

Electricity Demand Prediction by a Transformer-Based Model

A. M. Mahmood¹, M. M. Abdul Zahra², W. Hamed³, B. S. Bashar⁴, A. H. Abdulaal⁵, T. Alawsi⁶,
A. H. Adhab⁷

1- Department of Optical Techniques, AlNoor University College, Iraq.

Email: ahmed.m53@alnoor.edu.iq (Corresponding author)

2- Computer Techniques Engineering Department, Al-Mustaqbal University College, Hillah 51001, Iraq.

Email: musaddaqaahir@mustaqbal-college.edu.iq

3- Medical technical college, Al-Farahidi University, Baghdad, Iraq.

Email: hamed.h246@gmail.com

4- Al-Nisour University College, Baghdad, Iraq.

Email: bashar.s.eng@nuc.edu.iq

5- Medical Device Engineering, Ashur University College, Baghdad, Iraq

Email: alaa.Hussein@au.edu.iq

6- Scientific Research Center, Al-Ayen University, Thi-Qar, Iraq.

Email: taif.alawsi@alayen.edu.iq

7- Department of Medical Laboratory Technics, Al-Zahrawi University College, Karbala, Iraq.

Email: alihussien@g.alzahu.edu.iq

Received: 24 September 2022

Revised: 30 September 2022

Accepted: 10 October 2022

ABSTRACT:

The frighteningly high levels of power consumption at present are caused mainly by the expanding global population and the accessibility of energy-hungry smart technologies. So far, various simulation tools, engineering- and AI-based methodologies have been utilized to anticipate power consumption effectively. While engineering approaches forecast using dynamic equations, AI-based methods forecast using historical data. The modeling of nonlinear electrical demand patterns is still lacking for durable solutions, however, the available approaches are only effective for resolving transient dependencies. Furthermore, because they are only based on historical data, the current methodologies are static in nature. In this research, we present a system based on deep learning to anticipate power consumption while accounting for long-term historical relationships. In our approach, a transformer-based model is used for the prediction of electricity demand on data collected from the regional facilities in Iraq. According to the conducted experiments, our approach claims competitive performance, achieving an error rate of 2.0 in predicting 1-day-ahead of electricity demand in the test samples.

KEYWORDS: Electricity Demand, Machine Learning, Self-Attention, Power Consumption

1. INTRODUCTION

The demand for and consumption of energy and materials have increased significantly in recent decades due to population expansion, social development, and technological breakthroughs [1]. The overall annual energy consumption of Iraq has significantly increased from 179 GWh in 2010 to 341 GWh in 2015 and to 678 GWh in 2019 [2]. The demand for electricity has been shown to rise by 7% annually on average. With relation to this rising demand, enough energy is needed to meet the needs of the entire country while taking care of mother nature. To continuously improve their services, the utility providers must support improved programs

and keep a record of energy consumption [3].

Numerous statistical and conventional techniques have been employed to evaluate trends, define patterns, and predict the energy demand [4]. These techniques can be broadly divided into two categories: machine learning techniques and traditional techniques. Traditional ways to predicting energy usage comprise random time series and linear extrapolation strategies. By extending the time series patterns into the future, unpredictable time series function [5].

These approaches, which have been widely utilized in earlier publications, can solve linear problems more efficiently. Different learning methods, including

decision trees, Bayesian models, ensemble techniques, and neural networks, have gained popularity as a result of the development of artificial intelligence (AI) [6]. These AI-based techniques, which draw their main inspiration from the human brain, are particularly trustworthy since they can detect non-linearity in the data. Learning, validation, and testing steps are all included in artificial neural network (ANN) based learning models. An ANN is taught to create a mapping between input and output variables throughout the learning and validation stages [7].

The capacity of energy cloud services to forecast electricity demand across a broad variety of time spans is a key feature. Agents who create electricity buying and selling activities, energy suppliers, and electricity network administrators who forecast future demand all greatly benefit from it [8]. Brief power demand forecasting, intermediate electricity load prediction, and long-term electricity consumption forecasting are the three general categories into which demand forecasting can be divided [9].

Numerous studies have been conducted recently to determine the best methods for anticipating power consumption [10]. The researchers have utilized a variety of machine learning techniques for estimating energy consumption, including Support Vector Machines (SVM), Decision Trees, ANN, and Recurrent Neural Networks (RNN). Neural networks were discovered to be the most effective algorithms, offering a reasonable level of accuracy [11].

