Uluslararası İleri Doğa Bilimleri ve Mühendislik Araştırmaları Dergisi
Sayı 8, S. 65-73, 2, 2024

Araştırma Makalesi

https://as-proceeding.com/index.php/ijanser
ISSN: 2980-0811

International Journal of Advanced Natural Sciences and Engineering Researches
Volume 8, pp. 65-73, 2, 2024

Research Article

Mid-Term Power Load Forecasting of a Statistically Modified Long-term Data by using the LSTM

Khalid Alhashemi[*1], Okkes Tolga Altinoz[1]

[1]*Department of Electrical and Electronics, Ankara University, Turkey*
Email of the corresponding author: kwhalhashemi@ankara.edu.tr

(Received: 01 March 2024, Accepted: 08 March 2024)

(4th International Artificial Intelligence and Data Science Congress ICADA 2024, March 14-15, 2024)

ATIF/REFERENCE: Alhashemi, K. & Altinoz, O. T. (2024). Mid-Term Power Load Forecasting of a Statistically Modified Log-term Data by using the LSTM. *International Journal of Advanced Natural Sciences and Engineering Researches, 8(2), 65-73.*

Abstract **–** The surplus power produced by power plants, which is considered as generation losses, can be avoided by estimating the expected load consumption, which will lead to financial gains for companies producing electrical energy. An accurate estimation of the power load can yield a reliable determination for power system management and the accompanying reduction of gas emitted from power plants. This work aims to create an integrated deep learning model based on a time series index to estimate future values of electric power consumption by applying Long Short-Term Memory (LSTM) networks. The dataset used has taken directly from PJM Interconnection Organization, which is a regional transmission organization in the United States, the data is an hourly power consumption in megawatt for Chicago and much area of Northern Illinois state. A statistical test was used to evaluate the dataset. Three different statistics functions have used for resampling the dataset, mean, minimum, and maximum function. After fitting the proposed model, it will predict the power load for one year ahead on daily basis. When the minimum function has used in the resampling processing the model was able to attain a Mean Absolute Percentage Error (MAPE) of 3.84%, and the coefficient of determination (R-squared) of 0.8.

Keywords – Load Forecasting, Long Short-Term Memory, Machine Learning, Short-Term Load Forecasting, Time-Series Analysis

I. INTRODUCTION

Having a comprehensive understanding of the precise amount of electricity needed for power generation has consistently been a crucial aspect in mitigating excessive losses from overproduction and averting any potential power interruptions, at this point at which obtaining energy consumption forecasts becomes an essential manner. Accurate forecasting also has its advantages when it comes to approximating the expenses involved in generating electricity and even in calculating the earnings of a local utility company. The growing electricity demand has made it challenging to design efficient electrical system components. The principal and most pressing aim of every electrical industry is to constantly and steadfastly provide reliable and uninterrupted service to their customers.

A variety of techniques can be used to determine the estimate of energy consumption, such as regression techniques, exponential smoothing, the straight line method, and artificial neural networks. The electric load in reality is nonlinear and can be influenced by different factors, including temperature and humidity [1]. As a result, the accuracy requirements of the contemporary power management system cannot be satisfied by the forecasting accuracy of the traditional nonlinear forecasting model.

Researchers have separated ways of estimating electricity consumption into two categories which are: statistical methods and artificial intelligence techniques. Statistical methods are straightforward and simple to use, but they can only handle data that has a limited number of features, require high sample data stationarity, and have low prediction accuracy. while the artificial intelligence techniques have high prediction accuracy, controllable generalization errors, and can effectively handle a large number of features and various data structures.

Over the last few years, we have observed a notable change in the global electric supply networks, which have swiftly transitioned towards smart grids, data management by cloud computing, and the integration of Internet of Things (IoT) technology with power grids. An essential part of this shift is being played by deep learning techniques.

The fundamental principle of deep-learning technology involves stacking multiple layers of neural networks and utilizing vast amounts of precisely annotated datasets, such as ImageNet, MSCOCO, etc. Moreover, a variety of training techniques are utilized, such as the mixup data augmentation technique, the ReLU activation function, and residual neural networks [2].

