**Research Article**

# Electricity energy consumption forecasting using LSTM and NAR

**Parastou BEHGOUY[1],* , Abbas UĞURENVER[1]**

*[1]Institute of Graduate Studies, Department of Electrical and Electronic Engineering, Istanbul Aydın University, Istanbul, 34295, Türkiye*

**ABSTRACT**

Energy management systems (EMS) have gained importance in reducing energy waste and overconsumption. One critical component of effective EMS is long-term electricity consumption forecasting, which helps estimate energy demand and provides a general view of the future for macro-level planning. To achieve this, accurate predictions are essential for developing efficient energy plans. Consequently, machine learning models have garnered significant attention in recent decades due to their ability to address these challenges. These models, artificial neural networks with deep structures, are particularly effective, as they can analyze large volumes of data and provide accurate forecasts, even when the data exhibits complex, non-linear patterns. In this study, long-term electrical energy consumption has been forecasted using MATLAB software according to the past trend of consumption with long short-term memory (LSTM) and nonlinear autoregressive (NAR) neural networks. Monthly energy consumption data between 1973-2022 in USA was used as the data set.

The assessment metrics used for both networks included Mean Absolute Percentage Error (MAPE), Symmetric Mean Absolute Percentage Error (SMAPE), Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Bias Error (MBE), and the Coefficient of Determination ($R^2$). These metrics were calculated to evaluate the performance and accuracy of the networks. The fewer error values and the higher value of Coefficient of determination ($R^2$ =0.9318) in LSTM show more accuracy in this network when it's compared with NAR ($R^2$ =0.6387). The same situation with the Root Mean Square Error witch it shows that the LSTM gives better results (RMSE =8.8033) than the NAR (RMSE =20.1971).

The results show LSTM network has better performance and can produce more accurate output because of its structure which has short-term and long-term memories that make it able to remember information for long periods and has feedback connections. In fact, LSTM is a high-performance nonlinear predictor.

**Cite this article as:** Behgouy P, Uğurenver A. Electricity energy consumption forecasting using LSTM and NAR. Sigma J Eng Nat Sci 2026;44(1):153–161.

*Corresponding author.
*E-mail address: parastoubehgouy@stu.aydin.edu.tr

## INTRODUCTION

Global electricity consumption has been raised rapidly because of population growth and technology development [1], the energy management system (EMS) has gained importance to reduce waste of energy and energy overconsumption. Energy consumption forecasting has an important role in EMS [2,3] and accurate prediction is essential to achieve energy efficiency.

So far different type of techniques have been proposed as predictive model to forecast the future behavior of the system, Statistical approach such as SARIMA, ARMA, DLM, Holt–Winters, artificial neural networks including multilayer perceptron [4], physical models and also hybrid techniques. Machine learning (ML) as data-driven models has been realized the most accurate technique to obtain acceptable output in forecasting [5-7]. expresses artificial neural networks have perfect performance if there is sufficient data to train the models.

Recently, researchers have been attracted to deep learning networks because of their ability to receive and analyze data when they follow complex non-linear patterns and have a large volume [8,9]. Deep learning has provided significant progress in energy forecasting. These networks have created more innovative research opportunities that increase prediction accuracy [10].

Convolutional neural network with deep structure is presented in [11], which has been able to improve the electricity consumption estimation compared to the seasonal ARIMAX method. The results show that the accuracy of the deep neural network is higher than the ARIMAX. Another study shows the Solar radiation forecasting by using deep neural networks (ARIMAX method) in Eskisehir, Türkiye [12]. A hybrid LSTM model integrated with a stationary wavelet transform (SWT) has demonstrated efficacy in rectifying abnormalities in home energy use, hence enhancing forecast precision in both the short and long term [13]. Another study comparing LSTM, GRU, and Drop-GRU models revealed that LSTM excels above the other models in handling long-term dependencies and nonlinearities in energy data [14]. The integration of CNN with LSTM improves prediction by effectively capturing both geographical and temporal data, as evidenced in a study on multivariate energy use [15]. The efficacy of LSTM in time series analysis was demonstrated in a study that investigated data quality and hyperparameter optimization, highlighting the significance of preprocessing for precision [16,17] Indicates comparison of LSTM neural network and ARIMA for energy consumption forecasting. The result expresses this network is more accurate than statistical estimation methods such as ARIMA. Also, LSTM neural networks are more accurate than non-deep artificial neural networks for prediction. Overfitting may occur during the training process by adding layers in LSTM neural networks. This problem is a challenge in deep neural networks. In fact, if the number of layers increases regardless of the evaluation step error,

the prediction error will increase. The over-fitting issue could be solved by increasing data diversity and volume [18]. In conclusion, LSTM and its derivatives exhibit significant adaptability to various prediction tasks, offering solid solutions for dynamic energy systems. The results illustrate the capability of sophisticated machine learning algorithms to tackle issues in energy consumption predictions with significant accuracy and generalizability.