The electricity industry heavily relies on distributed energy sources like wind and solar power. Many areas have implemented deregulations of the electric energy business, which allow market players to raise competition and capture electricity prices, to promote clean energy, improve congestion management, and optimize resource allocation of the power market [12]. Day-ahead power price forecasting is a primary priority for effective bidding tactics and risk-control in accordance with market regulations. The market player can offer a suitable bidding strategy to maximize its payout if it has a favorable day in the future. However, it exhibits typical non-linearity and significant volatility as a result of the effect of rivals' bidding behavior and power system operating circumstances [13]. In response to these difficulties, there are three kinds of day-ahead electricity price prediction approaches that have been written about in the literature: physical methods, statistical methods, and machine learning methods.

Shallow learning models and deep learning models are the two main categories of machine learning techniques. The performance of shallow learning models, which are based on the idea of error reduction, is often better than that of physical and statistical approaches [14]. They have been one of the most popular techniques for predicting power costs because

of their significant qualities in extracting characteristics. Support vector regression, and ANNs are examples of shallow learning methods. In [15], the authors used a hybrid model based on feature selection and support vector machine-based approaches to forecast increases in the price of power. In [16], the authors used an artificial neural network along with data from the corresponding forecasting day's daily load and pricing to anticipate PJM market energy prices for the day ahead. According to the simulation findings, the suggested method's average mean absolute percentage error (MAPE) can reach 9.75%. To anticipate day-ahead power prices with tolerable performance, the authors in [17] created a hybrid predicting method based on a seasonal component auto-regressive approach and an ANN approach.

An adequate basis for managing and planning the energy supply and developing policy can be found in the forecasting of electricity consumption and the production of renewable energy [18]. In order to create a precise energy plan, several studies have created time-series statistical energy forecasting models for stable energy infrastructures. For instance, Thatcher [19] utilized a multiple linear regression (MLR) model with local history electricity needs and meteorological data to forecast a future local electricity demand curve for Australia. Using Italy's projected population and annual economic output, Bianco et al. [20] proposed a multiple regressive model to forecast annual energy consumption up to 2040.

In this paper, we propose a transformer-based model for predicting electricity demand. Our method utilizes multi-head self-attention to handle sequential data. The proposed approach is extensively evaluated on a newly collected dataset containing over 5000 samples gathered from electricity facilities located in Baghdad, Iraq. To clarify, we itemize the main contributions of the current study as follows:

1. We propose a novel transformer-based model for predicting electricity demand.
2. A new dataset is collected for training and evaluation purposes.
3. Our extensive evaluation and comparison with other state-of-the-art models, whose architectures are based on recurrent neural networks, show that the proposed approach is effective and reliable for being used in real energy facilities

The rest of this paper is outlined as follows. In section 2 we explain our methodology and its components. Section 3 details our experimental results. Section 4 contains conclusion.

2. METHODOLOGY

This section includes our proposed methodology. Firstly, we explain the sequence-to-sequence models,

followed by transformers' architecture explanation. Also, the preprocessing steps and our case study will be elaborated.

2.1. Sequence to Sequence Learning

When translating a sequence of words from one language into a sequence of distinct words in another, sequence to sequence models perform exceptionally well. Long-Short-Term-Memory (LSTM)-based models are a common option for this kind of model [21]. The LSTM modules may give meaning to the sequence with sequence-dependent data while remembering (or forgetting) the components it deems significant (or unimportant). For instance, sentences depend on word order because comprehension of the sentence depends on word order. The best option for this kind of data is LSTM [22].

The key to LSTMs' success is their assertion that they were among the first implementations to solve the technical issues and fulfill the promise of recurrent neural networks [23]. Bidirectional recurrent neural networks work on the principle of feeding each training sequence into two different recurrent nets that are both linked to the same output layer [24]. This implies that the it has comprehensive, sequential knowledge of all points before and after each point in a particular sequence. Additionally, there is no need to identify a (task-dependent) time-window or goal delay size because the net is free to use as much or as little of this context as is required [25]. As a result, learning ceases for layers in recurrent neural networks that get a modest gradient update. Typically, those are the older layers. RNNs can therefore forget what they have seen in longer sequences, resulting in a short-term memory, because these layers do not learn. You may read my earlier post here to learn more about the general workings of recurrent neural networks [26].

