Comparison the Performance of Different Optimization Methods in Artificial Intelligence Based Electricity Production Forecasting

Kübra KAYSAL
Department of Electrical Engineering,
Engineering Faculty,
Afyon Kocatepe University
Afyonkarahisar,Turkiye
kkaysal@aku.edu.tr

Fatih Onur HOCAOĞLU
Department of Electrical Engineering,
Engineering Faculty,
Afyon Kocatepe University,
Afyonkarahisar,Turkiye
hocaoglu@aku.edu.tr

Nihat ÖZTÜRK
Department of Electrical and
Electronics Engineering,
Technology Faculty,
Gazi University, Ankara, Turkiye
ozturk@gazi.edu.tr

***Abstract*— The purpose of this paper is to examine the performance of the LSTM method in Turkey's electricity production estimation and to determine the optimization technique that provides the best performance in the LSTM estimation method. In this study as a short term one hour's production forecast was made. For the forecasting, the last four years' real-time production data were analyzed hourly. Also, after the LSTM architecture was established in the study, the most suitable optimization method for the network was tried to be determined. The performance of each of them was calculated according to MAPE and R2 score criteria and the results were compared by using four different optimization techniques without changing the architectural structure created. According to the results obtained from Adam, Adamax, Nadam and RMSprop optimization techniques, proposed optimization method showed the best performance for this study with a 98% success rate.**

Keywords*— *short term load forecasting, optimization, energy forecasting.

I. INTRODUCTION

With the development of technology and population growth, the demand for energy production is increasing every day. For this reason, the accuracy of electricity forecasting has become increasingly important in ensuring supply-demand balance and investment planning in production, transmission and distribution systems. Excessive energy production causes storage problems and unnecessary use of energy, and insufficient use causes interruptions, and these situations prevent systems from working efficiently [1]. Forecasting of the energy production or demand is based on the study of past situations and data, making predictions about future situations. Production forecast depending on the time are evaluated as short, medium and long-term [2]. The most important criterion that should be provided by the developed algorithm is minimizing the errors. Optimization methods are used to minimize the error. Optimization methods are a step-by-step process. The amount of steps used in the process is called the learning coefficient, and choosing the optimal value in optimization is very important to ensure that the length of the estimate in terms of time and the minimum point are not missed. The most commonly used optimization methods in the literature are SGD, Adam, Adam, Adam and RMSPROP [3]. In there, authors used a single-layer and multi-layer artificial neural network, which is one of the deep learning methods, in their classification with a national data set and compared the performance of the results with various optimization techniques. According to the results, the best performance ratio was the man optimization technique with 92.31%.

In order to make more accurate predictions from the past to the present, various methods have been developed. In particular, the diversity of data brought about by the age of Artificial Intelligence, easy access to data and various intelligent techniques and forecasting methods have been observed in the literature studies that have achieved much more successful results.

Bodur and his colleagues used the RNN and LSTM model, one of the deep learning methods, for short-term load estimation [4]. They compared the performance of the two methods and observed that the LSTM method has a success rate of more than 90%. In reference 5, was used the LSTM method for daily solar radiation estimation and compared the results with machine learning methods [5]. For the method, the last 35 years of solar radiation data belonging to the province of Çorum are used. In the LSTM method, the MAPE value gave the lowest error result with 20.25% and was found to be more successful compared to machine learning methods. Kamber and colleagues forecasted the electricity demand with the LSTM method using Spain's one-year electricity generation data [6]. They compared the results with ARIMA, which is a time series analysis the results of both models showed similar performance. Kong at al made a short-term load estimate using data obtained from residential meters [7]. They compared the LSTM method for estimation with various machine learning methods according to the MAPE error criterion. LSTM performed best with a MAPE value of 36.52%. A short-term load estimate based on the LSTM method was made by Kwon [8]. According to the MAPE error criterion, LSTM was compared with some estimation methods and it was determined as the most successful method with MAPE 1.52% in the LSTM method. Using the LSTM method, Ma and his colleagues, who made a short-term load estimation, compared the model they developed with the Isolation Forest method according to the MAPE error criterion [9]. The MAPE value was found to be more successful in the LSTM method with 0.72%. In addition, the LSTM method has been compared with the ARMA, SARIMA and ARMAX used in the literature [10]. The MAPE value of 1.52 was found in LSTM and it is stated that it performs better than other methods.

