# Short-Term Load Forecasting Using an LSTM Neural Network

Mohammad Safayet Hossain and Hisham Mahmood
*Department of Electrical and Computer Engineering*
*Florida Polytechnic University*
Lakeland, FL, USA
mhossain7586@floridapoly.edu, hmahmood@floridapoly.edu

***Abstract*—In this paper, two forecasting models using long short term memory neural network (LSTM NN) are developed to predict short-term electrical load. The first model predicts a single step ahead load, while the other predicts multi-step intraday rolling horizons. The time series of the load is utilized in addition to weather data of the considered geographic area. A rolling time-index series including a time of the day index, a holiday flag and a day of the week index, is also embedded as a categorical feature vector, which is shown to increase the forecasting accuracy significantly. Moreover, to evaluate the performance of the LSTM NN, the performance of other machines, namely a generalized regression neural network (GRNN) and an extreme learning machine (ELM) is also shown. Hourly load data from the electrical reliability council of Texas (ERCOT) is used as benchmark data to evaluate the proposed algorithms.**

***Index Terms*—aggregated load, rolling prediction horizon, LSTM, neural network, artificial intelligence, deep learning.**

## I. Introduction

The short term load forecasting is an essential part of any reliable and economic energy management system. The bulk load is a combination of different local load profiles. The local weather of different parts of a large geographical area affects these individual load profiles. Recently, local weather variables and local electrical load profiles are considered in load forecasting models. A load forecasting method that partitions the aggregated load based on identified local weather areas is proposed in [1]. An artificial neural network (ANN) and auto-regressive integrated moving average (ARIMA) are used to forecast individual local load profiles before encoding them into the aggregated load. An ensemble load forecasting model is proposed in [2] where similar sub-profiles of an aggregated load demand are grouped together. The load from each individual group is predicted separately before merging it into the total aggregated load. A forecasting algorithm is proposed in [3] where ambient temperature and relative humidity are embedded as a heat index, and used as one of the input features. The approaches proposed in [1]–[3] use artificial neural network (ANN) as the forecasting engine. Recently the deep neural networks are accepted as a more effective option, for time series prediction, in comparison to the shallow ANNs [4], [5]. The utilization of Echo State Networks (ESN) for short term forecasting using recursive multi-step strategy is discussed in [6]. In a recursive multi-step forecasting method, the first value of the prediction horizon is predicted from the measured input sequence, then the engine uses the predicted values recursively to predict the rest of the horizon steps. On the other hand, in direct multi-step forecasting, a sequence of multiple time steps is used to predict the multi-step horizon as an output sequence. The direct multi-step forecasting strategy is more suitable, than the recursive one, for short prediction horizon [7]. An LSTM NN based forecasting method is proposed in [8] to predict a day ahead aggregated load by adopting a characteristic load decomposition technique. A combination of a deep NN and a shallow NN is used to forecast a day-ahead load profile in [9]. In day-ahead forecasting strategies, the prediction horizon starting point is commonly fixed at a certain time in the day, whereas in rolling horizon strategies, the prediction horizon is rolling forward in time as a sliding window by updating the starting point of the input-output sequence periodically. A wavelet neural network (WNN) based forecasting model considers weather forecast variables along with historical load data as input features [10]. Using temperature forecast data as a predictor is proposed in [11]. Since, the electrical load is partially correlated to weather, a reasonably accurate weather forecast can improve the prediction accuracy significantly. Weather forecast based models requires a reliable internet connection, which is in general very reliable in bulk power systems. However, in the case of remote area and developing countries microgrid based systems, reliability may be an issue.

In this paper, two forecasting models are developed, using LSTM NN, to predict short-term aggregated load. First, a single step ahead forecasting method is explored. Thereafter, a multi-step ahead algorithm is developed using a direct multi-step forecasting strategy. A rolling time-index series including a time of the day index, a holiday flag and a day of the week index, is embedded as a categorical feature vector in the model, which is shown to increase the forecasting accuracy significantly. It is shown that using shorter intraday rolling horizons can improve the prediction accuracy further in comparison to the common 24-hour horizon prediction. This may give high merits for employing dual prediction horizons, e.g. 3-step and 24-step horizons, in predictive energy management applications, especially in small microgrids where the load profile is more volatile. Finally, the algorithms are implemented in GRNN and ELM to evaluate the performance of the LSTM NN based models.

Section II presents the data set preparation. In Section III, the proposed forecasting algorithm is discussed along with the LSTM NN framework. Section IV illustrates performance evaluation metrics. Simulation results are shown and discussed in Section V, followed by the concluding remarks in Section VI.

