JIKO (Jurnal Informatika dan Komputer)
Vol. 5, No. 3, Desember 2022, hlm. 243-248
DOI: 10.33387/jiko

Akreditasi KEMENRISTEKDIKTI, No.105/E/KPT/2022
p-ISSN: 2614-8897
e-ISSN: 2656-1948

TRANSFORMER IMPLEMENTATION FOR SHORT TERM ELECTRICITY LOAD FORECASTING, CASE STUDY: BALI, INDONESIA

Arionmaro A Simaremare[1], Indra A Aditya[2], Didit Adytia[3], Isman Kurniawan[4]

[1,2]Generation Division, PLN Research Institute, Indonesia
[3,4]School of Computing, Telkom University, Bandung, Indonesia
Email: [1]arionmaro@pln.co.id, [2]ismankrn@telkomuniversity.ac.id, [3]adytia@telkomuniversity.ac.id

(Received: 18 November 2022, Accepted: 29 November 2022)

Abstract

Short term load forecasting is a crucial process in ensuring optimal and reliable operation of electric power system which is critical in sustaining highly technological economies. Various approaches and methods have been implemented in forecasting electricity load of a system including statistical methods such as auto regression and machine learning methods such as support vector machine and also deep learning methodology as recurrent neural network methodology which gain popularity in electricity load forecasting nowadays. In this paper, Transformer as a deep learning methodology is used to forecast hourly electricity load in Bali Area. Three lookback days scenario and ten days of forecast period are used to evaluate the performance of the Transformer models. This study suggest that although higher lookback days will give more complicated model due to increasing number of parameters involved, the best overall prediction performance are given by transformer model with 1 day of lookback period. The three model in this study also tend to have low prediction performance in predicting electricity load for weekend or holiday period. Future study using multivariate transformer model is suggested to improve the prediction performance of the transformer model in predicting electricity load in Bali area.

Keywords: *electricity load, transformer, time series, machine learning*

1. PENDAHULUAN

Electricity is critical to sustaining highly technologically advanced industrialization in all economies [1], [2]. In this modern era, electricity is essential to nearly all activities. As the years' pass, the global demand and consumption of electrical energy increase [3], but electrical energy generation, transmission, and distribution remain complex and expensive. Consequently, good grid management is crucial for reducing the cost of energy production and increasing the generation capacity to satisfy the rising demand for electric energy [4]. Consequently, effective grid management necessitates good load demand planning, an adequate maintenance schedule for generating, transmission, and distribution lines, and efficient load distribution via supply lines. Therefore, effective load forecasting will go a long way toward optimizing the planning process in the power generation industries [4], [5].

Numerous approaches and algorithms have been implemented for electrical load forecasting. In general, these approaches can be categorized into two primary groups, i.e., (i) statistical models and (ii) machine learning models [6], [7]. There are several popular statistical models that were used for electricity load forecasting, such as seasonal autoregressive [8], threshold autoregressive [9], ARIMA [10], SARIMA [11], etc. In contrast to the extensive use of statistical approaches, machine learning (ML) techniques have gained favor due to their efficacy, precision, and adaptability. In recent years, the availability of data (as a result of digitization) and the affordability of essential computer resources have made machine learning approaches an unavoidable option in this respect [12].

There are several ML techniques that has been implemented for electricity load forecasting. In 2010, Li et al. utilized several ML techniques to forecast annual residential energy demand in China. They found that support vector machine (SVM) model gives the best performance compared to other ML methods [13]. In 2013, Jain et al. also applied SVM techniques to investigate the temporal and spatial effects on the model performance. The results show the capability of the model in forecasting the energy demand of residential floors on an hourly scale [14]. In 2018, Ruiz-Abellón et al. predict the multivariate short-term electrical consumption by using regression tree methods. By adding calendar variables and temperatures, they obtained the best results from Random Forest method with short time of computational process [15]. Also, ML techniques are popular for prediction tasks in other fields [16].