In order to obtain more accurate and reliable forecast results, it is necessary to choose the forecasting methods and model architecture very well. Known that artificial intelligence covers Machine learning and Deep learning methods. In this study, one of the deep learning methods, LSTM (Long-Short Term Memory) method, was used and its performance was evaluated by various optimization methods. The results were found to be satisfactory.

II. LONG-SHORT TERM MEMORY(LSTM)

LSTM is a method developed as a solution to the problem of Gradient damping of recurrent neural networks (RNN). There are gates that decide what the model created should remember and what it should forget in this architectural structure that provides ease of processing long data entries. In LSTM if the input data is important, it takes this data to the next stage, contrary the memory forgets it. There are four sections in the LSTM architecture. These are the Forget Gate, which decides which information to forget or keep, the Cell State that provides data flow on the network, the Input Gate that updates the Cell State, and the Output Gate that decides the value that will be sent to the next layer [11]. An LSTM architectural structure is given in Figure1.

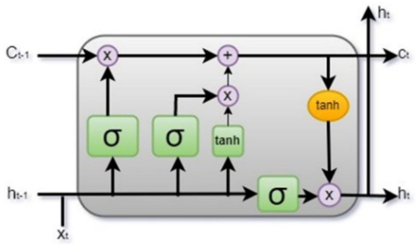

Fig. 1. LSTM model architectural structure

At the first step of the transfer mechanism among cells, it is decided which information should be kept and which information should be forgotten. This is accomplished by a sigmoid function of the neural network. If the output of the nerve has received a value of zero, the information is forgotten, if it has received a value, the information is retained. The result is found by Eq.-1 in this situation [12].

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

$h_{(t-1)}$ and x_t are the input values of the forgetting gate at the moment t. W_f is the weights of the forgetting gate, and b_f is the value of the bias [11].

The next step is for the input gate to decide what information to update. There are two cases here. In the first case, it is decided which values will be updated with a sigmoid function, while in the other case, the candidate state of the input cell is obtained with the *tanh* function. The formula obtained for the output i_t and the candidate state $\widetilde{C_t}$ of the input cell is given in Eq. 2 [12].

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

$$\widetilde{C_t} = tanh(W_c[h_{t-1}, x_t] + b_c) \quad (2.b)$$

Where W_i and W_c represent weights, b_i and b_c are bias values. Eq.3 gives the formula for the updated state of cells C_t for time *t*.

$$C_t = f_t * C_{t-1} + i_t * \widetilde{C_t}) \quad (3)$$

Obtaining the output value, $h_{(t-1)}$ and x_t are taken as input. The Eq.4 used to find the result obtained from the exit gate can given as follow,

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \quad (4)$$

Lastly $h_{(t)}$ is found with Eq.5.

$$h_t = o_t + tanh(C_t) \quad (5)$$

A. Optimization Algorithms

In order to minimize the error in artificial intelligence algorithms, various optimization methods are used. Some of them are SGD, Adamax, Adagrad, RMSProp, Adam and Adadelta methods [3].

B. Stochastic gradient descent algorithm

One of the popular algorithms is Stochastic Gradient Descent (SGD) which is used training a wide range of models in machine learning, support vector machines etc. For each training sample, the parameter is updated according to Eq. 6 given below by the SGD.

$$w_{t+1} = w_t - n\frac{\partial L}{\partial w_t} \quad (6)$$

C. Adaptive Movement Estimation (Adam) Algorithm

The objective of the method is to optimize the functions based on first-order gradient based on adaptive estimates of moments. This method is frequently used in the literature because it is easy to implement, computationally efficient, has little memory requirement and is suitable for the optimization of big data[3,14]. Below are the equations used during the implementation of the algorithm.