The application of deep-learning methodologies has led to immense strides in the area of video processing, image regression, robotics, autonomous vehicles, and medical diagnosis. However, other researchers have noted that deep learning techniques, such as Graph Neural Network (GNN), Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Generative Adversarial Networks (GAN), and Deep Belief Networks (DBNs), are highly beneficial for improving power load prediction through the optimal utilization of large amounts of time-series data [3]. The highly effective variant of deep-learning structure, RNNs, which have been meticulously engineered to handle temporal analysis and modeling, have already sparked considerable curiosity for their impressive flexibility in unearthing underlying non-linear and sequential relationships [4]. In recent years, RNNs have been highly praised for their exceptional performance in the domain of short-term load forecasting, largely due to their distinctive structural framework.

The RNN model encountered the vanishing gradient problem, resulting in disability to give satisfactory outcomes and highlighting a major weakness. The use of the same weights to evaluate performance at each stage of data preparation can lead to inaccurate outcomes [5]. However, the optimization of the activation function [6], and hyperparameters of the model such as number of layers, epoch, and number of layers [7] of the LSTM model has revolutionized the accuracy of time-based predictions in traditional RNNs, resulting in impressive achievements in the forecasting of the LSTM models.

II. MATERIALS AND METHOD

The cost of electricity generation and even the earnings of utility corporations can be estimated by precise midterm forecasting. It is of utmost significance to consider mid-term forecasting as a crucial aspect, as it brings forth a significant advantage, in addition to the aforementioned benefits. Furthermore, it plays a crucial role in enabling countries with a synchronized grid to effectively plan and allocate their financial resources.

A. *The LSTM Model*

LSTM networks, known as Long Short-Term Memory networks, have the exceptional ability to acquire and comprehend the intricacies of order dependency in sequence prediction tasks. This idea first came into light by Micheal C. Mozer in 1989, The core of his work was focused on back-propagation. Mozer worked on an equation of the activation context unit at the time step in which he considered the residual connection in the constant error carousel must set the real value to one [8]. Later in 1997, Moser’s paper was cited by Sepp Hochreiter and Jurgen Schmidhuber [9], these researchers introduced an efficient and gradient-based method called Long Short-Term Memory (LSTM), the initial version of the LSTM block included cells, input, and output gates have also proposed by them, moreover, they introducing the Constant Error Carousel (CEC) units. This proposed method has effectively tackled the issue of the vanishing gradient problem, which has long been a challenge in the field of deep learning and it has been the most successful RNN architecture and has been incredibly well-liked in several ensuing applications. In [10] authors have evaluated real data from residential smart meters and compared their results to various benchmarks, including the latest advancements in load forecasting, with comparing to other rival algorithms, LSTM model performs better in forecasting short-term load for individual residential households. Authors in [11] have utilized Recurrent Neural Networks RNN and their associated learning processes, they were able to achieve feasible meta-learning within large systems. The new system was created from the ground up and utilizes advanced learning algorithms to achieve high performance. Furthermore, the approach was proven to be successful in forecasting non-stationary time series. Through their experimentation in [12], the authors determined that the use of both neuro fuzzy logic and LSTM yielded more favorable results when compared to other approaches.

Recurrent Neural Networks (RNNs) are essentially different from traditional feedforward neural networks. In RNNs each hidden unit is not independent and their networks rely on sequences to establish the temporal correlations between past and present data. Figure 1 displays the unfolded perspective of a single RNN hidden unit. The neural network module A receives x_i as input, while h_i is the output. This cycle involves copying the same neural network several times and having each neural network module send the information to the subsequent one. This is how the information moves from one stage to the next. Except for input, the process was repeated at each stage of the RNN. This type of training approach ensures accuracy while significantly reducing the number of parameters the network has to learn and lowering training time.

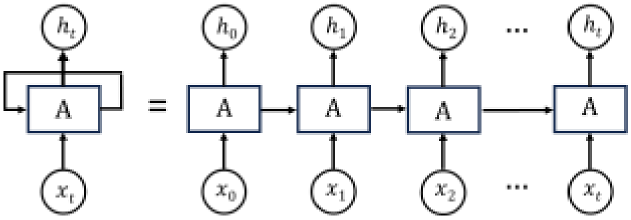

Fig. 1 Unfolded of basic RNN [1]

The time series index problem indicates that the choice made by an RNN at time t-1 can influence its subsequent decision at time t, this specific feature that RNNs possess makes them an excellent fit for resolving load prediction challenges encountered in the context of individual households [10]. It has been observed that acquiring long-term dependencies using RNNs poses a challenge, primarily attributed to the issue of vanishing gradients (opposite event), wherein the norm of the gradient for long-term components diminishes exponentially, in order to overcome this issue, the LSTM architecture with an extra forget gate was the solution.