Recently, deep learning (DL), as a sub-technique of ML, was frequently used for various energy-forecasting tasks because of its ability to model nonlinearity [17]. Singh et al. identified three primary characteristics that influenced the emergence of deep learning algorithms for short-term load forecasting tasks, which are its adaptability to be scaled on massive data, its capacity to do unsupervised feature learning, and its tendency for generalization [18]. Several review studies reported that among DL techniques, recurrent neural network (RNN) based methods give better performance than other deep algorithms [19]–[21]. Unfortunately, there are several drawbacks of RNN-based methods, such as the problem of gradient disappearance during time series processing [22] and the difficulty of processing in parallel mode, leading to time consumption [23]. To solve those drawbacks, the transformer method has been proposed by Vaswani et al. This method enables a parallel process for sequence type of input [24]. However, despite its advantage, implementing the transformer in electricity load forecasting is still very rare.

This study aim to implement time series-based Transformer method in electricity load forecasting. We used per hour electricity load data in Bali, Indonesia, for period of March 2020 until August 2021. The transformer prediction model was developed by varying the transformer block and head attention number. The performance of the model was evaluated by calculating several validation parameters, such as correlation coefficient (CC), root mean square of error (RMSE), and mean absolute percentage of error (MAPE).

2. MATERIAL AND METHODS

2.1. Data Set

The data set used in this study is per hour electricity load data in Bali, Indonesia, for period of March 2020 until August 2021, as presented in Figure 1. We splitted the data set into train, validation and test set with splitting scheme was illustrated in Figure 2. To develop the forecasting model, we utilized three look back values, i.e. 1, 3 and 7 days, and evaluate the contribution of the look back on the model performance. Here, the look back data was used as features to predict the electricity load of the next time. Hence, we tranformed the time series data of electricity load into set of feature and target.

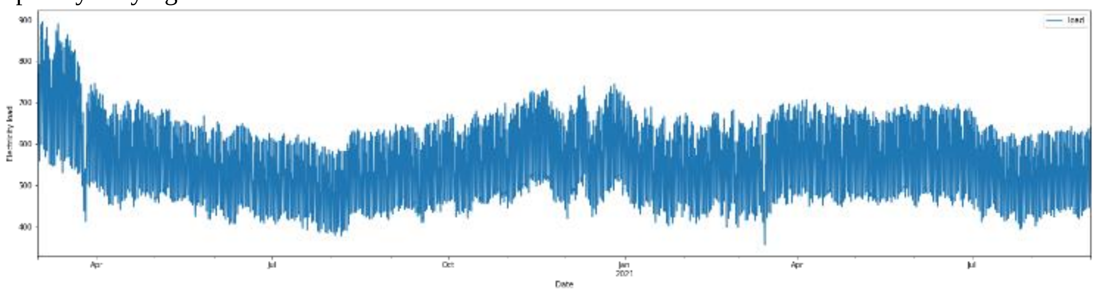

Figure 1. Time series plotting of electricity load

Figure 2. The scheme of data set splitting into train, validation, and test set

2.2. Transformer Model

The transformer model [24] was initially introduced for machine translation, but due to its great performance, it was rapidly implemented into other fields, including image generation, audio, text summarization, and music. The transformer does not employ recurrent or convolution but instead models with an attention module. The transformer utilizes the encoder-decoder technique. The input is initially entered into the encoder, and the output is then generated based on the encoded input and the prior outputs in the decoder.

Encoders are stacked in the part of the encoder. The number of encoders in a stack is a free parameter, which is typically six layers. A stack of decoders equal to the number of levels in the encoder is utilized in the decoder. Each encoding layer has its own parameters; these layers do not share a common weight. The transformer, unlike recurrent networks, has no difficulty with vanishing gradient and can reach any point in the past regardless of the distance between words. This technique permits the transformer to identify long-term dependencies. In addition, unlike recurrent networks, the transformer does not require sequential computing and can operate in complete parallel at high rates.

The vanilla Transformer [Vaswani et al., 2017] follows the encoder-decoder structure of most of the competitive neural sequence models. Each encoder and decoder consists of several identical building blocks. Each encoder block consists of a multi-head self-attention module and a position-wise feed-forward network (FFN), while each decoder block is put between the multi-head self-attention module and the position-wise feed-forward network (FFN).