$$w_{t+1} = w_t - \frac{n}{\sqrt{\hat{s}_t + \epsilon}} . \widehat{V_t} \quad (7)$$

$$\widehat{V_t} = \frac{v_t}{1-\beta_1^t} \quad (8)$$

$$\widehat{s_t} = \frac{s_t}{1-\beta_2^t} \quad (9)$$

$$V_t = \beta_1 V_{t-1} + (1-\beta_1)\frac{\partial L}{\partial w_t} \quad (10)$$

$$S_t = \beta_2 S_{t-1} + (1-\beta_2)[\frac{\partial L}{\partial w_t}]^2 \quad (11)$$

D. AdaMax algorithm

AdaMax is a generalisation of Adam based on the infinity norm. The modified norms are given in Eq.12 and 13. More broadly, is an extension to the Gradient Descent Optimization Algorithm [15].

$$S_t = \beta_2^\infty S_{t-1} + (1-\beta_2^\infty)[\frac{\partial L}{\partial w_t}]^2 \quad (12)$$

$$w_{t+1} = w_t - \frac{n}{S_t} . \widehat{V_t} \quad (13)$$

In there the initial values of S and V should be set to zero, *n=0.002*, β_1*=0.9,* and β_2*=0.999* [13].

E. Nesterov-accelerated Adaptive Moment Estimation (Nadam) algorithm

This is an extension of the Adam algorithm that incorporates Nesterov momentum and can result in better performance of the optimization algorithm [15].

$$w_{t+1} = w_t - \frac{n}{\sqrt{\hat{s}_t + \epsilon}} \cdot \widehat{V_t} \quad (14)$$

$$\widehat{V_t} = \left(\frac{\beta_1 * \beta_{1t}}{1-\beta_1}\right) + \left(\frac{1-\beta_1 * \frac{\partial L}{\partial w_t}}{1-\beta_1}\right) \quad (15)$$

$$\widehat{s_t} = \left(\frac{\beta_2 * \beta_{2t}}{1-\beta_2}\right) \quad (16)$$

There are three hyperparameters for the algorithm; they are: n: Initial step size (learning rate), a typical value is 0.002. Decay factor for first moment β_1 a typical value is 0.975. Second moment β_2 a typical value is 0.999. [15].

F. Root Mean Squared Propagation (RMSprop) algorithm

RMSprop is another algortihm which was inspired gradient descent algorithm.It uses a decaying average of partial gradients in the adaptation of the step size for each parameter. It performs the update by multiplying the current value by a learning coefficient [3,13].

$$w_{t+1} = w_t - \frac{n}{\sqrt{\hat{s}_t + \epsilon}} \cdot \frac{\partial L}{\partial w_t} \quad (17)$$

$$S_t = \beta S_{t-1} + (1-\beta)[\frac{\partial L}{\partial w_t}]^2 \quad (18)$$

III. METHOD AND SIMULATION RESULTS

In the study, real-time electricity generation data of the last four years of Turkey obtained from EPIAS were used. 80% of the data on 35075 hourly frequencies were organized as training and 20% as test data. Fig. 2 shows a graph of the hourly values of the data.

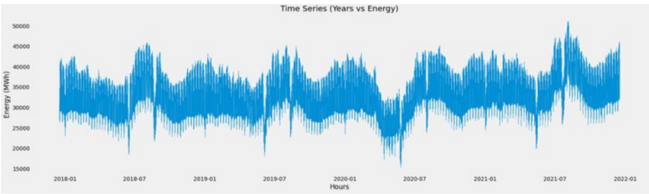

Fig. 2. Hourly data graph

After the data was uploaded, the scaling process was performed between 0 and 1. A three-layer LSTM model is designed. The training was conducted in 20 epoch. In the LSTM model, Adam, AdaMax, Nadam and RMSProp optimization techniques were used and their results were compared. The most commonly used MAPE (Mean Absolute Percentage Error) and R-squared (R^2) -approximation of the estimate to the truth- were used as performance measurements in the literature.

When the study results were evaluated, the results obtained with the Adam optimization technique performed more successfully than others. The comparison table obtained from the results is given in Table 1.

TABLE I. RESULTS OF OPTIMIZATION

LSTM Optimizers	MAPE	R^2Score
Adam Optimizer	0.014	%98
RMSprop Optimizer	0.022	%95
Adamax Optimizer	0.022	%95
Nadam Optimizer	0.043	%95

As a result of the study, the error-prediction graphs obtained from each optimization method are shown as in Fig. 3-6.