A typical Long Short-Term Memory (LSTM) unit is composed of a cell structure, an input gate, an output gate, and a forget gate. The outputs derived from the three gates are connected in a manner that is specifically linked to regulating the input, output and status units in the multiplication unit of the network. Three control gates: input, output, and forget gates are shown in Figure 2.

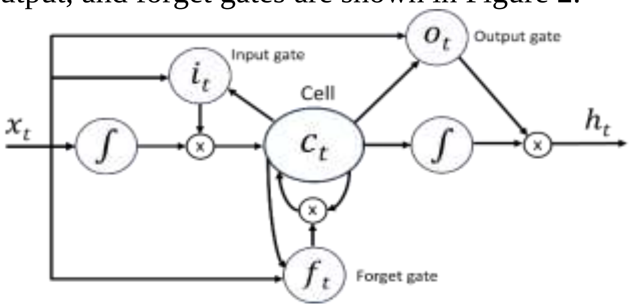

Fig. 2 Basic structure of LSTM [13]

At the initial moment of input processing, the output of the network will experience a continuous and ongoing impact as a result of this input. Consider $\{x_1, x_2, x_3, \ldots\ldots, x_T\}$ as the typical inputs sequence for an LSTM model, where each input vector of x_t is a real vector value. To establish temporal connections, the LSTM creates and preserves an internal memory cell state throughout its entire lifecycle. The intermediate output h_{t-1} and subsequent input x_t work together with the memory cell state S_{t-1} to decide which components of the internal state vector require to updating, erasing, or maintaining with taking into consideration the results from the previous state period and the inputs for the current state period. In addition to the internal state vector, the LSTM structure also includes input node c_t, input gate i_t, forget gate f_t, and output gate o_t. All nodes in the LSTM structure are represented by equation (1) to (6):

$$f_t = \sigma\left(W_{fx}x_t + W_{fh}h_{t-1} + b_f\right) \quad (1)$$

$$i_t = \sigma\left(W_{ix}x_t + W_{ih}h_{t-1} + b_i\right) \quad (2)$$

$$c_t = tanh\left(W_{cx}x_t + W_{ch}h_{t-1} + b_c\right) \quad (3)$$

$$o_t = \sigma\left(W_{ox}x_t + W_{oh}h_{t-1} + b_o\right) \quad (4)$$

$$s_t = \left(c_t \odot xi_t + s_{t-1} \odot f_t\right) \quad (5)$$

$$h_t = tanh(s_t) \odot o_t \quad (6)$$

W_{fx}, W_{fh}, W_{ix}, W_{ih}, W_{cx}, W_{ch}, W_{ox}, and W_{oh}, represent weight matrices, whereas b_f, b_i, b_c, and b_o represent bias vectors, $\odot$ refers to the pointwise multiplication of two vectors, σ refers to the sigmoid activation function which can represented by Equation 7:

$$\sigma_{(x)} = \frac{1}{1+e^{-x}} \quad (7)$$

tanh refers to the hyperbolic tangent function which can represented by Equation 8:

$$\tanh_{(x)} = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}} \quad (8)$$

ht in Equation (6) and h_{t-1} in Equation (1), (2), (3), and (4) are represent the output values at present time t and previous time $t-1$, respectively. Back to Figure 2, each × within the little circles signifies an element-wise multiplication performed between its inputs individually [14]. The large circular shapes with an S-shaped curve inside symbolize the use of a differentiable function, such as the sigmoid function, to a weighted sum. Furthermore, A dropout layer is consistently appended to each LSTM layer to inhibits overfitting in neural network modeling, which it will discuss later in the parameter and optimizer section.

