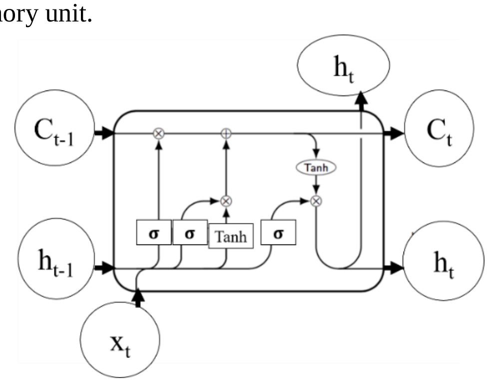


Fig. 1. LSTM architecture with its functional gates

LSTM network consists of four major components. They are - cell state, forget gate, input gate, and output gate. Here, fig. 1. depicts the fundamental LSTM architecture with its gates.

Cell State: Different gates carefully remove information from the cell state and also carry new data to the cell state.

Forget Gate: The main step is to make a decision about which information the cell state will eliminate. Basically, the sigmoid layer looks at $h_{t-1}, x_t$ and gives output a number ranging from 0 to 1 considering every value in cell state $C_{t-1}$.

$$f_t = \sigma\,(W_f\,.[h_{t-1}, x_t] + b_f) \tag{1}$$

here, $W_f$ is the weight matrix, $b_f$ is the biased value of forget gate, and $\boldsymbol{\sigma}$ is the sigmoid activation function.

Input Gate: A vector, $\overrightarrow{C_t}$ is created for the new stored value by the tanh layer that will be appended to the cell state.

$$i_t = \sigma\,(W_i\,.[h_{t-1}, x_t] + b_i) \tag{2}$$

$$\overrightarrow{C_t} = \tanh(\,W_c\,.[h_{t-1}, x_t] + b_c\,) \tag{3}$$

where, $i_t$ stands for input gate and $b_i$ is the biased value of input gate. Now, the previous value of $C_{t-1}$ will be transformed to current $C_t$ and the equation will be :

$$C_t = f_t * C_{t-1} + i_t * \overrightarrow{C_t} \tag{4}$$

Output Gate: Here, both the sigmoid and tanh layer perform in processing the values of cell state that are put through the tanh layer and thus, the desired information are obtained.

$$o_t = \sigma\,(W_{ou}\,.[h_{t-1}, x_t] + b_{ou}) \tag{5}$$

$$h_t = o_t * \tanh(C_t) \tag{6}$$

Here, $o_t$ refers to the output gate, $W_{ou}$ depicts weight matrix, $b_{ou}$ indicates biased value of output gate and $h_t$ is the output value for t duration.

That's how, LSTM extensively controls the information flow through its different gates.

### B. *Proposed Model*

To forecast the electrical load, a customized encoder-decoder based LSTM framework is implemented in our research. According to the design, there are three sections in the model. They are- encoder section, intermediate encoder vector and the decoder block. The proposed LSTM model for STLF is shown in fig. 2.

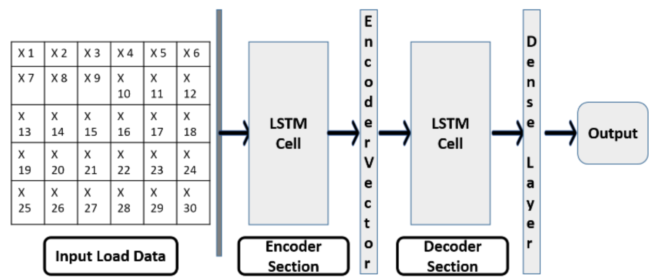


Fig. 2. Developed LSTM model for STLF

To implement the encoder model, one or more LSTM layers can be utilized. In a layer, 50 LSTM units are used. After normalization, converted array load elements are passed through the encoder block. Each LSTM unit then accepts elements of the input sequences, takes information of the load data and propagates them to the further section. Here, $H_t$ defines the encoder block's hidden state. Appropriate weights are applied to earlier hidden state $h_{t-1}$, then to input vector $x_t$.