## II. Preparation of Data Set

The aggregated load of the west region of ERCOT is used as the benchmark in this paper [12]. The hourly weather data from different local data centers of the same region is collected from the national renewable energy laboratory (NREL) website for the period of 2012-2015 [13]. The weather data set is created by averaging the extracted data. The correlation analysis between the electrical load and the meteorological variables temperature, relative humidity and pressure is performed, for the years of 2012 to 2014 using Pearson product-moment correlation coefficient formula. The correlation coefficient $\eta$ of two vector $x, y$ having length $m$ can be calculated as below:

$$\eta = \frac{\frac{1}{m}\sum_{j=1}^{m}(x_j - \overline{x})(y_j - \overline{y})}{\sqrt{\frac{1}{m}\sum_{j=1}^{m}(x_j - \overline{x})^2}\sqrt{\frac{1}{n}\sum_{j=1}^{m}(y_j - \overline{y})^2}} \tag{1}$$

As can be noticed from Table I, the electrical load is affected significantly by ambient temperature and relative humidity according to correlation analysis. On the other hand, the ambient pressure has negligible effect on the electrical load consumption. Therefore, the ambient temperature and relative humidity are considered as the weather predictors for the forecasting models in this paper. A sample profile of electrical load, temperature and relative humidity is illustrated in Fig. 1.

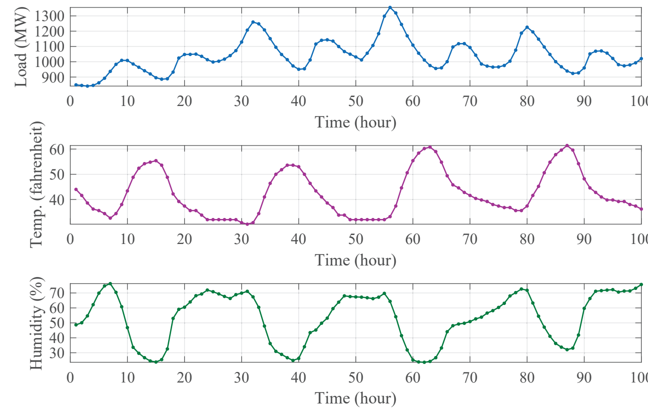


Fig. 1. Hourly sample of electrical load, temperature and relative humidity profile (1-Jan-2012 to 4-Jan-2012).

TABLE I
Correlation between Load and Weather Variables

| | *Temperature* | *Relative Humidity* | *Pressure* |
|---|---|---|---|
| *Load* | 0.4028 | -0.4111 | -0.0194 |

## III. LSTM NN Based Prediction Model

### A. LSTM Neural Network Structure

A typical regression type LSTM NN comprises multiple layers including a sequence input layer, an LSTM layer, and a regression output layer. The basic unit of an LSTM layer is called a *memory block*, which contains a single or multiple *memory cells* [14]. The internal architecture of a *memory block* is depicted in Fig. 2. The information flow through each *memory cell* is regulated by a so-called *forget gate* along with an *input gate* and an *output gate*. For a *memory block*, with $J$ number of *memory cells*, and an input vector $x(\tau) \in R^K$, the current state vector is $c(\tau) \in R^J$. The output vector $h(\tau) \in R^J$ is determined by utilizing the state vector from the previous step $c(\tau-1) \in R^J$ along with the output vector $h(\tau-1) \in R^J$. The input activation vector, forget activation vector and output activation vector are denoted by $i(\tau) \in R^J$, $f(\tau) \in R^J$ and $o(\tau) \in R^J$, respectively. Also, an intermediate state vector, or so called a candidate state vector, is introduced as $\widehat{c}(\tau) \in R^J$. The corresponding bias of $i(\tau)$, $f(\tau)$, $o(\tau)$ and $\widehat{c}(\tau)$ are denoted by $b_i$, $b_f$, $b_o$ and $b_c$ respectively. $W_f, W_i, W_c$ and $W_o$ are corresponding input weight matrices having $(J \times K)$ dimension. Whereas, $U_f, U_i, U_c$ and $U_o$ are corresponding recurrent connections weight matrices having $(J \times J)$ dimension. Each *memory block* is updated at time instant $\tau$ as follows:

$$f(\tau) = \sigma(W_f x(\tau) + U_f h(\tau-1) + b_f) \tag{2}$$

$$i(\tau) = \sigma(W_i x(\tau) + U_i h(\tau-1) + b_i) \tag{3}$$

$$\widehat{c}(\tau) = \phi(W_c x(\tau) + U_c h(\tau-1) + b_c) \tag{4}$$

$$o(\tau) = \sigma(W_o x(\tau) + U_o h(\tau-1) + b_o) \tag{5}$$

$$c(\tau) = f(\tau) \otimes c(\tau-1) + i(\tau) \otimes \widehat{c}(\tau) \tag{6}$$