B. *Construction of the Model*

The method proposed in this paper entails the concept of the LSTM algorithm within an improved model, with the aim of predicting future electric power based on a time series index on a daily basis for one year ahead. The dataset has split into nine years and one year, for training and testing set, respectively. There are some parameters that must be chosen, and they may be appropriate according to the dataset and the model in which it is intended to be built. The regression model has been set up and the errors have been measured to assess the precision of the predictions.

C. *Data Collection and Description*

The most difficult task for researchers developing a forecasting model is locating a high-quality dataset devoid of errors in typing, missing values, and overfitting. Therefore, in order to determine whether the dataset is useful for providing an accurate prediction, some statistical analysis must be performed. There are many different analyses that can be applied to the dataset, e.g., testing the probability value (p-value), that indicates the likelihood that the events in your data may have occurred in the event of the null hypothesis [15]. Generalized Sequential Pattern (GSP) is one of several data mining techniques that can also be used with LSTM. Since it deals with a large data set, it is used to identify patterns in sequential data. The goal of GSP mining is to discover patterns in data that occur over time, especially when it comes to electricity consumption which is greatly seasonally affected by many factors such as temperature during seasons, weekends or holidays days.

Selecting a reliable dataset is the first step in an effective prediction procedure. It is impossible to forecast a time series with accuracy unless the dataset has a temporal component.

In this paper, the dataset used is an hourly recorded of the power demand in megawatt MW that was taken directly from PJM Interconnection organization official website. PJM is a regional transmission organization that coordinates the movement of electricity in United States. PJM includes many electric utilities from which electricity is generated, transmit, and distributed. Due to regional changes throughout time, data could only be available for specific areas that is controlled by sub electric utilities. Therefore, in this paper the dataset was taken from the territory that Commonwealth Edison (ComEd) electric utility is responsible of. ComEd provides and services the electricity in Chicago and much of Northern Illinois State. The dataset is available at: https://dataminer2.pjm.com/feed/hrl_load_metered.

The dataset contains 87648 observations, starting from 01/01/2014 00:00 to 12/31/2023 23:00. with no null or missing values, the mean of the consumption is 11021.65 MW, the minimum and maximum consumption value was recording as 6775.822MW in 03/05/2020 at 07:00, and 22467.01MW in 24/08/2023 at 17:00, respectively.

Figure 3 illustrates the graphical representation of the full dataset indicates a small upward trend. However, this visual may not always convey an appropriate impression.

The histogram of the data distribution of the power recorded over the time is shown in Figure 4, at the first glance these data clearly seem are not normally distributed, in another words, they are somehow skewed (not symmetric). For better data distribution, the histogram must be balanced in both left and right tail when the mean is approximately equal to median. Moreover, the histogram shows the distribution of data around the vertical red line which is present the mean of the data.

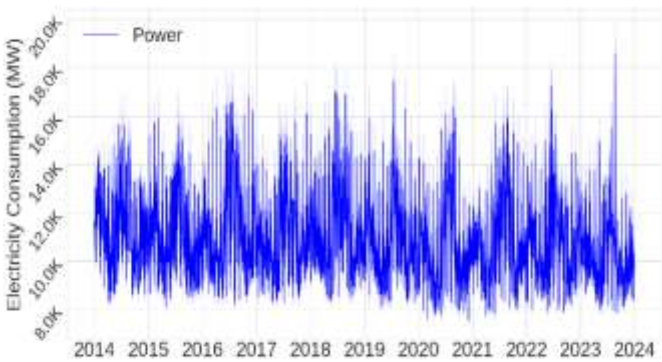

Fig. 3 Graphical representation of the dataset

The significance of the null hypothesis (H_0), has been evaluated by computing the p-value, which is the likelihood that the observed difference will occur by chance. The p-value provides evidence for reject or accept the null hypothesis (H_0), the less p-value is, the best dataset is, a dataset with p-value <= 0.05 is reject the null hypothesis (H_0) [16], and then data does not have a unit root and is stationary. The p-value for the dataset utilized in this study is 7.46×10^{-10}.