A better RNN is a long short-term memory neural network. Since RNN may result in issues with potential to eliminate and explosion, LSTM is suggested to analyze time series data with a large time period. The memory capacity of the LSTM RNN's hidden layer, which is made up of memory/forget gates, input gates, and output gates replace, is such that it can store lengthy historical time series. The network weights of the CNN, LSTM RNN, and DBN, the weights, and the biases, are trained and updated according on the BP algorithm utilizing gradient descent approaches with such a mini-batch form. This improves the accuracy and stability of the day-ahead electricity demand prediction [27].

The RNN is identical to the n-gram model (shown in Fig. 1), with the exception that the results of all past computations will affect the results of the current input. The internal state of the RNN serves as a memory. Speech processing and language generation processing both benefit greatly from it. This graphic demonstrates

how the input at a given time ($t+6$) depends on both the input at that time and the hidden state of each preceding step.

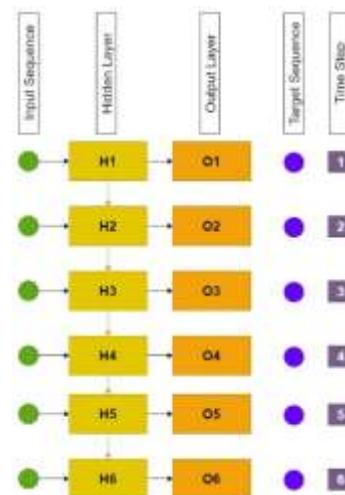


Fig. 1. The overall architecture of an RNN-based model

As we will merely pass the words individually on our model, it allows the network to remember a history of previously learnt parameters and use it to forecast the next output, which gets around the issue of word order and eliminates the need for calculation. Although RNN-based models have advantages, they have the following conundrums: 1) Vanishing or exploding gradients 2) Since the results are dependent on earlier calculations, we are unable to parallelize the computations [28].

2.2. Transformers

Transformers were introduced in [29]. These models utilize the mechanism called self-attention and have helped to improve the overall performance of neural networks in a variety of tasks such as machine translation, speech recognition, language models, and image captioning [30]. Through weight factorization, weight quantization, weight pruning, and knowledge distillation, existing efforts aim to increase memory efficiency in transformers. With these approaches, training or inference times are sped up by lowering the memory or computing requirements, but scaling to lengthy sequences is still difficult since the time complexity is still quadratic with respect to the length of the sequence. On the other hand, we demonstrate, both conceptually and practically, that our strategy minimizes the memory and temporal complexity of transformers [31].

Based on Fig. 2, the first step in determining self-attention is to divide each input vector of the encoder into three separate vectors (in this case, the embedding of each word). As a result, we generate a Query vector, Key vector, and Value vector for each word. By

multiplying the embedding by three trained matrices throughout the training phase, these vectors are produced. You will see that the dimensions of these new vectors are fewer than those of the embedding vector [32].

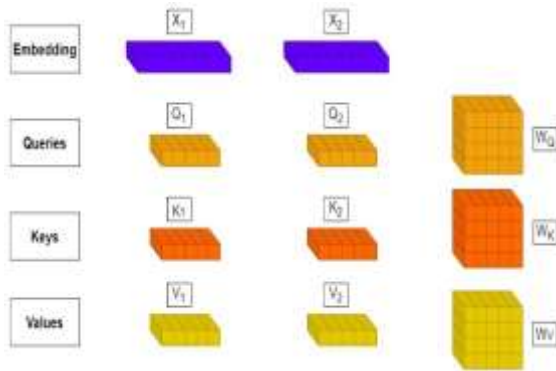


Fig. 2. The inputs of transformer module.