$$h(\tau) = o(\tau) \otimes \phi(c(\tau)) \tag{7}$$

The sigmoid function ($\sigma$) and the hyperbolic tangent function ($\phi$) are used as the activation functions. A special sign "$\otimes$" is used to indicate the element-wise multiplication. The element-wise functions $\sigma$ and $\phi$ are defined as follows:

$$\sigma(z) = \frac{1}{1+e^{-z}} \tag{8}$$

$$\phi(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}} \tag{9}$$

### B. LSTM NN Based Single-step Forecasting Algorithm

The input array of the LSTM NN comprises a number of matrix cells. The input matrix cell at a time step $(\tau - 1)$ is prepared as follows:

1) The number of look back time points, i.e. the input sequence length $M$ is selected.
2) The input feature sequences are organized as follows: the hourly electrical load of past $M$ time steps is

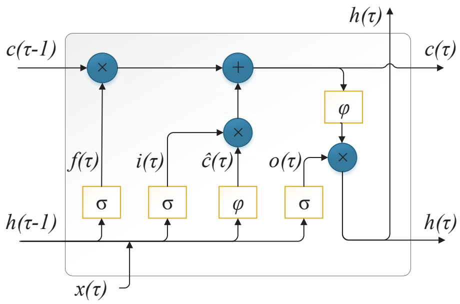


Fig. 2. Architecture of LSTM NN memory block.

$L = \{l(\tau - M), l(\tau - M + 1), ..., l(\tau - 1)\} \in R^M$; the hourly temperature of past $M$ time steps is set as $F = \{f(\tau - M), f(\tau - M + 1), ..., f(\tau - 1)\} \in R^M$; the hourly relative humidity of past $M$ time steps is set as $R = \{r(\tau - M), r(\tau - M + 1), ..., r(\tau - 1)\} \in R^M$.

3) The features are normalization using the min-max method as below:

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{10}$$

4) Finally, the normalized input features are organized in a $(3 \times M)$ input matrix array as $X(\tau - 1) = \{\widetilde{L}^T, \widetilde{F}^T, \widetilde{R}^T\}^T$.

Thereafter, a deep learning network is attained by stacking two LSTM layers as illustrated Fig. 3 (a). The first LSTM layer accepts the input matrix array sequentially and updates the *memory block* $M$ times for each full input sequence sample. The second LSTM layer *memory block* is updated with first layer synchronously. The last updated output in the second layer, which correspond to the last time step in the sequence, is sent to the output layer to produce a scalar output. This output is the predicted next step load, $Y(\tau - 1) = \{l(\tau)\} \in R$.

### C. LSTM NN Based Multi-step Forecasting Algorithm

The conventional day-ahead forecasting models commonly use previous day data as input where the model takes advantage of the general pattern resemblance between the input sequence and prediction horizon. However, in the case of intraday rolling horizons, the model learns from the preceding intraday sequence which carries less resemblance to the prediction horizon than that in the case of a day-ahead prediction, as shown in Fig. 4. The multi-step ahead model is proposed using two approaches in this paper. The hourly load with ambient temperature and relative humidity are considered as the predictors in first approach, which is called Version-1. To improve the machine learning process, a holiday flag, a time of the day index, and a day of the week index are considered as categorical input features in Version-2. The model framework is illustrated in Fig. 3 (b). The input matrix cell array at a time step $(\tau - 1)$ is prepared as follows:

1) The number of look back time steps $M$ is selected so that the input sequence length matches that of the considered prediction horizon.