In this research, long-term electrical energy consumption is forecasted by LSTM and NAR neural networks using MATLAB software. Monthly energy consumption data from 1973-2022 in USA was used as the data set and the accuracy of the implemented networks was compared. There were different types of studies like the work of this paper. For example, in [19] the LSTM and NAR were used to forecast monthly smoothed sunspot number time series. In [20], The LSTM and NAR have been used to predict the wind turbine's generated power. Also, in [21] the LSTM and NAR are designed and compared to find the most accurate solution to forecast daily water consumption.

This paper consists of the following sections:

The predictive models were presented in Section 2. In Section 3, the data set and experimental results were expressed, evaluation metrics are proposed and calculated in Section 4 and finally, there is the conclusion in Section 5.

## PREDICTIVE MODELS

In this paper, two neural networks with different structures (LSTM and NAR) were implemented as multi-step predictive models which are accurate models for analyzing time series data and forecasting future values.

### LSTM Neural Network

LSTM is a type of Recurrent Neural Network (RNN) which could solve RNN's long-term memory problem. The gates of LSTM control the flow of information and data and indicate which data is unimportant and must be deleted and which ones should be kept, so this network allows to pass only the important data to achieve the accurate output [22,23]. Figure 1 shows LSTM cell structure [24].

LSTM includes 3 gates, input, output and forget gates. Every cell has two inputs, $x_t$ and $h_{t-1}$ also $h_t$ and $c_t$ as output. $c_{t-1}$ is connected to $c_t$ directly which is possible to add or delete some data, in fact, unnecessary data is deleted by forget gate. c is the short form of cell state which is long-term memory. The data is always between -1 and 1 by using the tanh. The below equations make up LSTM network, where $W_o$, $W_g$, $W_i$ and $W_f$ are the weight matrices. $b_o$, $b_g$, $b_i$ and $b_f$ are bias vectors and $\sigma$ indicates sigmoid activation functions. [24,25]

$$f_t = \sigma\left(W_f \times [x_t, h_{t-1}] + b_f\right) \qquad (1)$$

$$i_t = \sigma\left(W_i \times [x_t, h_{t-1}] + b_i\right) \qquad (2)$$

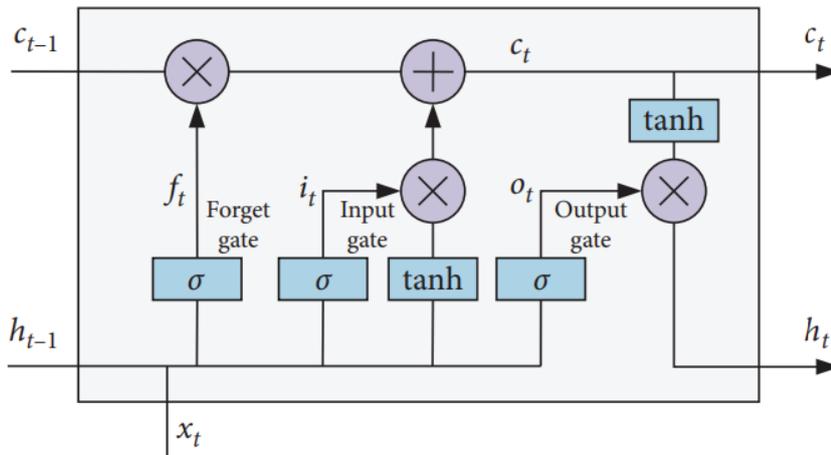$$g_t = \tanh\left(W_g \times [x_t, h_{t-1}] + b_g\right) \qquad (3)$$

**Figure 1.** LSTM structure.

$$c_t = f_t \times c_{t-1} + f_t \times g_t \qquad (4)$$

$$o_t = \sigma \left( W_o \times [x_t, h_{t-1}] + b_o \right) \qquad (5)$$

$$h_t = o_t \times \tanh(c_t) \qquad (6)$$

The training process is maintained until the minimum error or maximum epoch is achieved. According to the obtained results, the LSTM neural network is a successful deep learning approach based on a sophisticated network structure. It uses memory units in the model to capture time-series correlations.

**NAR (Nonlinear Auto-Regressive) Neural Network LSTM**

NAR neural network is an effective model to recognize time series patterns and nonlinear trends. In fact, it's a dynamic network where the future values are forecasted only from past values. This network can be implemented as a time series forecasting model [26,27]. Figure 2 shows the NAR neural network structure [28].
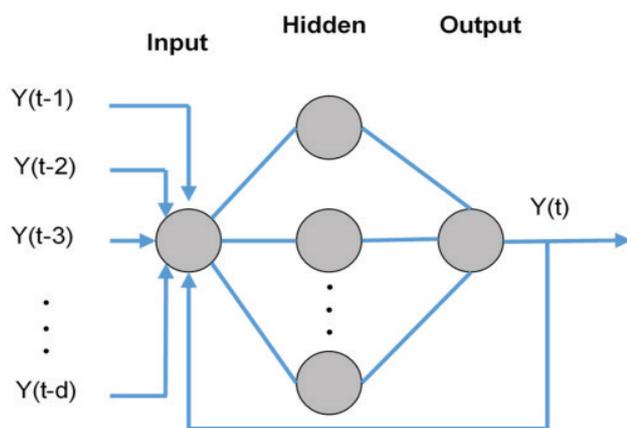
This network can be expressed by the below function [28] where y(t), f are forecasted value and nonlinear approximation function which is determined by NAR network during training and $\epsilon$ (t) is the error.