The self-attention module applied by using the scale dot-product attention with Query-Key-Value (QKV) as formulated as:

$$Attention(\mathbf{Q},\mathbf{K},\mathbf{V}) = softmax\left(\frac{\mathbf{QK^T}}{\sqrt{D_K}}\right)\mathbf{V} \quad (1)$$

where D_k represent the key dimension. Transformer employs multi-head attention (MHA) with H distinct sets of learned projections instead of single attention function formulated as follows:

$$MultiHeadAttention(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \\ = Concat(head_1, \ldots, head_H)\mathbf{W}^O \quad (2)$$

where $head_i = Attention\left(\mathbf{Q}\boldsymbol{W}_i^Q, \mathbf{K}\boldsymbol{W}_i^K, \mathbf{V}\boldsymbol{W}_i^V\right)$. Then, MHA is linked to feed-forward network that is fully connected module as follows:

$$FFN(\mathbf{H}') = ReLU\left(\mathbf{H}'\mathbf{W}^1 + \mathrm{b}^1\right)\mathbf{W}^2 + \mathrm{b}^2 \quad (3)$$

where **H**' means the output from the previous layer, $\mathbf{W}^1$, $\mathbf{W}^2$, b^1, b^2 are trainable parameters.

In this study, the transformer model was developed by optimizing two architecture parameters, transformer block, and head attention number. Both parameters were considered hyperparameters that were optimized by using a manual tuning approach. A total of six transformer models were developed that were derived from the combination of two transformer block values (1 and 2) and three head number values (1, 2, and 3). Besides both parameters, other hyperparameters of the transformer model were set to fix, as presented in Table 1.

Table 1. The Fix Hyperparameter of Transformer Model

Hyperparameter	Value
Head size	128
Convolution layer (FF)	4
Act. func. in FF	relu
Hidden node in FC	128
Act. func. in FC	relu
Learning rate	1 x 10^{-4}
Epoch	200
Patience callback	10

2.3. Model Validation

To validate the model, we used Transformer model to predict maximum 10 days of the test set. However, the prediction did not performed in one batch using all the test. Instead, we conducted the prediction hour by hour by utilizing the predicted value for the next hour prediction. Hence, we consider the accumulation of error to investigate the ability of the model to overcome the issue. Also, this approach represent the prediction power of the model and point out the limitation of model applicability in the term of time series data. Validation parameters that were calculated to evaluate the model performance consists of correlation coefficient (CC), root mean square of error (RMSE), and mean absolute percentage of error (MAPE). The calculation of those parameters are formulated as follows:

$$\mathrm{CC} = \frac{\sum_{t=1}^{N}(X_t^a - \overline{X^a})(X_t^p - \overline{X^p})}{\sqrt{\left[\sum_{t=1}^{N}(X_t^a - \overline{X^a})^2\right]\left[\sum_{t=1}^{N}(X_t^p - \overline{X^p})^2\right]}} \quad (4)$$

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{t=1}^{N}(X_t^a - X_t^p)}{N}} \quad (5)$$

$$\mathrm{MAPE} = \frac{100\%}{N}\sum_{t=1}^{N}\left|\frac{X_t^a - X_t^p}{X_t^a}\right| \quad (6)$$

where X_t^a represent actual electricity load at time t, $\overline{X^a}$ represent the average value of actual electricity load, X_t^p represent predicted electricity load at time t, $\overline{X^p}$ represent the average value of predicted electricity load, and N is total number of data.

3. RESULT AND DISCUSSIONS

3.1. Model Fitting

Table 2 shows the model architecture that gives the best performance regarding the correlation coefficient, RMSE, and MAPE for different lookback periods. From Table 2, we can see that the higher the number of lookbacks, the more complicated the model is regarding the number of transformer blocks and head number. This is also aligned with the parameter number involved in the model, where the higher the lookback period, the more parameters are involved.

In the model fitting process, we monitor the value of loss-value for the train and validation dataset. From Figure 3, we can see how the loss value for the train and validation dataset progress, starting very high value in the first stage of model fitting and getting lower as the model fitting continues. We can also see how the loss-value for test and validation datasets are close to each other at the end of the model-fitting process, which indicates a good model-fitting process. In addition, although the epoch number set in the model fitting is 200, the model fitting process stopped at the epoch number of around 160 due to the patience callback parameter, which is set at a value of 10. The resulting time series plot of the electricity load for training, validation, and test dataset is given in Figure 4.

3.2. Model Evaluation

After the best model for each lookback is obtained, the model is then tested against the test dataset to evaluate how each model performs for each lookback and prediction length. Figure 5 shows the plot of electricity load forecasting using the Transformer model for ten day forecasting period for each lookback number. As we can see from Figure 5, qualitatively, we can see that although all three lookback setting performs poorly for the first forecasting day, lookback day 3 gives the best fitting performance for the ten days prediction period. Lookback 1, in general, fails to predict the electricity load at the high-value region, while on the other hand, lookback 3 fails to predict the electricity load at the lower-value region. Lookback 7, on the other hand, fails to predict the electricity load both in the higher and lower load region. Upon further evaluation, we can see that all electricity load forecasting models fail to predict electricity load for the weekend or holiday period. This might be caused by the missing information in the feature regarding the day position in a week.