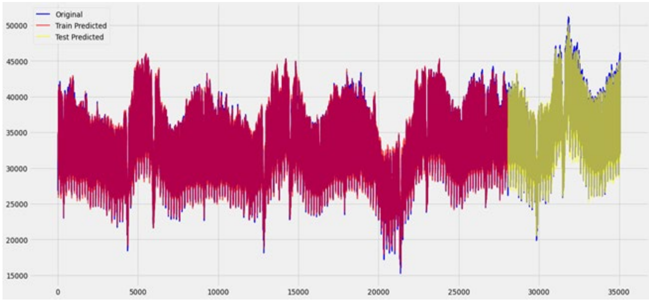

Fig. 3. Adam optimization error prediction graph.

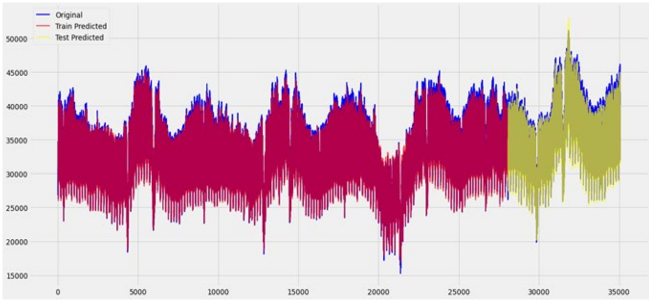

Fig. 4. AdaMax optimization error prediction graph.

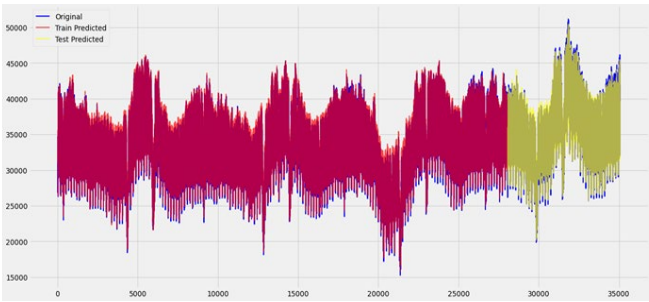

Fig. 5. Nadam optimization error prediction graph.

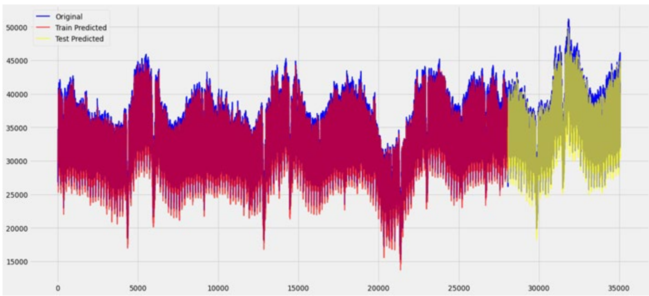

Fig. 6. RMSprop optimization error prediction graph.

IV. DISCUSSION AND COMPARISON

Table 2 gives a comparison of the results obtained from this study and some of the similar studies conducted in the literature. When similar studies were examined, it was seen that the data set used in the study was sufficient for forecast. In addition, the *MAPE* value turned out to be quite low compared to other studies, which proves that the error rate in this estimate is quite low. Another benchmark, the R^2 Score value, is understood to be quite successful compared to other studies with a 98% success rate.

TABLE II. THE COMPARISON OF THE RESULTS PROPOSED LSTM TECHNIQUE AND SOME STUDIES IN THE LITERATURE.