$$y(t) = f\left( y(t-1), y(t-2), y(t-3), \ldots\ldots y(t-n) \right) + \epsilon(t) \qquad (7)$$

**EXPERIMENTAL RESULTS**

Monthly electricity consumption data from 1973-2022 in USA was used as the data set and input of the models which is shown in Figure 3.

The data set is normalized by the following equation in order to achieve a similar scale by the min-max method, where $X$ and $X_n$ are actual and normalized values, $X_{MIN}$ and $X_{MAX}$ refer to minimum and maximum data's values [29].
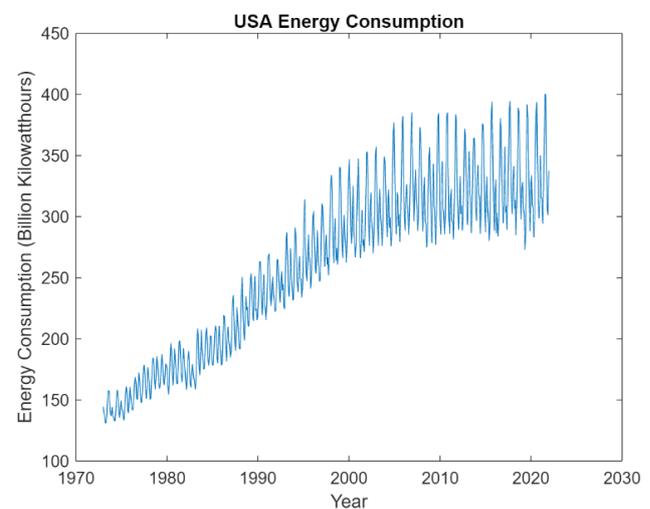


**Figure 2.** NAR Neural Network Structure [28].



**Figure 3.** Monthly electricity energy consumption.

$$X_N = \frac{X - X_{MIN}}{X_{MAX} - X_{MIN}} \qquad (8)$$

In LSTM, the time series data is divided into two parts for training (90%) and testing (10%) and in NAR into three parts, train (80%), Validation (10%) and test (10%). ANN parameters are usually set by previous experience in trial-and-error procedure, and we found that the best value for the hidden units is 300 in both networks. We trained our model depending on the dataset number that we have. At the first we divide our training to three parts. In the first part we chose a small number of hidden units which is between 100-150 and we got underfitting in our model. And then we tried to increase our hidden units to 450 units and then we got overfitting situation. So, the best situation was to choose 300 hidden units.

Figure 4 shows the proposed layers structure for models. Table 1 represents the training details for the LSTM network, where the Gradient Decay Factor is equal to 0.9, as it is the optimum value between 0.9 and 0.99. The Initial Learning Rate falls within the range of 0.001 to 0.01, the Learning Rate Schedule specifies the temporal variations of the learning rate throughout the training process, and the Learning Rate Drop Factor is usually set to a value between 0.1 and 0.5. The Learning Rate Drop Period specifies the

frequency of learning rate updates, often adjusted by the drop factor. It indicates the number of epochs or training iterations after which the learning rate should decrease. The Max Epochs parameter is set to 250, considering the specific requirements of the LSTM modeling process. The Mini Batch Size, which denotes the amount of training data utilized in each gradient descent update, is set to 128 to provide more stable gradients and faster convergence. The Verbose Frequency is set to 50, meaning the training progress will be displayed after every 50 epochs. The Validation Frequency specifies how often the model's performance is assessed on a validation dataset throughout the training process.