Table 2. Model Architecture

Look back	Transformer block	Head number	Val score	Parameter number
1	2	1	0.007	5,139
3	1	2	0.006	11,274
7	2	3	0.007	27,155

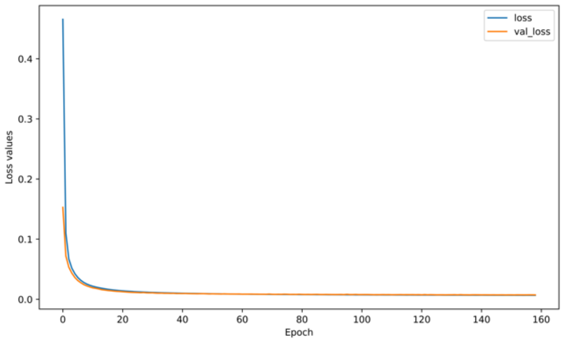

Figure 3. Learning curve of the best transformer architecture for look back 1 day

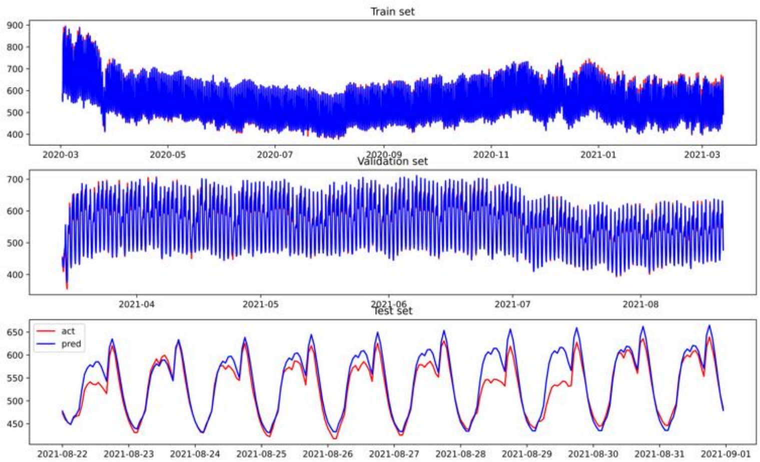

Figure 4. Time series plot of electricity load forecasting for transformer with look back 1 day

Table 3 provides the quantitative performance for all three lookback models for ten days of the prediction period. From Table 3, we can see that all three lookback models give poor prediction performance for the first prediction day in terms of RMSE and MAPE value while still giving a good correlation coefficient value. Even though it is expected to have high fitting performance for a prediction length of one day, high vertical error for the first day for all three lookback day models indicates that the model performs poorly in predicting electrical model for weekend/holiday periods due to lower than average electricity load at those time period. This indication is also emphasized by low prediction performance for all three models for the seventh and eighth days of the prediction period, which also includes the weekend period (Saturday and Sunday). Overall, the model for lookback day 1 gives the best performance in terms of CC, RMSE, and MAPE, followed by the lookback 3 models, especially for prediction lengths of 3 and 5 days. Lookback 7 model gives the lowest performance of all three models. This may suggest that in electricity load forecasting, a more complicated model and a higher number of parameters cannot guarantee a better prediction performance.