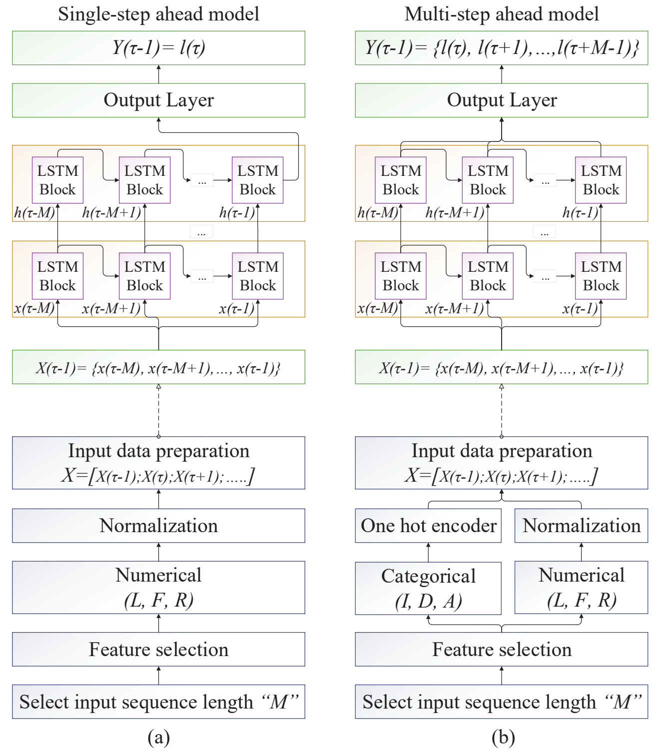


Fig. 3. Framework of load forecasting model. (a) single-step ahead, (b) multi-step ahead.

2) The numerical input features are organized as follows: the hourly electrical load of past $M$ time steps is set as $L = \{l(\tau - M), l(\tau - M + 1), ..., l(\tau - 1)\} \in R^M$; the hourly temperature of past $M$ time steps is set as $F = \{f(\tau - M), f(\tau - M + 1), ..., f(\tau - 1)\} \in R^M$; and, the hourly relative humidity of past $M$ time steps is set as $R = \{r(\tau - M), r(\tau - M + 1), ..., r(\tau - 1)\} \in R^M$.

3) The incremental time of the day indices of past $N$ time steps are set to $I \in R^M$, where $I = \{i \in N : 1 \leq i \leq 24\}$. The day of the week indices are set to $D \in R^M$ where $D = \{d \in N : 1 \leq d \leq 7\}$. The holiday flag is set to $A \in R^M$ where $A = \{1, 2\}$.

4) The numerical features are normalized using the min-max normalization method. Whereas, the categorical features are formulated using the one-hot encoder technique [15]. The one hot encoder is a technique for mapping a categorical feature, with a Q number of possible values, into a vector with Q number of elements, where only the element corresponding to the current feature value is "1", while the remaining elements are "0's".

5) Finally, the normalized input features are combined into $(36 \times M)$ input matrix array $X(\tau - 1) = \{\widetilde{L}^T, \widetilde{F}^T, \widetilde{R}^T, \widetilde{I}^T, \widetilde{D}^T, \widetilde{A}^T\}^T$.

A deep learning configuration is attained using two LSTM layers as in the single-step ahead model. The first LSTM layer accepts the input matrix sample sequentially, one feature

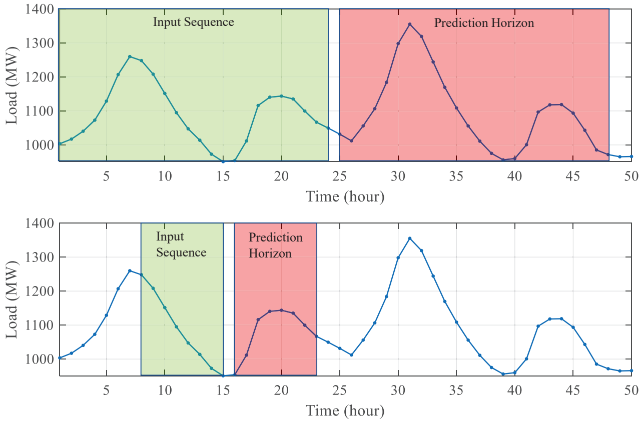


Fig. 4. Resemblance levels between input sequence and prediction horizon patterns for a day-ahead and an intraday forecasting.

vector at a time, while the *memory block* is updated $M$ times to complete a full input sequence. The *memory block* of second LSTM layer is updated synchronously by accepting the sequence of first LSTM layer cell output vector. In the second LSTM layer, the output vector of each update of the *memory block* is sent to the output layer to generate the predicted output sequence $Y(\tau-1) = \{l(\tau), l(\tau+1), ..., l(\tau+M-1)\} \in R^M$.

In the multi-step ahead model the length of input sequence and output sequence have to be same. The output layer receives the *memory block* update at each time step of the sequence from the last hidden layer. On the other hand, in the single-step ahead model the input sequence length can be varied until getting the best forecasting accuracy. In this case, the output layer receives the *memory block* update at the last time step of the sequence, from last hidden layer. Hence, it is worth emphasizing that the single step ahead model is not merely a multi-step model with a single step horizon.