In Table 2, For NAR, 80 percent of the data used for training and 10 percent for testing and 10 percent for validation. The epoch number is decided by using empirical testing and chosen to be 1000 epoch to capture the intricate patterns and dependencies in the data. As can be shown from the results in NAR the Gradient is too small and the model uncertainty (MU) is too big, and this indicates that the LSTM performance is better than NAR in this research.
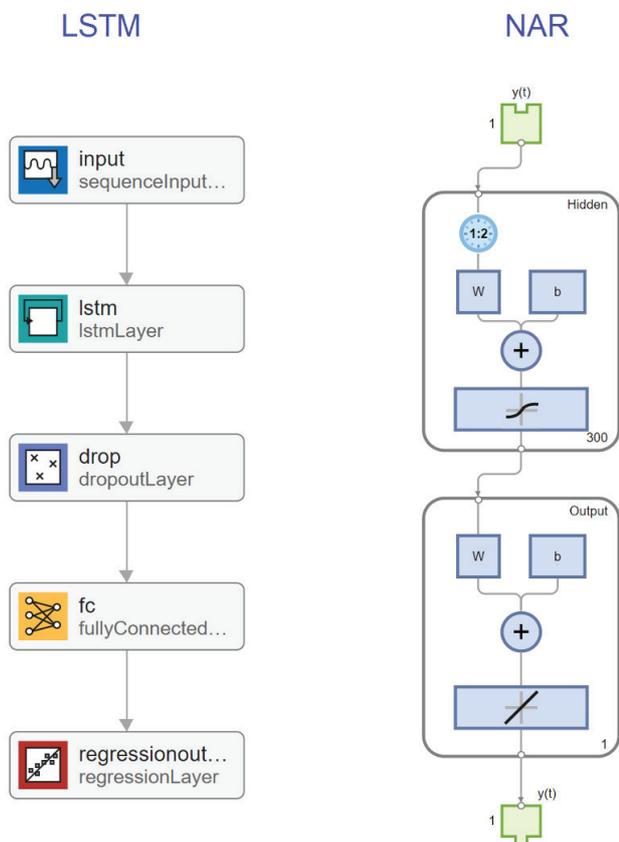
**Table 1.** LSTM Training

| Parameters | Value |
|---|---|
| Gradient Decay Factor | 0.9000 |
| Initial Learn Rate | 0.0050 |
| Learn Rate Schedule | Piecewise |
| Learn Rate Drop Factor | 0.2000 |
| Learn Rate Drop Period | 125 |
| Gradient Threshold Method | l2norm |
| Gradient Threshold | 1 |
| Max Epochs | 250 |
| Mini Batch Size | 128 |
| Verbose Frequency | 50 |
| Validation Frequency | 50 |

**Table 2.** NAR Training

| Parameters | Value |
|---|---|
| Training | %80 |
| Validation | %10 |
| Testing | %10 |
| Epoch | 1000 |
| Time | 03:14 |
| Performance | 260.7 |
| Gradient | 1e-7 |
| MU | 1e+10 |
| Validation Check | 6 |



**Figure 4.** Proposed layers structure for LSTM and NAR networks.

Figure 5 shows how the RMSE changes across iterations during the training phase in LSTM. It starts high, and at the 100th iteration, it stabilizes, indicating that the model has reached convergence. By the 250th iteration, it fits all the trained data.

Figure 6 shows the performance of NAR in training. The Mu is almost stable at 0.05 throughout the training, indicating no adjustment was needed to facilitate learning stability. Also, the gradient is declining from $4.34\times106$ to 2.59 which means that the NAR model is near to local or global minimum.

In the first experiment, energy forecasting was done with LSTM network. Figure 7 shows a comparison of actual data and forecast, also the error which is the difference between these two values. In figure 7, the blue line represents the real data, and the red line shows the estimated data. According
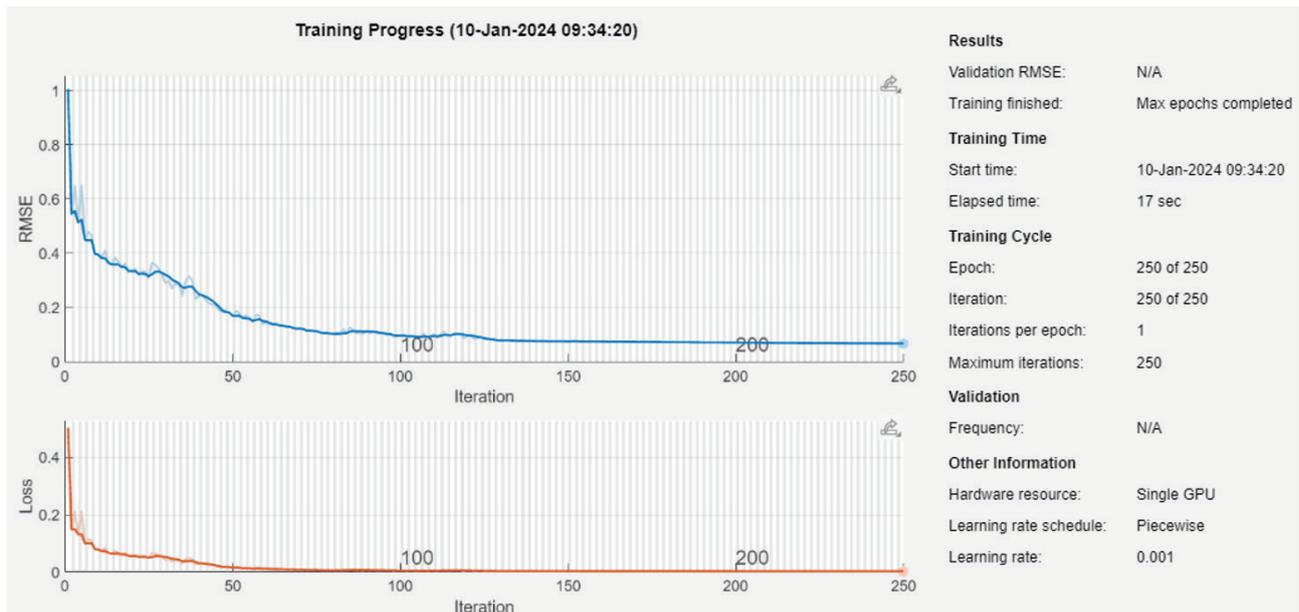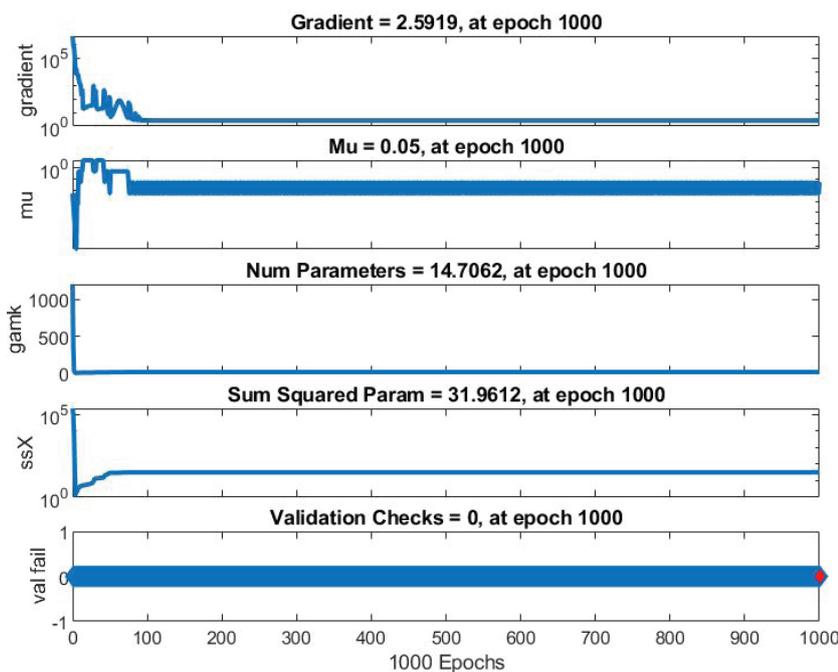