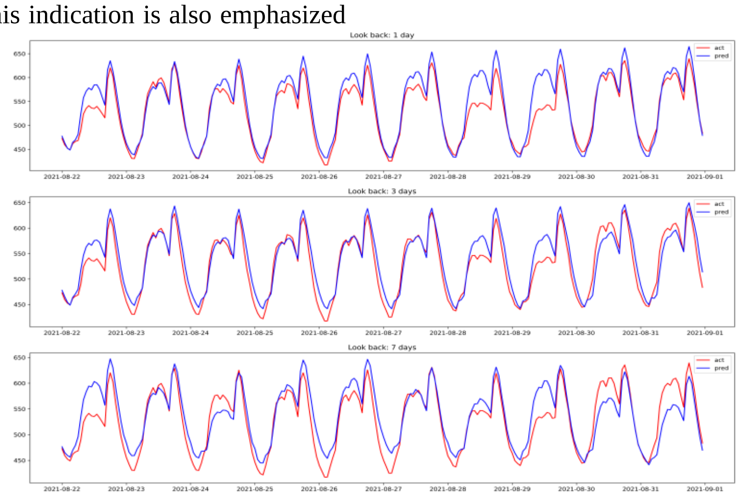

Figure 5. The comparison of time series plot of electricity load forecasting

Table 3. Quantitative Performance of Transformer Models

Look back	Prediction Length			
	1	3	5	10
CC				
1	0.969	0.971	0.982	0.965
3	0.983	0.972	0.971	0.957
7	0.959	0.907	0.934	0.903
RMSE				
1	26.235	17.642	17.532	23.536
3	23.912	19.086	18.979	20.360
7	38.473	28.021	27.742	27.655
MAPE				
1	3.768	2.330	2.560	3.016
3	3.711	2.897	2.898	3.112
7	5.847	4.304	4.365	4.141

4. CONCLUSION

Machine learning methodology, such as support vector machine, and deep learning methodology, such as recurrent neural network, has been widely implemented in short-term electricity load forecasting. In this study, the Transformer methodology, as one of the deep learning methods, is used to forecast hourly electricity load in the Bali area. This study uses three lookback scenarios and ten days of prediction time to evaluate the performance of the transformer models. This study found that a higher lookback period will require a more complex transformer model architecture due to the increasing number of parameters involved. However, the complexity of the model does not necessarily increase the performance of electricity load prediction. This study found that the transformer model with one day of lookback period gives the best overall forecasting performance while, on the contrary, the transformer model with 7 days of lookback period gives the worst overall forecasting performance. A future study using a multivariate instead of a univariate transformer model is suggested to improve the prediction accuracy of the transformer model in electricity load forecasting.

5. REFERENCES

[1] A. Dedinec, S. Filiposka, A. Dedinec, and L. Kocarev, 2016. “Deep belief network based electricity load forecasting: An analysis of Macedonian case,” Energy, vol. 115, pp. 1688–1700.

[2] A. Hussain, M. Rahman, and J. A. Memon. 2016. “Forecasting electricity consumption in Pakistan: The way forward,” Energy Policy, vol. 90, pp. 73–80, doi: 10.1016/j.enpol.2015.11.028.

[3] J. Steinbuks, J. de Wit, A. Kochnakyan, and V. Foster. 2017. ” Forecasting Electricity Demand: An Aid for Practitioners”, Available at: [https://openknowledge.worldbank.org/handle/10986/26189], Accessed: Nov. 13, 2022.

[4] M. Uz Zaman, A. Islam, and N. Sultana. 2022. “Short term load forecasting based on Internet of Things (IoT),” 2018, Available at: [http://dspace.bracu.ac.bd/xmlui/handle/10361/10170] Accessed: Nov. 13, 2022.

[5] C. Kumar and M. Veerakumari. 2012. “Load Forecasting of Andhra Pradesh Grid using PSO, DE Algorithms,” Int J Adv Res Comput Eng Technol., pp 179–184.

[6] J. Zhang, Y. M. Wei, D. Li, Z. Tan, and J. Zhou. 2018. “Short term electricity load forecasting using a hybrid model,” Energy, vol. 158, pp. 774–781, doi: 10.1016/J.ENERGY.2018.06.012.

[7] E. A. Madrid and N. Antonio. 2021. “Short-Term Electricity Load Forecasting with Machine Learning,” Information 2021, vol. 12(2), p. 50, doi: 10.3390/INFO12020050.

[8] G. A. N. Mbamalu and M. E. El-Hawary. 1993. “Load Forecasting Via Suboptimal Seasonal Autoregressive Models and Iteratively Reweighted Least Squares Estimation,” IEEE Transactions on Power Systems, vol. 8(1), pp. 343–348, doi: 10.1109/59.221222.

[9] S. R. Huang. 1997. “Short-term load forecasting using threshold autoregressive models,” IEE Proceedings: Generation, Transmission and Distribution, vol. 144(5), pp. 477–48, doi: 10.1049/IP-GTD:19971144.