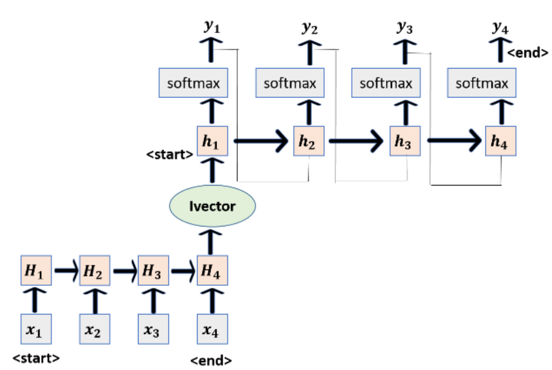


Fig. 3. Sequential parameters of the developed LSTM model

The encoder vector aims to incorporate all input element information in order to aid the decoder in making predictions with great accuracy. It serves as the model's initial hidden layer for the decoder part.

The decoder block reads the encoder vector and tries to predict the output sequence. In this block, 50 LSTM units are used according to our design and 100 timedistributed dense units are added after the decoder block. The hidden state of decoder is defined by $h_t$. Now, each LSTM unit creates an output value along with a hidden state of its own after getting previous hidden state from earlier LSTM block according to the above fig. 3. Here, total of 20,200 parameters are trainable for getting updated in weights and like encoder block, return sequences are given true.

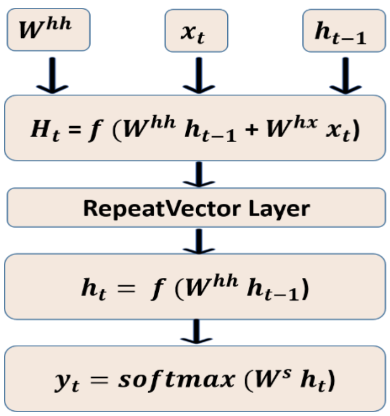


Fig. 4. Functional calculation of each block in proposed model

Eventually, we have computed the outputs by combining the present $h_t$ with the weight $W^s$. Here, *softmax* function is used to generate a probability vector that will aid in the prediction of the ultimate result of load values. All the formulations that we have used in our encoder-decoder model are represented sequentially in the fig. 4 mentioning the input vectors, weights and the hidden states.

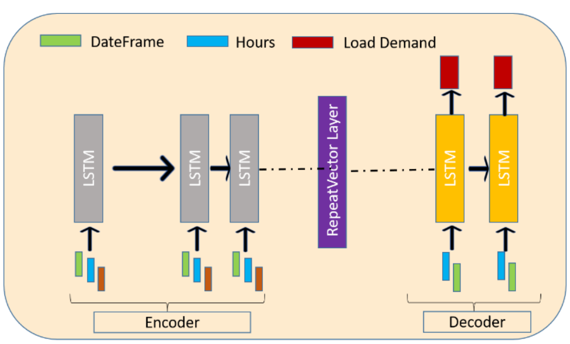


Fig. 5. Processed outcome from encoder-decoder section

Analyzing historical load data, date frame and hours in the encoder block, predicted electrical load is observed from the decoder's LSTM units. Considering multiple steps time series, forecasting of (T, T+1, T+2) months' demand is possible from the given observation of (T-1, T-2, T-n) months' demand. Here, required time steps for input pattern considered as three and one for output.

## C. *Dataset and Data Framing*

For STLF, electrical load data of the last three and half years with hourly information are collected from the website of a public company called Power Grid Company of Bangladesh (PGCB) [22]. As mentioned, the data from January 1, 2018 to March 15, 2021 have been used in the dataset. And, the total number of used observations is 28,080.

The whole dataset is segmented into some features such as: date, year, month, day, weekday, hours and the load demand. Noteworthy, the dataset is split into many weeks which start from Sunday and end on Saturday. Here, in the 'Week Day' column, the number '1' represents Sunday and '7' refers to Saturday. Ten data from the dataset is shown below in TABLE I. as samples:

TABLE I. Characteristics of sample data from the dataset

| Year | Month | Day | Week Day | Hour | Demand (MW) |
|---|---|---|---|---|---|
| 2021 | 3 | 1 | 2 | 0 am | 8942.65 |
| | | | | 1 am | 8495 |
| | | | | 2 am | 7992.93 |
| | | | | 3 am | 7791 |
| | | | | 4 am | 7584 |
| | | | | 5 am | 7475.5 |
| | | | | 6 am | 7543.8 |
| | | | | 7 am | 7890.84 |
| | | | | 8 am | 8383 |
| | | | | 9 am | 8704.22 |

Each observation is taken within an interval of an hour. Decomposition of these large load data leads to an improved forecasting accuracy and ensure the causal insights. To implement this, the *seasonal_decompose* function of *statsmodel* library is used.