**Figure 5.** LSTM Training Progress.



**Figure 6.** NAR training progress.

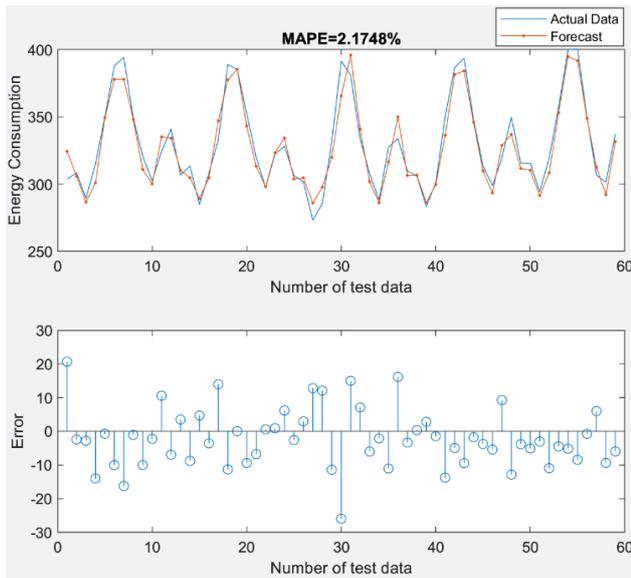**Figure 7.** Forecasting, error (LSTM).



**Figure 8.** Forecasting, error (NAR).

to the result, MAPE is about 2.1748% which shows high accuracy of the model. The second graph shows the amount of errors in each test data point.

As the second method, NAR neural network was used, and the performance of the model and error graphs are shown in Figure 8, although the MAPE is low which is around 4.7544% and the deviations are minimal, still the LSTM method shows better results than NAR.

## EVALUATION METRICS

The assessment of the presented networks was done by calculating 6 errors: MAPE, SMAPE, RMSE, MAE, MBE and R² which are expressed by the below formulas [30,31].

$$MAPE = \frac{1}{m} \sum_{i=1}^{m} \left| \frac{Y_i - X_i}{Y_i} \right| \tag{9}$$

$$SMAPE = \frac{1}{m} \sum_{i=1}^{m} \left| \frac{|X_i - Y_i|}{\frac{|X_i| + |Y_i|}{2}} \right| \tag{10}$$