[10] J. F. Chen, W. M. Wang, and C. M. Huang. 1995. “Analysis of an adaptive time-series autoregressive moving-average (ARMA) model for short-term load forecasting,” Electric Power Systems Research, vol. 34(3), pp. 187–196, doi: 10.1016/0378-7796(95)00977-1.

[11] A. Tarsitano and I. L. Amerise. 2017 “Short-term load forecasting using a two-stage sarimax model,” Energy, vol. 133, pp. 108–114, doi: 10.1016/J.ENERGY.2017.05.126.

[12] N. Shirzadi, A. Nizami, M. Khazen, and M. Nik-Bakht. 2021. “Medium-Term Regional Electricity Load Forecasting through Machine Learning and Deep Learning,” Designs, vol. 5(2), p. 27, doi: 10.3390/DESIGNS5020027.

[13] Q. Li, P. Ren, and Q. Meng, “Prediction model of annual energy consumption of residential buildings,” 2010 International Conference on Advances in Energy Engineering, ICAEE 2010, pp. 223–226, 2010, doi: 10.1109/ICAEE.2010.5557576.

[14] R. K. Jain, K. M. Smith, P. J. Culligan, and J. E. Taylor. 2014. “Forecasting energy consumption of multi-family residential buildings using support vector regression: Investigating the impact of temporal and spatial monitoring granularity on performance accuracy,” Appl Energy, vol. 123, pp. 168–178, doi: 10.1016/J.APENERGY.2014.02.057.

[15] M. del Carmen Ruiz-Abell n, A. Gabaldón, and A. Guillamón. 2018. “Load Forecasting for a Campus University Using Ensemble Methods Based on Regression Trees,” Energies, vol. 11(8), p. 2038, doi: 10.3390/EN11082038.

[16] S. S. A. Umri, M. S. Firdaus, and A. Primajaya. 2021. “Analisis Dan Komparasi Algoritma Klasifikasi Dalam Indeks Pencemaran Udara Di Dki Jakarta,” Jurnal Informatika dan Komputer, vol. 4(2), doi: 10.33387/jiko.

[17] S. Bouktif, A. Fiaz, A. Ouni, and M. A. Serhani. 2018. "Optimal Deep Learning LSTM Model for Electric Load Forecasting using Feature Selection and Genetic Algorithm: Comparison with Machine Learning Approaches," Energies vol. 11(7), p. 1636, doi: 10.3390/EN11071636.

[18] Nilakanta Singh, K., Robindro Singh, K. 2021. A Review on Deep Learning Models for Short-Term Load Forecasting. In: Choudhary, A., Agrawal, A.P., Logeswaran, R., Unhelkar, B. (eds) Applications of Artificial Intelligence and Machine Learning. Lecture Notes in Electrical Engineering, vol.778. Springer, Singapore. https://doi.org/10.1007/978-981-16-3067-5_53

[19] M. N. Fekri, H. Patel, K. Grolinger, and V. Sharma. 2021. "Deep learning for load forecasting with smart meter data: Online Adaptive Recurrent Neural Network," Appl Energy, vol. 282, p. 116177, doi: 10.1016/J.APENERGY.2020.116177.

[20] Y. Tian, L. Sehovac, and K. Grolinger. 2019. "Similarity-Based Chained Transfer Learning for Energy Forecasting with Big Data," IEEE Access, vol. 7, pp. 139895–139908, doi: 10.1109/ACCESS.2019.2943752.

[21] L. Sehovac and K. Grolinger. 2020. "Deep Learning for Load Forecasting: Sequence to Sequence Recurrent Neural Networks with Attention," IEEE Access, vol.8, pp. 36411–36426, doi: 10.1109/ACCESS.2020.2975738.

[22] C. Wang, Y. Wang, Z. Ding, T. Zheng, J. Hu, and K. Zhang. 2022. "A Transformer-Based Method of Multienergy Load Forecasting in Integrated Energy System," IEEE Trans Smart Grid, vol. 13(4), pp. 2703–2714, doi: 10.1109/TSG.2022.3166600.

[23] A. L'heureux, K. Grolinger, and M. A. M. Capretz. 2022. "Transformer-Based Model for Electrical Load Forecasting," Energie, vol. 15(14), p. 4993, doi: 10.3390/EN15144993.

[24] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser and Illia Polosukhin. 2017. "Attention Is All You Need," Adv Neural Inf Process Syst, vol. 2017-December, pp. 5999–6009, doi: 10.48550/arxiv.1706.03762.