Here, 65% data (January 1, 2018 to January 30, 2020) is chosen to train the proposed model and the rest 35% data for the test purposes. The data from February, 2020 to March, 2021 is kept to evaluate the model's performance. Basically, this last year contains the test data of 55 weeks. Recombining all these pre-process techniques can accurately interpret the model for better prediction.

## D. *Process Flow of the Proposed Method*

First of all, collecting a standard dataset having a good amount of data is necessary to build and train the model. Generally, historical load data is to be checked whether there are null values or not. Again, normalizing the load values is needed to make the neural network more stable. To make the load values normalized, *MinMaxScaler* imported from Scikit-learn, is used in the proposed model. Scaling the load values between 0 to 1 ensures speedy learning and also faster convergence. When the training step gets done, tuning the hyper-parameters may give better training results to some extent. Finally, the forecasting outcomes of the scheme are considered for comparison with the RF regression. The structured flow chart depicting the used model is represented below (fig. 6.) :

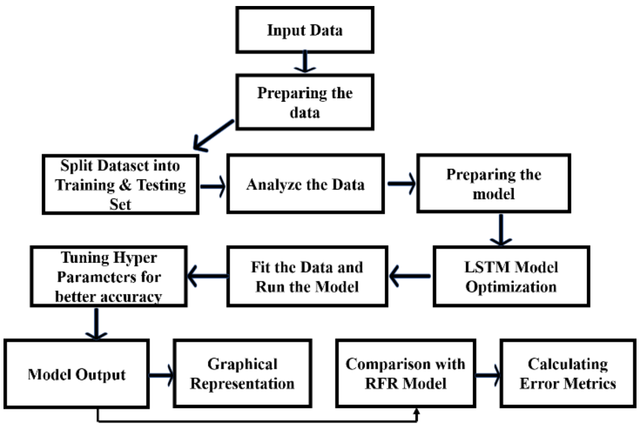


Fig. 6. Flow diagram of the implemented LSTM model

## E. *Building and Training the Proposed Model*

Here, the introduced LSTM based load forecasting model is developed by a free source high-level API library called Keras. Noteworthy, the LSTM model has to learn two sorts of traits; short term and long term, contained by the input data profiles and these features direct the LSTM model to perform accurately after training and testing. The hyper-parameters that are used to train the network are presented as follows :

TABLE II. Used hyper-parameters in the model

| Name | Hyper-parameters |
|---|---|
| Model | Sequential |
| Activation Function | Rectified Linear Unit (RELU) |
| Optimizer | Adam |
| Types of used layers | Dense and LSTM |
| No. of hidden layer | 3 for LSTM |
| No. of dense layer | 2 |
| No. of iteration (epochs) | 50 |
| Batch Size | 64 |
| Validation Split | 0.2 |
| Total parameters | 50,851 (All trainable) |
| Training time | 28.15 minutes |

All these hyper-parameters help significantly to reduce the over-fitting problem as the bias value is kept low and also, the low variance is selected in the proposed model.

# III. RESULTS AND ANALYSES

## A. *Performance Calculation Using Evaluation Metrics*

To measure the performance of implemented LSTM framework, error metrics or evaluation metrics are used frequently. There are three major evaluation metrics which are used as the performance indicators. Among them, the common two error metrics used for comparison purposes in our study are 'Mean Absolute Percentage Error (MAPE)' and 'Root Mean Squared Error (RMSE)'. A model can forecast more accurately if these error metrics show comparatively lower values. The mathematical expressions of the performance evaluation metrics are given below:

$$MAPE = \frac{\sum_{L=1}^{N}\left|\frac{(F_L - A_L)}{F_L}\right|}{N} \times 100\% \qquad (7)$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{L=1}^{N}(F_L - A_L)^2} \qquad (8)$$

here, $N$ refers the total data points in predicted load, $F_L$ indicates the forecasted demand values and $A_L$ represents the magnitude of actual loads at the time $t$.