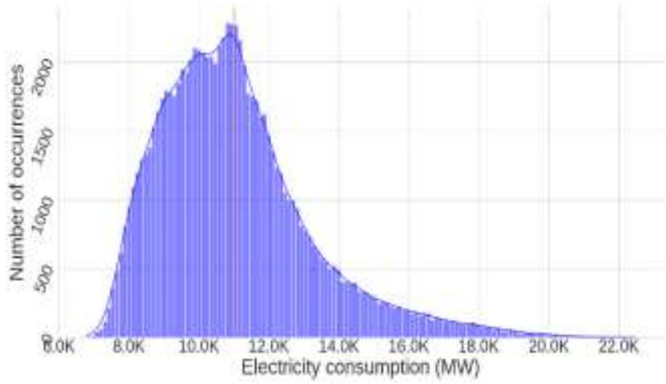

Fig. 4 Data distribution

D. *The Parameter and Optimizer*

All of the parameters that required to build the LSTM model or any model in the field of the machine learning must be satisfied in order to produce a successful model that produces the best outcomes with the least amount of inaccuracy. As previously mentioned, the dataset is recorded hourly within the scope of the start of each hour. Resampling has performed for each individual day by using three statistics functions, i.e. the mean, minimum, and maximum value. The prediction process has done three times with each mentioned statistics function.

As previously mentioned, the dataset was recorded hourly within the scope of the start of each hour. Resampling was performed for each individual day by using three statistics functions i.e., the mean, minimum, and maximum value. Therefore, the prediction process has been done with each mentioned statistics function. The data that has split into a training set is for the whole days between the years (2014 to 2022), thus it now has 3287 observations, in contrast, the data that has split into a test set is for whole days of the year 2023, thus it has 365 observations.

The subset of machine learning techniques built on artificial neural networks is known as deep learning. The utilization of multiple layers in the network is indicated by the word "deep ". Computational models consisting of several processing layers can acquire representations of data with multiple levels of abstraction through deep learning [17]. The essential component of deep learning is that these feature layers are not created by human engineers, rather, a general-purpose learning process is used to learn them from data. In this study, many layers have been involved to confirm the forecasting process to achieve optimal predictions. The first step for building the model with LSTM approach is pile up several layers that are the same or different such that one's output feeds into a subsequent one, it had done by using a sequential model that permits the accurate and sequential construction of a neural network from input to output by going through several neural layers one after the other. The first layer is a LSTM layer with 100 memory units and it returns the sequences. This is carried out to make sure that sequences, rather than merely randomly distributed input layers, are sent to the following LSTM layer. To avoid overfitting, a layer called the dropout layer is incorporated into neural network architectures. In this procedure, individual nodes are treated as if they were not a part of the network structure at all by being eliminated using a probability in several training iterations. Dropout layer has incorporated four time with each LSTM layer.

Dense Layer is also incorporated into neural network architectures, it is a fully connected and simple layer of neurons in which each neuron receives input from all the neurons of the previous layer. The dense Layer is used as the final stage of the neural network structure. The model employed the low-memory footprint and computationally efficient optimizer which is the Adaptive Moment Estimation (Adam) optimization technique that leverages momentum and adaptive learning rates to accelerate convergence, during the training of the model, the Mean Absolute Error (MSE) behaves as the loss function with each iteration.

The error components of MSE can be represented by Equation (9):

$$\text{MSE} = \frac{\sum_{i=1}^{n}(\hat{yi} - yi)^2}{n} \quad (9)$$

Where y_i denotes the actual value of the power consumption from the dataset, $\hat{yi}$ denotes the predicted value based on previous values, and n denotes number of samples.

The model was successfully trained and optimized through 88 epochs, each consisting of a batch size of 32. Over the course of an adequate number of training epochs, the model will ultimately produce results that exhibit progressive enhancement.

E. *Performance Metrics*

Any forecasting model may be assessed to evaluate its predicted outcomes as long as real data is available. The forecasting model's accuracy has been evaluated by using MAPE and (R-squared), their error components can be represented by Equations (10) and (11), respectively:

$$\text{MAPE} = \left(\frac{1}{n}\sum_{i=1}^{n}\frac{(yi - \hat{yi})}{yi}\right) * 100 \quad (10)$$

$$\text{R-squared} = 1 - \frac{SS_{\text{res}}}{SS_{total}} \quad (11)$$

$$SS_{\text{res}} = \sum_i (yi - \hat{yi})^2 \quad (12)$$

$$SS_{\text{total}} = \sum_i (yi - \hat{y})^2 \quad (13)$$

$$\hat{y} = \frac{1}{n}\sum_{i=1}^{n} yi \quad (14)$$

Where SS_{res} denotes the sum of squares of the residuals. SS_{total} denotes the total sum of squares. $\hat{y}$ denotes the mean of the observed data.