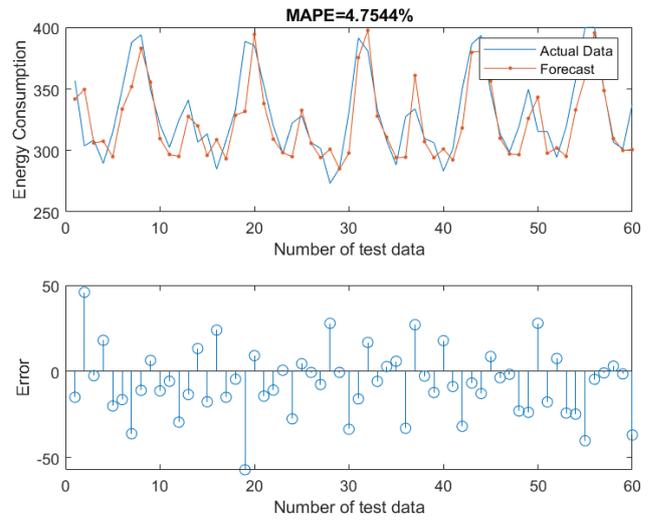$$MSE = \frac{1}{m} \sum_{i=1}^{m} (X_i - Y_i)^2 \tag{11}$$

$$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^{m} (X_i - Y_i)^2} \tag{12}$$

$$MAE = \frac{1}{m} \sum_{i=1}^{m} |X_i - Y_i| \tag{13}$$

$$MBE = \frac{1}{m} \sum_{i=1}^{m} X_i - Y_i \tag{14}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{m} (X_i - Y_i)^2}{\sum_{i=1}^{m} (\bar{Y} - Y_i)^2} \tag{15}$$

Where Xi, Yi and m are the actual value, forecasted value and the number of test data respectively. The results are expressed in Table 3. The fewer errors and the higher value of R² in LSTM show this model is more accurate.

In this study, the number of hidden units which were used in the networks for achieving optimal accuracy was 300. LSTM can produce more accurate output because of feedback connections and having short-term and long-term memories. The accuracy that we can see by using LSTM and NAR shows to us clearly that the proposed LSTM model achieves better

**Table 3.** Results

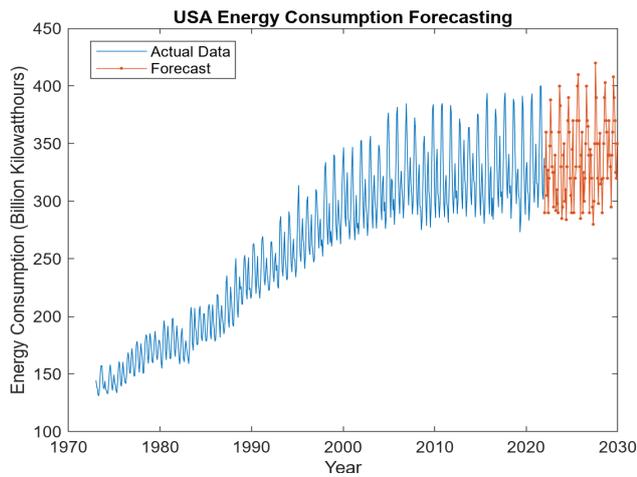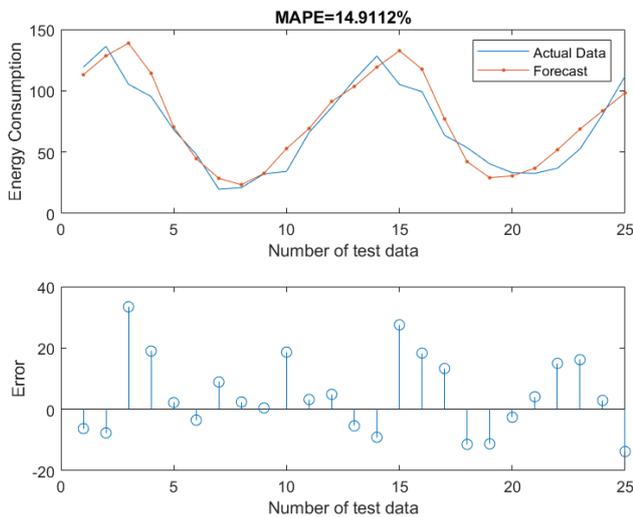| Neural Network | MAPE | SMAPE | RMSE | MAE | MBE | R2 |
|---|---|---|---|---|---|---|
| LSTM NN | 2.1748 | 2.1265 | 8.8033 | 7.0646 | -2.3907 | 0.9318 |
| NAR NN | 4.7544 | 4.8278 | 20.1971 | 15.8327 | -6.9090 | 0.6387 |

**Figure 9.** Forecasting with LSTM.



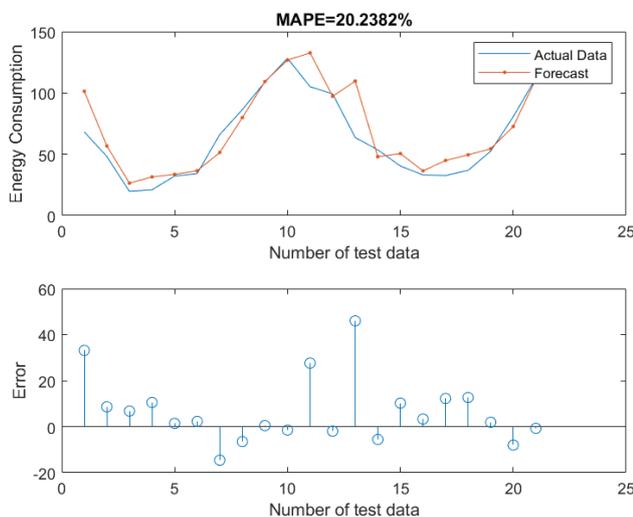**Figure 10.** Forecasting with LSTM.