## B. *Forecasting Outcomes*

The proposed encoder-decoder based LSTM scheme gives forecasting outcomes for various time frames such as 24 hours, 7 days, and one month with an hourly interval. There are 28,080 data in total in the dataset and they are taken at an hourly interval. According to the model design, 18,252 load data (i.e. from January 2018 to January 2020) is allocated in training the implemented framework and rest 9,828 data for the test purposes. The prediction outcomes of the proposed model on different criteria are represented graphically in fig. 7 to fig. 14.

Here, the RF regression model is used to compare with the forecasting results of the LSTM model. Basically, RF is used in non-liner characteristics data to minimize high variance.

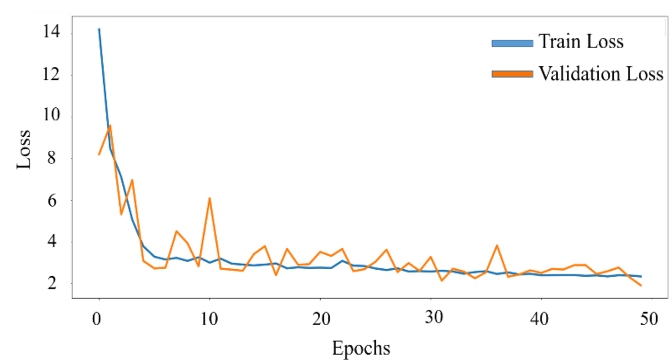


Fig. 7. Training loss curve of proposed model

In fig. 7. train loss and validation loss curves are decreasing with the number of epochs increasing. Here, a little bit of fluctuation also takes place along with some steady values. It is observed that validation loss gets smaller than the training loss when the epochs come closer to 50. And, the overfitting problems are reduced with the epochs going forward. Now, in fig. 8. prediction curves are shown using the test data from Feb 1, 2020 to March 15, 2021. In the prediction result, we observed that actual and predicted curves are overlapped in the majority portion and a few portion of the graph shows non-overlapping region which indicates the better prediction accuracy.

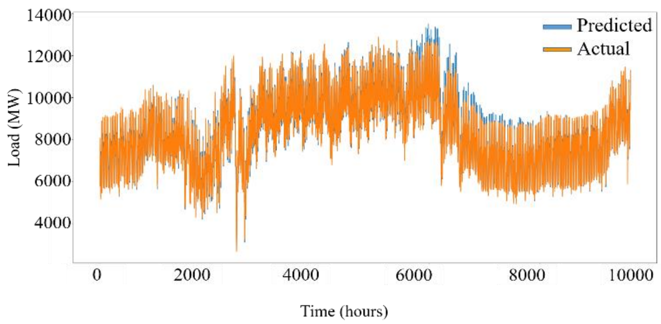


Fig. 8. Prediction curve using test data

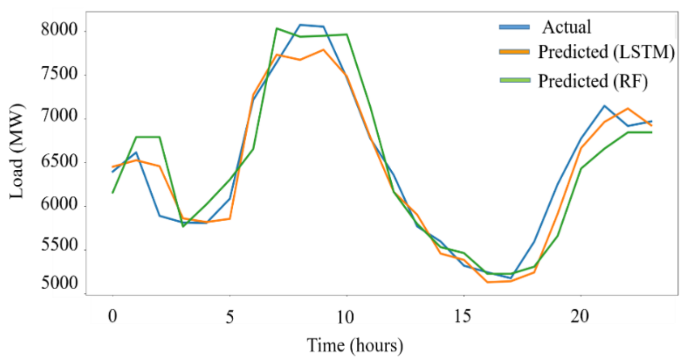


Fig. 9. Load forecasting diagram using proposed model

In fig. 9. it is seen that proposed model forecasts the actual electrical load pattern properly. In fact, the observation from the above figure indicates that the prediction trend of the LSTM network is close enough to the actual load pattern. For example, on Feb 4, 2020, at 11 am, our predicted result shown the forecasted demand 7100 MW (almost) which is nearer to the actual demand (7157 MW). Again at 12 pm, on that day, forecasted demand found 6600 MW (almost) while actual demand was 6687 MW. On the contrary, from the above figure, the prediction curve done by RF regression model shows frequent deviation at different points from the actual load pattern. The error value obtained from the RF model is 11.03 which is very high than the proposed model.