MAPE is often used in time series and regression models to assess prediction accuracy [18]. It indicates the error values as percentages. In other words, a MAPE of five percent signifies a five percent difference between the actual and predicted values. Whereas, R-squared is a number between 0 and 1 that presents how well the model fits the dependent variables or the “outcomes”.

III. RESULTS

The results obtained in this paper for prediction of the power consumed for one year have been done in three scenarios, the convenience method for frequency conversion and resampling of time series data was done by taking the mean, minimum, and maximum value of 24 recordings throughout one day to convert the data from the hourly basis into a less frequency to a daily basis. The efficiency of the model is relied on calculating the MAPE and R-squared for each scenario.

The Mape was measured for each resample process, it achieved 4.15%, 3.84%, and 5.94% for the statistics functions of mean, minimum, maximum, respectively.

As seen using the minimum function in the resampling process gives the higher prediction accuracy than other functions. These results calculate how accurate the forecasted quantities were in comparison with the real quantities. In the other hand, R-squared achieved 0.8, 0.71, and 0.76 for the statistics functions of mean, minimum, maximum, respectively. In a regression model, R-squared is a statistical metric that indicates how much of the variation in the dependent variable can be explained by the independent variable, what this means is the R-squared indicates how well the data matches the regression model.

Figures 5, 6, and 7 show the comparison results between the actual power consumption in MW, for the 2023 year for the territory that ComEd company provides the electricity in Chicago and much of Northern

Illinois state, in the United States (the blue curve), versus the estimated power consumption by using the statistics functions of mean, minimum, maximum for the resampling process, respectively.

It can be seen from these figures there are many small errors between the real and forecasted power loads. However, the scores earned in the model are considered reliable and quite accurate.

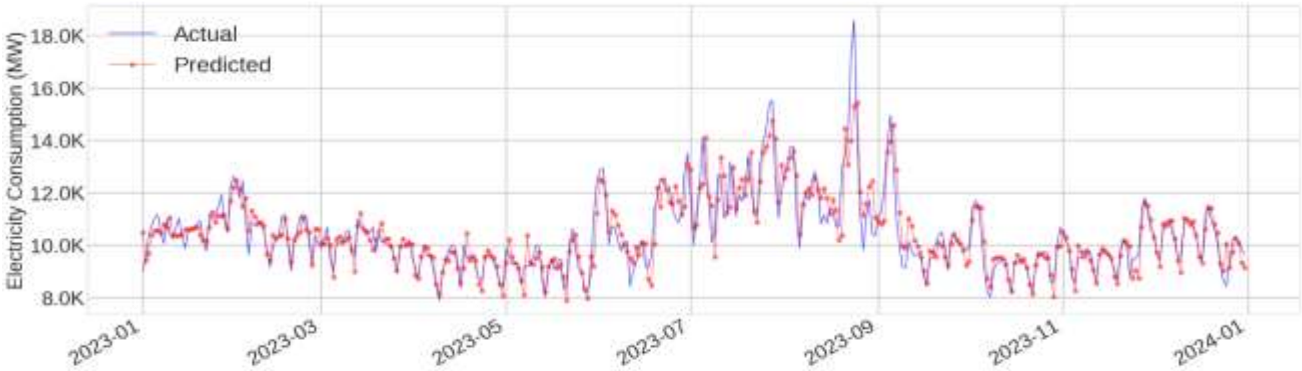

Fig. 5 Forecasting power load by using the mean function in the resampling process vs actual values