**Figure 11.** Forecasting with NAR.

performance in terms of estimating a long data when compared with the non-linear auto regressive (NAR). The evaluation was based on the coefficient of determination ($R^2$) which is a number between 0 and 1 that measures how well a statistical model predicts an outcome and its shows that LSTM have a better performance (0.9318) than NAR (0.6387). Also, by looking at table 3 we can see that the MAPE, SMAPE, RMSE, MAE, MBE all these values indicate that LSTM is better than NAR. Figure 9 shows forecasting with LSTM till 2030 which can give a general view for planning.

For this study LSTM and NAR were tested also with different datasets and different periods. The data were taken from Danish energy agency For Use of Electrical Energy Consumption from 2000 to 2020 and the results indicate that LSTM have a better Coefficient of determination ($R^2$) = 0.8591 and better Mean Absolute Percentage Error MAPE=14.9 than NAR which it has a Coefficient of determination ($R^2$) = 0.7683 and MAPE=20.23 as we can see from figures 10 and 11.

## CONCLUSION

In this research, long-term electricity energy consumption was forecasted using MATLAB software with two predictive models, LSTM which is a kind of RNN and NAR (nonlinear autoregressive) neural network. The monthly energy consumption data from 1973-2022 in USA was used as a time series data set and input of the networks.The experimental results of both networks were compared and evaluation metrics in this study were different errors such as MAPE, SMAPE, RMSE, MAE, MBE and $R^2$ for measuring forecasting accuracy. The lower value of the errors and the higher value of $R^2$ in LSTM (0.9318) shows this model is more accurate than NAR (0.6387). Also, if we compared the Symmetric Mean Absolute Percentage Error in LSTM (2.1265) which gives better performance than the NAR (4.8278) and here also the lower value means the better performance. These results indicate that LSTM method is more efficient than NAR because LSTM has short-term and long-term memories which can keep information for long periods and has feedback connections. In fact, this network has shown better performance.

For the future work, we will use the same data with different NN estimation methods like, for example NARX. Also, this work can be used to estimate the stock market. Also, to increase the accuracy of the model optimization algorithms like particle swarm optimization algorithm hyperparameters of LSTM can be used.

## AUTHORSHIP CONTRIBUTIONS

Authors equally contributed to this work.

## DATA AVAILABILITY STATEMENT

The authors confirm that the data that supports the findings of this study are available within the article. Raw

data that support the finding of this study are available from the corresponding author, upon reasonable request.

## CONFLICT OF INTEREST

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## ETHICS

There are no ethical issues with the publication of this manuscript.

## STATEMENT ON THE USE OF ARTIFICIAL INTELLIGENCE

Artificial intelligence was not used in the preparation of the article.

## REFERENCES

[1] Kim TY, Cho SB. 2019. Predicting residential energy consumption using CNN-LSTM neural networks. Energy 2019;182:72–81. [CrossRef]

[2] Kim JY, Cho SB. Electric Energy Consumption Prediction by Deep Learning with State Explainable Autoencoder. Energies 2019;12:739. [CrossRef]

[3] Djenouri D, Laidi R, Djenouri Y, Balasingham I. Machine learning for smart building applications: Review and taxonomy. ACM Comput Surv 2019;52:1–36. [CrossRef]

[4] Silva FLC, Costa K, Rodrigues PC, Salas R, Gonzales JLL. Statistical and artificial neural networks models for electricity consumption forecasting in the Brazilian industrial sector. Energies, 2022;15:588. [CrossRef]

[5] Ajayi RO, Alaka H, Sulaimon I, Sunmola F, Ajayi S. Building energy consumption prediction for residential buildings using deep learning and other machine learning techniques. J Build Eng 2022;45:103406. [CrossRef]

[6] Fathi S, Srinivasan R, Fenner A, Fathi S. Machine learning applications in urban building energy performance forecasting: A systematic review. Renew Sustain Energy Rev 2020;133:110287. [CrossRef]

[7] Bourhnane S, Abid MR, Lghoul R, Dine KZ, Elkamoun N, Benhaddou D. Machine learning for energy consumption prediction and scheduling in smart buildings. SN Appl Sci 2020;2:1–10. [CrossRef]

[8] Donmez I, Aslan Z. Role of air pollution on COVID-19 in Istanbul. Sigma J Eng Nat Sci 2023;4:793–806. [CrossRef]