Now, the load forecasting outcomes of proposed LSTM model for the different months of 2020 are shown below from the fig. 10 to fig. 14 from which a clear indication is, the forecasted load demand patterns are more adjacent to the actual trends rather than the comparable RF model. In fig. 10, first 4 days, 20th, 21st, 22nd and the last 2 days of January gave a perfect forecasting trends that are almost similar to actual patterns and considering winter, load fluctuation is seen comparatively low.

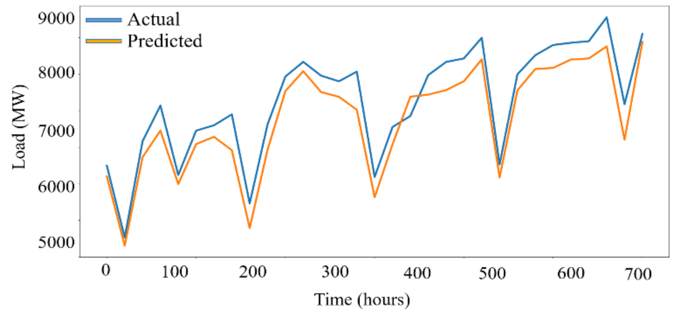


Fig. 10. Load Forecasting using LSTM for January, 2020

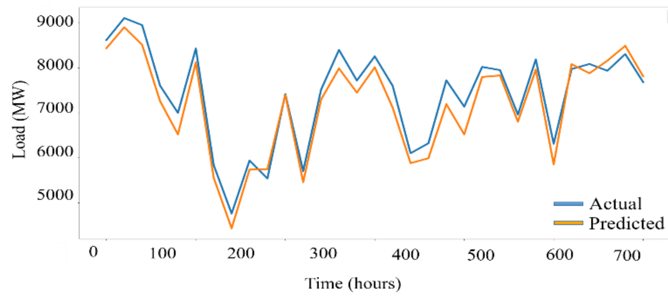


Fig. 11. Load Forecasting using LSTM for March, 2020

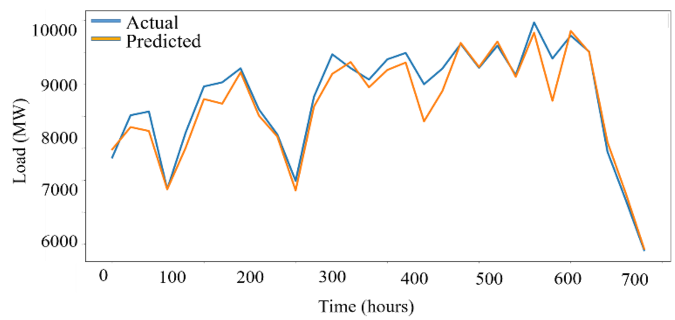


Fig. 12. Load Forecasting using LSTM for June, 2020

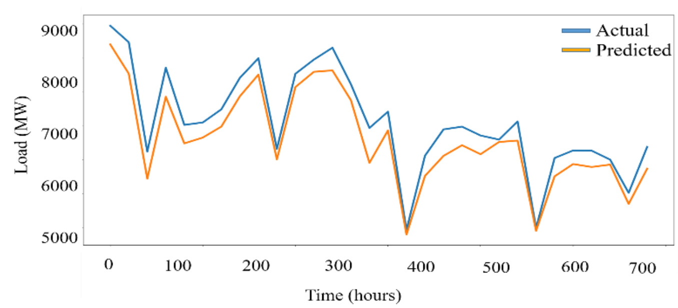


Fig. 13. Load Forecasting using LSTM for September, 2020

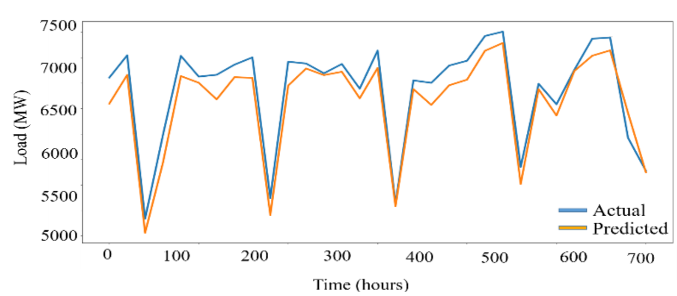


Fig. 14. Load Forecasting using LSTM for December, 2020

In fig. 11, the load demand of 7th and 8th days of March falls below 5000 MW and the forecasted curve can accurately measure the demand on those days nearly to 3900 MW which is less than the actual. From fig. 12, the load demand of last 5 days of June is continuously decreasing starting from 9500 MW and here, the predicted load pattern is also too close to the actual. For September in fig. 13, the major load fluctuations occur on 6th-7th, 13th and 19th of the month and the proposed scheme shows adjoining load patterns on those days. Again, on 2nd, 8th-10th, 14th-16th and 20th-24th December, the forecasting load curves are perfectly fitted to the actual patterns seen in fig. 14. Noteworthy, summer peak demand fluctuates roughly 20% of the typical demand while in winter, peak level reaches nearly 50% of the average demand.

TABLE III. Comparison between proposed LSTM and RF

| Month/ Year | Random Forest | | Proposed LSTM Method | |
|---|---|---|---|---|
| | MAPE % | RMSE | MAPE % | RMSE |
| Jan, 2020 | 8.71 | 576.63 | 2.26 | 217.32 |
| March,2020 | 18.14 | 1157.82 | 1.98 | 196.92 |
| June, 2020 | 7.38 | 488.89 | 2.01 | 220.69 |
| Sept, 2020 | 17.55 | 973.21 | 2.09 | 207.56 |
| Dec, 2020 | 5.22 | 344.38 | 2.05 | 190.01 |
| Overall | 11.03 | 727.68 | 2.29 | 235.62 |