2.3. Preprocessing

The dataset gathered in the actual world is prone to a variety of errors, such as incomplete data, distortion, missing information, and raw format. These inconsistencies or inaccuracies in the raw data might result in subpar data analysis [33]. Data must thus be pre-processed in order to guarantee the validity of the process of discovering knowledge from the data. Typically, the data preparation step contains the following sub-phases: Cleansing Data Filling in missing numbers, removing noise, identifying outliers, and resolving inconsistencies in the data are all parts of data cleaning. Data Transformation: This sub-phase entails a variety of techniques, such combining many files into a single format that may be used, scaling the attribute to adhere to particular characteristics, etc. Data reduction seeks to retain the majority of the data's qualities while eliminating redundancies. It seeks to provide a condensed representation of the data, either by sampling or by limiting the number of characteristics. Data discretization: To make data acceptable for analysis, discretization entails using a variety of techniques, such as binning (to limit the number of values of a variable by splitting the range of attribute values into intervals), concept hierarchies, and more.

2.4. Case study

The energy requirement dataset for Baghdad each day (timestamped) is used in the current study. It is an Iraqi city that has been carefully designed. The city occupies a about 123 x 11 m-squared area of land. Baghdad's demand for energy is rising quickly, at a rate of 3.41 units per year. We have data on the electrical consumption that was sampled every 15 minutes at regular intervals beginning in January 2015. Table 2 provides statistical information on numerous factors

(including average, maximum, lowest, and peak electricity demand for the chosen day in every season).

3. RESULTS AND DISCUSSION

3.1. Evaluation Metrics

The effectiveness of the regression models can be evaluated using a variety of performance metrics. The following assessment metrics were employed in this study:

1. Correlation Coefficient (CC): It is a metric that gauges how strongly real-time measurements of a characteristic correlate with predictions.

$$\vartheta = \frac{\text{Covariance}(y, \bar{y})}{\sigma_y \cdot \sigma_{\bar{y}}} \quad (1)$$

2. Root Mean Squared Error (RMSE): It is a metric that gauges how strongly real-time measurements of a characteristic correlate with predictions.

$$RMSE = \sqrt{\frac{1}{2n} \sum_{i=1}^2 \sum_{j=1}^n (y_{j,i} - \bar{y}_{j,i})^2} \quad (2)$$

3.2. Performance Results

The performance of the proposed model is evaluated using the metrics introduced in section 3.1. Table 1 demonstrates the minimum and maximum prediction for power demand achieved by the model.

Table 1. The samples of times roman type sizes and styles used for formatting a technical work.

Season	Day	Min demand	Max demand
Spring	Monday	89.62	121.25
	Tuesday	88.72	132.63
	Wednesday	89.15	129.62
	Thursday	90.12	130.15
	Friday	88.45	121.62
	Saturday	90.01	132.21
	Sunday	89.72	128.66
Summer	Monday	88.61	122.32
	Tuesday	89.32	132.63
	Wednesday	90.15	128.64
	Thursday	90.12	130.15
	Friday	89.45	122.32
	Saturday	91.01	131.30
	Sunday	90.71	130.63
Autumn	Monday	90.65	124.25
	Tuesday	87.34	134.63
	Wednesday	89.67	128.62
	Thursday	91.53	129.15

	Friday	90.54	124.62
	Saturday	89.40	132.21
	Sunday	90.90	128.66
Winter	Monday	89.54	124.25
	Tuesday	88.64	126.63
	Wednesday	90.32	130.62
	Thursday	90.65	129.15
	Friday	89.54	132.34
	Saturday	88.32	131.40
	Sunday	91.91	131.65

Regression models are fed all learning algorithm during the training phase to quickly create an input to output routing. While fine-tuning hyper-parameters, the validation data is used to offer an unbiased assessment of a model fit on training data. The proposed regression model's effectiveness is assessed on the testing set after training and validation evaluations.

Moreover, we present Fig. 3 to illustrate the RMSE value and time the regressor model takes to predict a value.

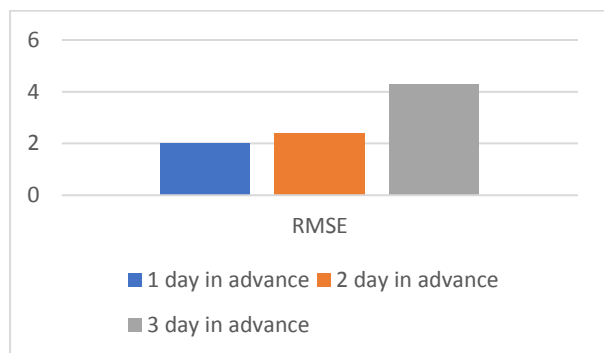


Fig. 3. The comparison of RMSE value with respect to the time (in advance) the model can predict.