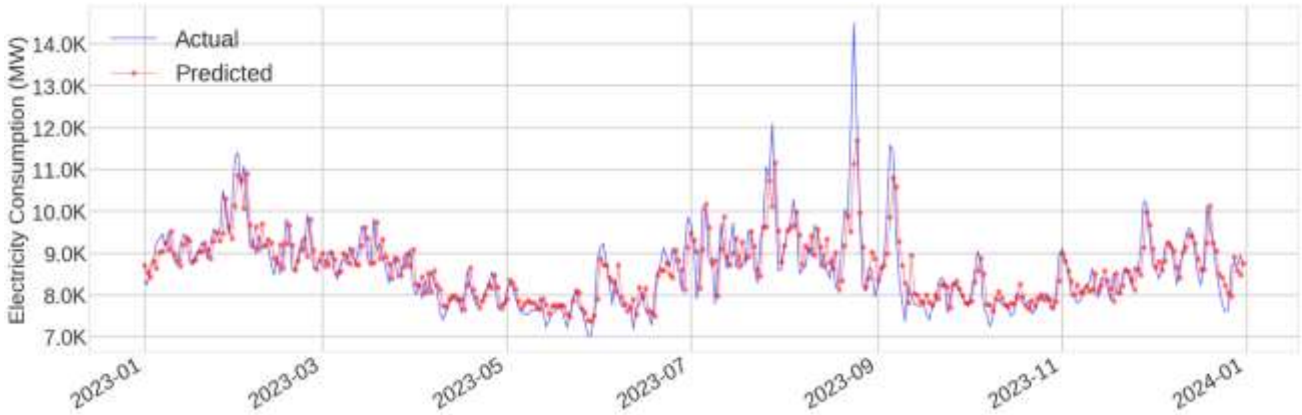

Fig. 6 Forecasting power load by using the minimum function in the resampling process vs actual values

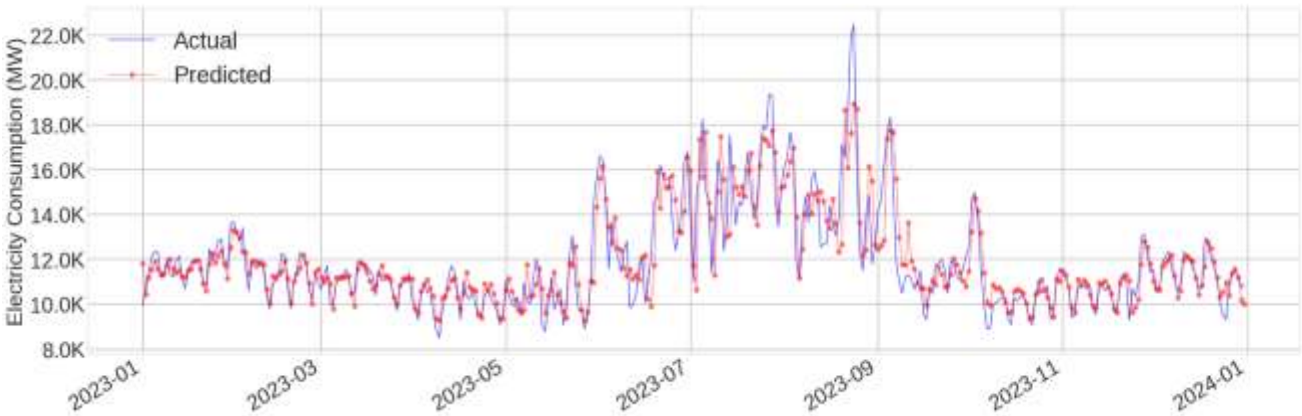

Fig. 7 Forecasting power load by using the maximum function in the resampling process vs actual values

The proposed model achieved MAPE of 3.84% when a minimum function was used for the resampling of the data, in this scenario R-squared was 0.71. In another hand, the best R-squared value was achieved as 0.8 when a mean function was used for the resampling of the data, and its MAPE was 4.15%, that is no necessarily means the minimum function is best than the mean function, because the R-squared used for measuring the fit, or accuracy of the model, but what it actually indicates is about variance, in the constant of MAPE, which indicates the percentage of the difference between the actual and predicted values.

IV. CONCLUSION

Mid-term power load forecasting plays an important role in the stability, and sustainability of synchronization between the power stations, and reduction of the gas emissions. In this study, A subtype of Recurrent Neural Networks (RNN), which is the LSTM model has been proposed. Statistical operations are significant in evaluating the dataset, also resampling the data, and even evaluating the forecasting model. The right choice of the hyperparameters helps to increase the accuracy of the forecasting. In this study, three functions were used for resampling the data. Using the minimum function gave the best forecasting results. In general, this model that used the LSTM has good predictability with a rapid response to anticipating declines that occur throughout the season during the year.