From the TABLE III, comparing the months' predictions, LSTM scheme shows higher performance than the RF model. For instance, in March 2020, the MAPE value in the LSTM method is observed 1.98% whereas in the RF model, it is more than 9 times larger (18.14%). For September 2020, RMSE value in LSTM is 207.56 which is comparatively smaller than RF (4.7 times smaller). Again, the overall error rate in the LSTM model is very low compared to the RF (MAPE 2.29%) and the accuracy is 97.71%. So, in terms of error metric and accuracy, it represents that the forecasting results of the proposed LSTM scheme are satisfactory and more efficacious for STLF purposes.

## IV. CONCLUSION

This research paper comes up with a novel approach of implementing the customized encoder-decoder based LSTM framework in STLF. Encoder-decoder based LSTM prediction scheme has already been used for sequence prediction in the natural language processing but still unfamiliar in the arena of electrical load forecasting. Here, the sequence learning capability of the proposed model strengthens its viability to be used in STLF where the load nature is volatile. To get more insights about load fluctuation of BPS occurring from time to time, the prediction outcomes may have great consequences. Again, a large load data of BPS till 2021 and the forecasting outcomes will give a better understanding on recent load trends that can lead to a well-planned and optimized future grid. However, the proposed LSTM model shows the lower MAPE values compared to the RF model. This LSTM network also performs well on the different months' prediction (accuracy near about 98% in each) and its overall error metric rate is also notably low (MAPE 2.29%). Hence, considering all the validation cases, it can undoubtedly be said that the above mentioned LSTM based framework outperforms in every prediction result and thus, the high accuracy of the proposed scheme indicates its robustness to renovate BPS network by predicting future load in different time horizons.

## REFERENCES


[1] P. T. Mana, and J. Hansen, "Distributed economic dispatch using framework for network co- simulation" in Proc. 2018 *IEEE Power & Energy Society General Meeting*, Portland, OR, USA, 2018, pp.15.

[2] P. Vähäkyla, E. Hakonen, and P. Léman, "Short-term forecasting of grid load using Box-Jenkins techniques," *Int. J. Elect. Power Energy Syst*., vol. 2, no. 1, pp. 2934

[3] G. Juberias, R. Yunta, J. G. Moreno, and C. Mendivil, "A new ARIMA model for hourly load forecasting," in *Proc. IEEE Transmiss. Distrib Conf*., vol. 1, Apr. 1999, pp. 314319.

[4] H. Li, Y. Zhao, Z. Zhang, and X. Hu, "Short-term load forecasting based on the grid method and the time series fuzzy load forecasting method," in *Proc. Int. Conf. Renew. Power Gener.*, Oct. 2015, pp. 16.