Based on Fig. 3 we can deduce that almost uniformity in each option can be seen in the overall outcomes of several predicting methods. Given that market players could have varied risk attitudes, probabilistic predicting performance with different confidence level are examined using multiple predicting techniques during each season to show efficacy and feasibility.

4. CONCLUSION

This research presents a unique transformer-based framework for predicting day-ahead power demand that consists of an integrated development environment module, a deep learning-based point prediction module, an error compensation module, and a probabilistic prediction module. Compared to other disciplines, the reliable and precise projection of power consumption has received little attention. However, during the last few years, there have been considerable improvements

made in the creation of accurate and effective electrical prediction models. Due to their success in solving non-linear issues, artificial intelligence-based solutions have performed well and received considerable attention. Our research reveals the effectiveness of the suggested methodology and its superior performance when compared to other cutting-edge methods.

REFERENCES

- [1] Saoud, Lyes Saad, Hasan Al-Marzouqi, and Ramy Hussein. "Household Energy Consumption Prediction Using the Stationary Wavelet Transform and Transformers." *IEEE Access*, Vol. 10, pp. 5171-5183, 2022.
- [2] Sun, Lisha, and David Lubkeman. "Agent-based modeling of feeder-level electric vehicle diffusion for distribution planning." *IEEE Transactions on Smart Grid*, Vol. 12, No. 1, pp. 751-760, 2020.
- [3] L'Heureux, Alexandra, Katarina Grolinger, and Miriam AM Capretz. "Transformer-Based Model for Electrical Load Forecasting." *Energies*, Vol. 15, No. 14, p. 4993, 2022.
- [4] Al-Bahrani, Loau Tawfak, Mehdi Seyedmahmoudian, Ben Horan, and Alex Stojcevski. "Distribution transformer load behavior, burden, and characteristics of residential consumers: A case study of Baghdad City." *Energy and buildings*, Vol. 210, p. 109693, 2020.
- [5] Dong, Ming. "A data-driven long-term dynamic rating estimating method for power transformers." *IEEE Transactions on Power Delivery*, Vol. 36, No. 2, pp. 686-697, 2020.
- [6] Nourani, Vahid, Zahra Razzaghzadeh, Aida Hosseini Baghanam, and Amir Molajou. "ANN-based statistical downscaling of climatic parameters using decision tree predictor screening method." *Theoretical and Applied Climatology*, Vol. 137, No. 3, pp. 1729-1746, 2019.
- [7] Johannessen, Nils Jakob, Mohan Kolhe, and Morten Goodwin. "Relative evaluation of regression tools for urban area electrical energy demand forecasting." *Journal of cleaner production*, Vol. 218, pp. 555-564, 2019.
- [8] Hong, Shin-Ki, Sung Gu Lee, and Myungchin Kim. "Assessment and mitigation of electric vehicle charging demand impact to transformer aging for an apartment complex." *Energies*, Vol. 13, No. 10, p. 2571, 2020.
- [9] Cao, Jianwen, Bizhong Xia, and Jie Zhou. "An Active Equalization Method for Lithium-ion Batteries Based on Flyback Transformer and Variable Step Size Generalized Predictive Control." *Energies*, Vol. 14, No. 1, p. 207, 2021.
- [10] Gjelij, Marjan, Nataly Bañol Arias, Chresten Traeholt, and Seyedmostafa Hashemi. "Multifunctional applications of batteries within fast-charging stations based on EV demand-prediction of the users' behaviour." *The Journal of Engineering* Vol. 2019, No. 18, pp. 4869-4873, 2019.
- [11] Gao, Yuan, and Yingjun Ruan. "Interpretable deep learning model for building energy consumption