Paper	Year	Method	MAPE	R^2 Score	Data Set
Kong et al. [7]	2019	LSTM	% 8.18	-	29808 real data
Muzaffar and Afshari [10]	2019	LSTM	% 1.52	-	10000 real data
Kwon et al. [8]	2020	LSTM	% 1.49	-	8760 real data
Ma et al. [9].	2019	LSTM	% 0.88	-	35712 real data
Seyyarer et al. [3]	2020	LSTM (Adam)	-	%92	-
Farsi et al. [16]	2021	LSTM	% 3.11	%96.6	17518 real data
Proposed	2022	LSTM	% 0.14	%98	35075 real data

Nowadays, as technology is progressing rapidly, the world's population is growing at an even faster pace. These such situations also increase the need for electrical energy every day. It is very important that electricity is produced in sufficient quantity in accordance with supply and demand. To ensure a good balance of production, transmission and distribution is to make an accurate and reliable forecast. Compared to traditional forecasting methods, it is possible to make predictions with less error, high accuracy and reliability with more intelligent forecasting systems based on Artificial Intelligence and accurate optimization techniques. In this study, it was observed that the energy production estimation of LSTM and Adam optimization technique, one of the forecasting methods based on artificial intelligence, also achieved successful results.

REFERENCES

[1] T. Akman, C. Yılmaz and Y.Sonmez, "Analysis of Electrical Load Forecasting Methods," Gazi Journal of Engineering Sciences, 4(3):168-175, 2018.

[2] F.O. Hocaoğlu, K. Kaysal and A. Kaysal, "Hybrid Model for Load Forecasting (ANN and Regression)". APJES III-II, 33-39, 2015.

[3] E. Seyyarer, F. Ayata, T. Uçkan and A. Karcı, "Applications and Comparison of Optimization Algorithms used in Deep Learning", Anatolian Journal of Computer Sciences, vol. 5, no:2, pp:90-98, 2020.

[4] I. Bodur, E. Celik and N. Ozturk, "A Short-Term Load Demand Forecasting based on the Method of LSTM," 2021 10th International Conference on Renewable Energy Research and Application (ICRERA), 2021, pp. 171-174. İstanbul, Turkiye

[5] A. Kara, "Global Solar Irradiance Time Series Prediction Using Long Short-Term Memory Network", GU J. Sci, Part C, vol.7(4), pp.882-892, 2019.

[6] E. Kamber, S. Körpüz, M. Can, H.Y. Aydoğmuş, M. Gümüş, "Short Term Electrical Load Forecasting Based on Artificial Neural Networks", Journal of Industrial Engineering, vol. 32(2), pp.364-379, 2021.

[7] 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, 2019.

[8] B.S. Kwon, R.J. Park, and K.B. Song, "Short-Term Load Forecasting Based on Deep Neural Networks Using LSTM Layer", Journal of Electr. Eng. Technol. vol. 15, pp. 1501–1509, 2020.

[9] Y. Ma, Q. Zhang, J. Ding, Q. Wang and J. Ma, "Short Term Load Forecasting Based on iForest-LSTM," 2019 14th IEEE Conference on Industrial Electronics and Applications (ICIEA), 2019, pp. 2278-2282, 2019.

[10] S. Muzaffar and A. Afshari, "Short-Term Load Forecasts Using LSTM Networks", Energy Procedia, 10th International Conference on Applied Energy (ICAE2018), vol.158, pp.2922-2927, August 2018, Hong Kong,China.

[11] J. Wang, J. Li, X. Wang, et al. "Air quality prediction using CT-LSTM", Neural Computing & Applications vol. 33, pp.4779–4792, 2021.

[12] H. Tian, P. Wang, K. Tansey, J. Zhang, S., Zhang, and H. Li, "An LSTM neural network for improving wheat yield estimates by integrating remote sensing data and meteorological data in the Guanzhong Plain, PR China", Agricultural and Forest Meteorology , vol. 310, 108629, 2021.

[13] S.Ruder, "An overview of gradient descent optimization algorithms", https://ruder.io/optimizing-gradient-descent/ (accessed Jan. 26, 2022).

[14] D.P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization", https://arxiv.org/abs/1412.6980 (accessed Jan. 26, 2022).

[15] J. Brownlee, "Gradient Descent Optimization with Nadam From Scratch", https://machinelearningmastery.com/gradient-descent-optimization-with-nadam-from-scratch (accessed Jan. 26, 2022).

[16] B. Farsi, M. Amayri, N. Bouguila and U. Eicker, "On Short-Term Load Forecasting Using Machine Learning Techniques and a Novel Parallel Deep LSTM-CNN Approach," in IEEE Access, vol. 9, pp. 31191-31212, 2021.