REFERENCES

[1] Chen, Z., Zhang, D., Jiang, H., Wang, L., Chen, Y., Xiao, Y., Liu, J., Zhang, Y. and Li, M, “Load Forecasting Based on LSTM Neural Network and Applicable to Loads of ‘Replacement of Coal with Electricity,’” *Journal of Electrical Engineering & Technology*, vol. 16, no. 5, pp. 2333–2342, Apr. 2021.

[2] X.B. Jin, W.Z. Zheng, J.L. Kong, X.Y. Wang, Y.T. Bai, T.L. Su, and S. Lin, “Deep-Learning Forecasting Method for Electric Power Load via Attention-Based Encoder-Decoder with Bayesian Optimization,” *Energies*, vol. 14, no. 6, p. 1596, Mar. 2021.

[3] N.-R. Son, S. Yang, and J. Na, “Deep neural network and Long Short-Term memory for electric power load forecasting,” *Applied Sciences*, vol. 10, no. 18, p. 6489, Sep. 2020.

[4] X.-B. Jin, N.-X. Yang, X. Wang, Y. Bai, T. Su, and J. Kong, “Integrated Predictor Based on Decomposition Mechanism for PM2.5 Long-Term Prediction,” *Applied Sciences*, vol. 9, no. 21, p. 4533, Oct. 2019.

[5] S. Sivamohan, S. N. Sridhar, and S. Krishnaveni, “An Effective Recurrent Neural Network (RNN) based Intrusion Detection via Bi-directional Long Short-Term Memory,” *2021 International Conference on Intelligent Technologies (CONIT)*.

[6] H. Zhou and Y. Tian, “Research on Time Performance Optimization of LSTM model for pedestrian volume prediction,” *2022 IEEE 10th Joint International Information Technology and Artificial Intelligence Conference* (ITAIC), Jun. 2022.

[7] I. S. Kervanci and F. Akay, “LSTM Hyperparameters optimization with Hparam parameters for Bitcoin Price Prediction,” *Sakarya University Journal of Computer and Information Sciences*, vol. 6, no. 1, pp. 1–9, Apr. 2023.

[8] M. C. Mozer, “A focused backpropagation algorithm for temporal pattern recognition.,” *Complex Systems*, vol. 3, Jan. 1989, [Online]. Available: https://dblp.uni-trier.de/db/journals/compsys/compsys3.html#Mozer89

[9] S. Hochreiter and J. Schmidhuber, “Long Short-Term memory,” *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.

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

[11] S. Hochreiter, A. S. Younger, and P. R. Conwell, “Learning to learn using gradient descent,” *in Lecture Notes in Computer Science*, 2001, pp. 87–94.

[12] S. Motepe, A. N. Hasan, and R. Stopforth, “Improving load forecasting process for a power distribution network using hybrid AI and deep learning algorithms,” *IEEE Access*, vol. 7, pp. 82584–82598, Jan. 2019.

[13] A. Graves, A. -r. Mohamed and G. Hinton, "Speech recognition with deep recurrent neural networks," *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, Vancouver, BC, Canada, 2013, pp. 6645-6649.

[14] Z. C. Lipton, J. Berkowitz, and C. Elkan, “A Critical Review of Recurrent Neural Networks for Sequence Learning,” *arXiv Preprint arXiv:1506.00019.*, May 2015.

[15] Ali Grami, "Hypothesis Testing," in *Probability, Random Variables, Statistics, and Random Processes: Fundamentals & Applications*, Wiley, 2019, pp.287-310

[16] G. Zhao, “A test of non null hypothesis for linear trends in proportions,” *Communications in Statistics - Theory and Methods*, vol. 44, no. 8, pp. 1621–1639, Apr. 2013.

[17] Y. LeCun, Y. Bengio, and G. E. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015.

[18] H. ÇetiNer and İ. ÇetiNer, “Analysis of Different Regression Algorithms for the Estimate of Energy Consumption,” *European Journal of Science and Technology*, no. 31, pp. 23–22, Dec. 2021.