[9] Gozzoglu A, Ozgonenel O, Gezegin C. Modeling and implementation of demand-side energy management system. Sigma J Eng Nat Sci 2024;42:1628–1645. [CrossRef]

[10] Wang H, Lei Z, Zhang X, Zhou B, Peng J. A review of deep learning for renewable energy forecasting. Energy Convers Manage 2019;198:111799. [CrossRef]

[11] Cai M, Pipattanasomporn M, Rahman S. Day-ahead building-level load forecasts using deep learning vs. traditional time-series techniques. Appl Energy 2019;236:1078–1088. [CrossRef]

[12] Qasem M, Başaran Filik Ü. Solar radiation forecasting by using deep neural networks in Eskişehir. Sigma J Eng Nat Sci 2021;39:159–169. [CrossRef]

[13] Yan K, Li W, Ji Z, Qi M, Du Y. A Hybrid LSTM Neural Network for Energy Consumption Forecasting of Individual Households. IEEE Access 2019;7:157633–157642. [CrossRef]

[14] Mahjoub S, Alaoui LC, Marhic B, Delahoche L. Predicting Energy Consumption Using LSTM, Multi-Layer GRU and Drop-GRU Neural Networks. Sensors 2022;22:4062. [CrossRef]

[15] Chung J, Jang B. Accurate prediction of electricity consumption using a hybrid CNN-LSTM model based on multivariable data. PLoS One 2022;17:e0278071. [CrossRef]

[16] Durand D, Aguilar J, R-Moreno MD. An Analysis of the Energy Consumption Forecasting Problem in Smart Buildings Using LSTM. Sustainability 2022;14:13358. [CrossRef]

[17] Mpawenimana I, Pegatoquet A, Roy V, Rodriguez L, Belleudy C. A comparative study of LSTM and ARIMA for energy load prediction with enhanced data preprocessing. IEEE Sens Appl Symp 2020;1–6. [CrossRef]

[18] Shi H, Xu M, Li R. Deep learning for household load forecasting—A novel pooling deep RNN. IEEE Trans Smart Grid 2017;9:5271–5280. [CrossRef]

[19] Ramadevi B, Bingi. Time Series Forecasting Model for Sunspot Number. IEEE Xplore 2022;1-6. [CrossRef]

[20] Sunder M, Abishek R, Maiti M, Bingi K, Devan PAM, Assaad M. Forecasting of Wind Turbines Generated Power with Missing Input Variables. 2022 International Conference on Future Trends in Smart Communities (ICFTSC), Kuching, Sarawak, Malaysia, 2022, pp. 98-103. [CrossRef]

[21] Boudhaouia A, Wira P. Comparison of machine learning algorithms to predict daily water consumptions. IEEE Xplore 2021;1–6. [CrossRef]

[22] Torres JF, Álvarez FM, Troncoso A. A deep LSTM network for the Spanish electricity consumption forecasting. Neural Comput Appl 2022;34:10533–10545. [CrossRef]

[23] Jang J, Han J, Leigh SB. Prediction of heating energy consumption with operation pattern variables for non-residential buildings using LSTM networks. Energy Build 2022;255:111647. [CrossRef]

[24] Wang B, Zhang L, Ma H, Wang H, Wan S. Parallel LSTM-based regional integrated energy system multienergy source-load information interactive energy prediction. Complexity 2019;1–13. [CrossRef]

[25] Jang G, Kim N, Ha T, Lee C, Cho S. Base station switching and sleep mode optimization with LSTM-based user prediction. IEEE Access 2020;8:222711–222723. [CrossRef]

[26] Sarkar R, Julai S, Hossain S, Chong WT, Rahman M. A comparative study of activation functions of NAR and NARX neural network for long-term wind speed forecasting in Malaysia. Math Probl Eng 2019;3:1–14. [CrossRef]

[27] Khaleghi S, Karimi D, Beheshti SH, Hosen S, Behi H, Berecibar M, et al. Online health diagnosis of lithium-ion batteries based on nonlinear autoregressive neural network. Appl Energy 2021;282:116159. [CrossRef]

[28] Taherdangkoo R, Tatomir A, Taherdangkoo M, Qiu P, Sauter M. Nonlinear autoregressive neural networks to predict hydraulic fracturing fluid leakage into shallow groundwater. Water 2020;12:841. [CrossRef]

[29] Ghimire S, Huy TN, Prasad R, Deo RC, Perez DC, Sanz SS, Bhandari B. Hybrid convolutional neural network-multilayer perceptron model for solar radiation prediction. Cogn Comput 2023;15:645–671. [CrossRef]

[30] Chicco D, Warrens MJ, Jurman G. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. PeerJ Comput Sci 2021;7:e623. [CrossRef]

[31] Jiménez JAB, Estévez J, Marín APG. New machine learning approaches to improve reference evapotranspiration estimates using intra-daily temperature-based variables in a semi-arid region of Spain. Agric Water Manag 2021;245:106558. [CrossRef]