## IV. Performance Evaluation Metrics

Three popular statistical metrics, the mean absolute error (MAE), the root mean square error (RMSE) and the mean absolute percentage error (MAPE), are used to evaluate the forecasting algorithms [16].

$$MAE = \frac{1}{n}\sum_{k=1}^{n}|x_{predicted}(k) - x_{actual}(k)| \quad (11)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{k=1}^{n}(x_{predicted}(k) - x_{actual}(k))^2} \quad (12)$$

$$MAPE = \frac{1}{n}\sum_{k=1}^{n}|\frac{x_{predicted}(k) - x_{actual}(k)}{x_{actual}(k)}| \times 100\% \quad (13)$$

## V. Simulation Results and Discussion

The forecasting machines are trained using the data set from the period of 2012-2014, while the 2015 data set is used to test the algorithms. The first and the second LSTM layers contain 55 neurons and 50 neurons, respectively.

TABLE II
SINGLE-STEP AHEAD FORECASTING PERFORMANCE

| Look back steps | MAE (MW) | RMSE (MW) | MAPE (%) |
|---|---|---|---|
| **2** | 19.47 | 25.71 | 1.71 |
| **4** | 18.59 | 24.68 | 1.64 |
| **6** | 18.53 | 24.32 | 1.65 |
| **8** | 17.61 | 23.62 | 1.55 |
| **10** | 17.51 | 22.91 | 1.54 |
| **12** | 17.11 | 22.65 | 1.52 |
| **14** | 16.64 | 22.16 | 1.48 |
| **16** | 16.82 | 22.42 | 1.49 |
| **18** | 16.02 | 21.62 | 1.42 |
| **20** | 16.75 | 22.34 | 1.49 |
| **22** | 16.61 | 22.41 | 1.47 |

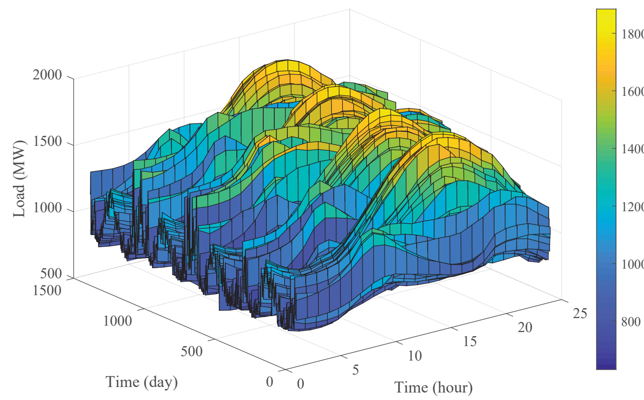


Fig. 5. Load profile of ERCOT (West) for the period of 2012-2015.

### A. Single-step ahead Algorithm

The forecasting model performance is sensitive to the input sequence length. Therefore, the performance of the algorithm with different input sequence lengths, i.e. different number of time steps in the look-back time window, is explored to determine the most effective sequence length. As illustrated in Table II, 18-step sequences achieve the lowest error. Hence, 18-step look-back windows are used to train and test the proposed algorithm. The daily load profile varies with the season as shown in Fig. 5. Hence, the model forecasting accuracy is investigated through different seasons as shown in Table III. To investigate the seasonal effect, the actual load profile of each month of each season is smoothed using a moving average technique to create a base load profile ($P_{base}$). The base load profile is subtracted from the actual load profile to quantify the volatility of the profile as $P_{fluctuation}$. Finally, a signal-to-noise ratio is calculated as the volatility measure, by considering $P_{base}$ as the signal and $P_{fluctuation}$ as the noise. The lowest forecasting error appears to be in the Summer according to Table III, which has the least volatility measure as indicated in Fig. 6. The performance of the single-step ahead model is shown in Fig. 7.

### B. Multi-step ahead Algorithm

The performance comparison of the two versions based on different intraday prediction horizons is shown in Table IV. As

TABLE III
SINGLE-STEP AHEAD LOAD FORECASTING PERFORMANCE BY SEASON

| Season | MAE (MW) | RMSE (MW) | MAPE (%) |
|---|---|---|---|
| Spring (Mar-May) | 15.19 | 20.27 | 1.51 |
| Summer (Jun-Aug) | 14.51 | 18.14 | 1.14 |
| Autumn (Sep-Nov) | 15.68 | 20.53 | 1.36 |
| Winter (Dec-Feb) | 18.77 | 26.31 | 1.59 |