[5] K.-B. Song, Y.-S. Baek, D. H. Hong, and G. Jang, "Short-term load forecasting for the holidays using fuzzy linear regression method," *IEEE Trans. Power Syst*., vol. 20, no. 1, pp. 96101, Feb. 2005.

[6] W. Christiaanse, "Short-term load forecasting using general exponential smoothing," *IEEE Trans. Power App. Syst*., vol. PAS-90, no. 2, pp. 900911, Mar. 1971.

[7] C.-M. Lee and C.-N.Ko, "Short - term load forecasting using lifting scheme and ARIMA models," *Expert Syst. Appl.*, vol. 38, no. 5, pp. 59025911, May 2011.

[8] T. Hong, P. Pinson, Y. Wang, R. Weron, D. Yang, and H. Zareipour, "Energy forecasting: A review and outlook," *IEEE Open Access J. Power Energy*, vol. 7, pp. 376388, 2020.

[9] H. H. Çevik and M. Çunkas, " Short - term load forecasting using fuzzy logic and ANFIS," *Neural Comput. Appl.*, vol. 26, no. 6, pp. 13551367, Aug. 2015.

[10] H. Z.Wang, J.Zhou, and K. Liu, "Summary of research on the short -term load forecasting method of the electric power system," *Elect. Automat*., vol. 37, no. 1, pp. 13, Jan. 2015.

[11] L.Medsker and L. Jain, "Recurrent neural networks : design and applications", 2001.

[12] H. A. Shaikh, M. A. Rahman, and A. Zubair, "Electric load forecasting with hourly precision using long short - term memory networks" in *Proc. 2019 International Conference on Electrical, Computer and Communication Engineering*, Cox's Bazar, Bangladesh, 2019, pp.1 4.

[13] W. Kong, Z. Y. Dong, Y. Jia, D. J. Hill, Y. Xu, and Y. Zhang, "Short -Term Residential Load Forecasting based on LSTM Recurrent Neural Network," *IEEE Trans. Smart Grid*, no. September, 2017.

[14] J. Zheng,C. Xu, Z. Zhang, and X. Li, "Electric load forecasting in smart grids using long-short-term-memory based recurrent neural network," in *Proc. of Annual Conf. on Information Sciences and Systems*, Mar. 2017, pp. 1–6.

[15] W. Kong, Z.Y.Dong, and D.J.Hill, "Short-Term Residential Load Forecasting based on Resident Behaviour Learning", in *IEEE Transactions on Power Systems* ,2017.

[16] Z. Chen, D. Zhang, H. Jiang, L. Wang, Y. Chen, Y. Xiao, J. Liu, Y. Zhang, and M. Li, "Load Forecasting Based on LSTM Neural Network and Applicable to Loads of Replacement of Coal with Electricity" in *Journal of Electrical Engineering & Technology (2021)* 16:2333–2342

[17] Ermatita, I. Pahendra and E. Darnila, "Peak Load Forecasting Based on Long Short Term Memory" in *2019 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS)*.

[18] A. Z. U. Din, Y. Ayaz, M. Hasan, J. Khan, "Bivariate short term electric power forecasting using LSTM network" in *2019 International Conference on Robotics and Automation in Industry (ICRAI).*

[19] M. A. Mahmud, "Isolated area load forecasting using linear regression analysis: practical approach," *Energy and Power Engineering*, Vol.3. No. 4, pp. 547-550, September 2011.

[20] A. Islam, S. R. Hasib and M. S. Islam, "Short term electricity demand forecasting for an isolated area using two different approaches," *Journal of Power Technologies*, vol. 93, no. 4, pp. 185-193, 2013.

[21] H.-Al-Shaikh, Md. Asifur Rahman, and Ahmed Zubair, "Electric Load Forecasting with Hourly Precision Using Long Short-Term Memory Networks" in *International Conference on Electrical, Computer and Communication Engineering (ECCE)*, 2019.

[22] "Report on Forecasted Generation & Demand of national grid", Power Grid Company of Bangladesh Ltd., Available on : http://pgcb.gov.bd/site/page/0dd38e19-7c70-4582-95ba-078fccb609a8/-