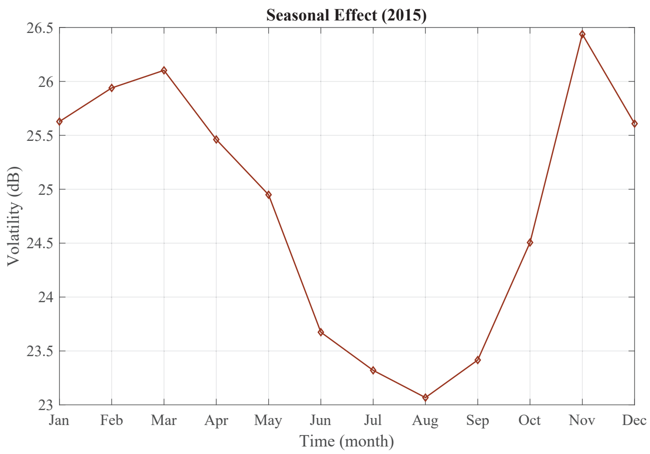


Fig. 6. Profile volatility measure of monthly electrical load profiles in 2015.

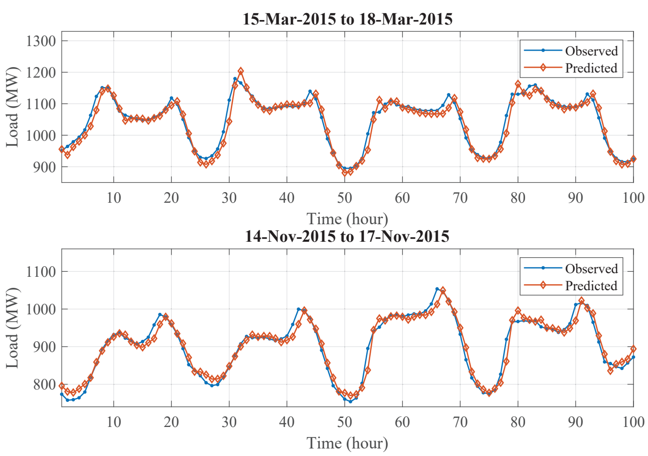


Fig. 7. Single-step ahead load forecasting performance.

expected, Version-1 reaches its highest accuracy when 24-hour prediction horizons are used. The forecasting performance drops at shorter prediction horizons. The forecasting accuracy of Version-2 is better than Version-1 for all prediction horizons, which verifies the contribution of the added categorical features. The prediction accuracy drops with increasing the prediction horizon. The seasonal effect on the forecasting accuracy of Version-2 model is shown in Table V, for 24-hour horizons. The model achieves its highest accuracy in the Summer, as in the case of the single-step ahead model. Two examples of 12-hour and 24-hour ahead forecasts using Version-2 model are shown in Fig. 8 and Fig. 9, respectively.

TABLE IV
COMPARISON BETWEEN TWO VERSIONS OF THE MULTI-STEP AHEAD MODEL

| Hours | Version-1 | | | Version-2 | | |
|---|---|---|---|---|---|---|
| | MAE | RMSE | MAPE | MAE | RMSE | MAPE |
| 3 | 78.89 | 86.98 | 6.69 | 29.52 | 31.89 | 2.61 |
| 6 | 82.37 | 97.71 | 7.19 | 42.22 | 47.31 | 3.69 |
| 12 | 82.25 | 100.55 | 7.11 | 55.42 | 63.81 | 4.79 |
| 24 | 65.42 | 76.63 | 5.71 | 64.58 | 75.02 | 5.59 |

TABLE V
SEASONAL PERFORMANCE USING 24-HOUR HORIZONS WITH VERSION-2 MODEL

| Season | MAE (MW) | RMSE (MW) | MAPE (%) |
|---|---|---|---|
| Spring (Mar-May) | 63.25 | 74.92 | 6.15 |
| Summer (Jun-Aug) | 49.34 | 57.22 | 3.79 |
| Autumn (Sep-Nov) | 49.76 | 64.06 | 4.61 |
| Winter (Dec-Feb) | 103.39 | 118.28 | 8.67 |

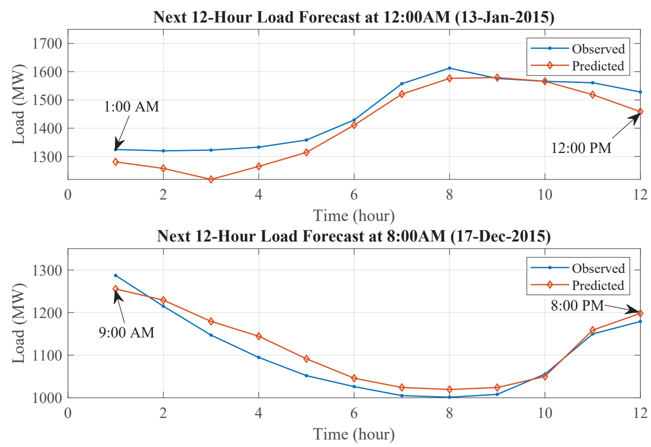


Fig. 8. 12-hour ahead load forecast in Winter 2015.

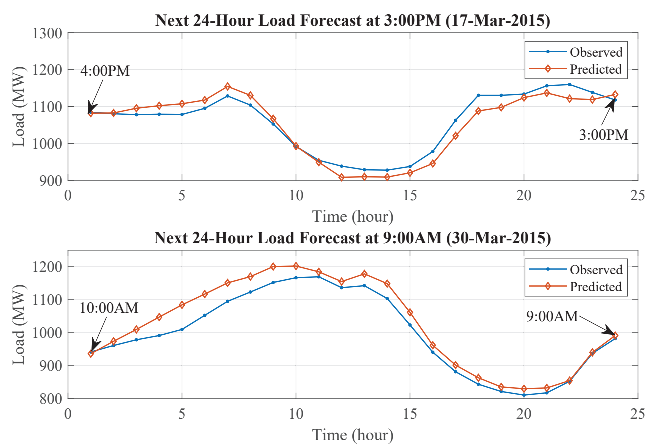


Fig. 9. 24-hour ahead load forecast in Spring 2015.

### C. Forecasting Engine Performance Evaluation

The LSTM NN based forecasting models are benchmarked by implementing the same algorithms in GRNN and ELM.

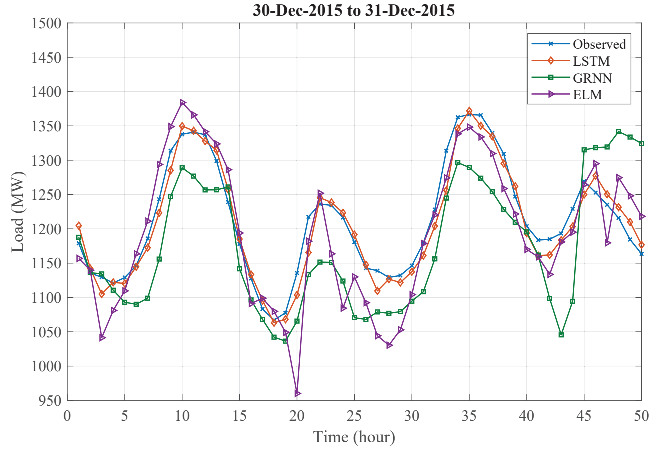


Fig. 10. Single-step ahead load forecasting comparison.

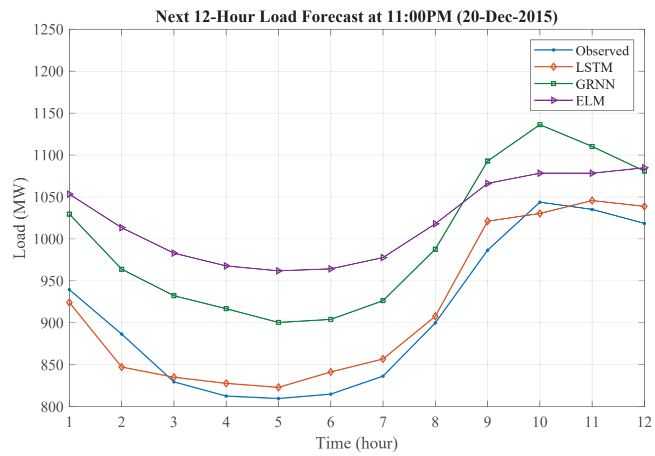


Fig. 11. Multi-step ahead load forecasting comparison.

TABLE VI
FORECASTING PERFORMANCE COMPARISON

| **Engine** | **Single-step ahead Model** | | | **Multi-step ahead Model** | | |
|---|---|---|---|---|---|---|
| | **MAE** | **RMSE** | **MAPE** | **MAE** | **RMSE** | **MAPE** |
| **LSTM NN** | 17.11 | 22.65 | 1.52 | 55.42 | 63.81 | 4.79 |
| **GRNN** | 34.35 | 45.98 | 3.05 | 61.62 | 68.45 | 5.33 |
| **ELM** | 36.59 | 44.52 | 3.44 | 73.82 | 80.56 | 6.86 |

The single step ahead model is implemented with a 12-step look-back window. The forecasting comparison for two days in Winter is presented in Fig. 10. The multi-step model is implemented for 12-hour prediction horizons. Fig. 11 shows the performance of the engines for a 12-hour prediction horizon. Finally, the performance comparison is shown in Table VI, which shows the superiority of the LSTM network.

## VI. CONCLUSION

Two intelligent forecasting algorithms using LSTM NNs are proposed and evaluated in this paper. The effect of the input sequence length on the accuracy of the single step model is investigated. It is shown that the highest accuracy can be achieved at a certain sequence length, where increasing the size of the look back window further reduces the accuracy. Also, the seasonal effect on the load profiles is investigated using the introduced volatility measure. For multi-step prediction, it is shown that the addition of the categorical time-related indices can further improve the prediction accuracy of intraday rolling horizons in comparison to the common 24-hour horizon prediction. Finally, the superiority of LSTM NNs, in comparison to GRNNs and ELMs, is illustrated.

## REFERENCES

[1] D. Liu, L. Zeng, C. Li, K. Ma, Y. Chen and Y. Cao, "A Distributed Short-Term Load Forecasting Method Based on Local Weather Information," *IEEE Systems Journal*, vol. 12, no. 1, pp. 208-215, March 2018.

[2] Y. Wang, Q. Chen, M. Sun, C. Kang and Q. Xia, "An Ensemble Forecasting Method for the Aggregated Load With Subprofiles," *IEEE Transactions on Smart Grid*, vol. 9, no. 4, pp. 3906-3908, July 2018.

[3] J. Xie, Y. Chen, T. Hong and T. D. Laing, "Relative Humidity for Load Forecasting Models," *IEEE Transactions on Smart Grid*, vol. 9, no. 1, pp. 191-198, Jan. 2018.

[4] Y. Bengio, A. Courville and P. Vincent, "Representation Learning: A Review and New Perspectives," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, pp. 1798-1828, Aug. 2013.

[5] Z. Deng, B. Wang, Y. Xu, T. Xu, C. Liu and Z. Zhu, "Multi-Scale Convolutional Neural Network With Time-Cognition for Multi-Step Short-Term Load Forecasting," *IEEE Access*, vol. 7, pp. 88058-88071, 2019.

[6] B. A. Hverstad, A. Tidemann, H. Langseth and P. ztrk, "Short-Term Load Forecasting With Seasonal Decomposition Using Evolution for Parameter Tuning," *IEEE Transactions on Smart Grid*, vol. 6, no. 4, pp. 1904-1913, July 2015.

[7] A. F. Atiya, S. M. El-Shoura, S. I. Shaheen and M. S. El-Sherif, "A comparison between neural-network forecasting techniques-case study: river flow forecasting," in *IEEE Transactions on Neural Networks*, vol. 10, no. 2, pp. 402-409, March 1999.

[8] K. Park, S. Yoon and E. Hwang, "Hybrid Load Forecasting for Mixed-Use Complex Based on the Characteristic Load Decomposition by Pilot Signals," *IEEE Access*, vol. 7, pp. 12297-12306, 2019.

[9] K. Chen, K. Chen, Q. Wang, Z. He, J. Hu and J. He, "Short-Term Load Forecasting With Deep Residual Networks," *IEEE Transactions on Smart Grid*, vol. 10, no. 4, pp. 3943-3952, July 2019.

[10] X. Sun et al., "An Efficient Approach to Short-Term Load Forecasting at the Distribution Level," *IEEE Transactions on Power Systems*," vol. 31, no. 4, pp. 2526-2537, July 2016.

[11] K. Methaprayoon, W. Lee, S. Rasmiddatta, J. R. Liao and R. J. Ross, "Multistage Artificial Neural Network Short-Term Load Forecasting Engine With Front-End Weather Forecast," *IEEE Transactions on Industry Applications*, vol. 43, no. 6, pp. 1410-1416, Nov.-dec. 2007.

[12] Electric Reliability Council of Texas (ERCOT). [Online]. Available: *http://www.ercot.com/gridinfo/load*.

[13] National Renewable Energy Laboratory (NREL). [Online]. Available: *https://www.nrel.gov/research/data-tools.html*.

[14] F. A. Gers, J. Schmidhuber and F. Cummins, "Learning to forget: continual prediction with LSTM," *1999 Ninth International Conference on Artificial Neural Networks ICANN 99*. (Conf. Publ. No. 470), Edinburgh, UK, 1999, pp. 850-855 vol.2.

[15] 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 Transactions on Smart Grid*, vol. 10, no. 1, pp. 841-851, Jan. 2019.

[16] R. Jiao, T. Zhang, Y. Jiang and H. He, "Short-term non-residential load forecasting based on multiple sequences LSTM recurrent neural network," *IEEE Access*, vol. 6, pp. 59438-59448, 2018.