- prediction based on attention mechanism." *Energy and Buildings*, Vol. 252, p. 111379, 2021.
- [12] Pradhan, Pravakar, Iftekhar Ahmad, Daryoush Habibi, Ganesh Kothapalli, and Mohammad AS Masoum. "Reducing the impacts of electric vehicle charging on power distribution transformers." *IEEE Access*, Vol. 8, pp. 210183-210193, 2020.
- [13] Pouladi, Parsa, Abbas Afshar, Mohammad Hadi Afshar, Amir Molajou, and Hamid Farahmand. "Agent-based socio-hydrological modeling for restoration of Urmia Lake: Application of theory of planned behavior." *Journal of hydrology*, Vol. 576, pp. 736-748, 2019.
- [14] Van den Berg, M. A., I. Lampropoulos, and T. A. AlSkaif. "Impact of electric vehicles charging demand on distribution transformers in an office area and determination of flexibility potential." *Sustainable Energy, Grids and Networks*, Vol. 26, pp. 100452, 2021.
- [15] Tang, Binh, and David S. Matteson. "Probabilistic transformer for time series analysis." *Advances in Neural Information Processing Systems*, Vol. 34, pp. 23592-23608, 2021.
- [16] Lahooti Eshkevari, Alireza, Ali Mosallanejad, and Mohammadsadegh Sepasian. "In-depth study of the application of solid-state transformer in design of high-power electric vehicle charging stations." *IET Electrical Systems in Transportation*, Vol. 10, No. 3, pp. 310-319, 2020.
- [17] Pradhan, Pravakar, Iftekhar Ahmad, Daryoush Habibi, Ganesh Kothapalli, and Mohammad AS Masoum. "Reducing the impacts of electric vehicle charging on power distribution transformers." *IEEE Access*, Vol. 8, pp. 210183-210193, 2020.
- [18] Tziouvani, Lysandros, Lenos Hadjidemetriou, Charalampos Charalampous, Maria Tziakouri, Stelios Timotheou, and Elias Kyriakides. "Energy management and control of a flywheel storage system for peak shaving applications." *IEEE Transactions on Smart Grid*, Vol. 12, No. 5, pp. 4195-4207, 2021.
- [19] Ha, Sooji, Daniel J. Marchetto, Sameer Dharur, and Omar I. Asensio. "Topic classification of electric vehicle consumer experiences with transformer-based deep learning." *Patterns*, Vol. 2, No. 2, p. 100195, 2021.
- [20] Wu, Binrong, Lin Wang, and Yu-Rong Zeng. "Interpretable wind speed prediction with multivariate time series and temporal fusion transformers." *Energy*, Vol. 252, p. 123990, 2022.
- [21] Rocha, Oscar David Ariza, Kateryna Morozovska, Tor Laneryd, Ola Ivarsson, Claes Ahlrot, and Patrik Hilber. "Dynamic rating assists cost-effective expansion of wind farms by utilizing the hidden capacity of transformers." *International Journal of Electrical Power & Energy Systems*, Vol. 123, p. 106188, 2020.
- [22] Zhou, Feng, Peng Zhao, Min Lei, Changxi Yue, Jicheng Yu, and Siyuan Liang. "Capacitive voltage transformer measurement error prediction by improved long short-term memory neural network." *Energy Report*, Vol. 8, pp. 1011-1021, 2022.
- [23] Baroni, Bruno R., Wadaed Uturbey, André MG Costa, and Samuel P. Rocha. "Impact of photovoltaic generation on the allowed revenue of the utilities considering the lifespan of transformers: A Brazilian case study." *Electric Power Systems Research*, Vol. 192, p. 106906, 2021.
- [24] Brinkel, N. B. G., W. L. Schram, T. A. AlSkaif, I. Lampropoulos, and W. G. J. H. M. Van Sark. "Should we reinforce the grid? Cost and emission optimization of electric vehicle charging under different transformer limits." *Applied Energy*, Vol. 276, p. 115285, 2020.
- [25] Choi, Eunjeong, Soohwan Cho, and Dong Keun Kim. "Power demand forecasting using long short-term memory (LSTM) deep-learning model for monitoring energy sustainability." *Sustainability*, Vol. 12, No. 3, p. 1109, 2020.
- [26] Zweistra, Marisca, Stan Janssen, and Frank Geerts. "Large scale smart charging of electric vehicles in practice." *Energies*, Vol. 13, No. 2, p. 298, 2020.
- [27] Wörner, Ralf, Inna Morozova, Danting Cao, Daniela Schneider, Martin Neuburger, Daniel Mayer, Christian Körner et al. "Analysis and prediction of electromobility and energy supply by the example of Stuttgart." *World Electric Vehicle Journal*, Vol. 12, No. 2, p. 78, 2021.
- Yan, Chun, Meixuan Li, and Wei Liu. "Transformer fault diagnosis based on BP-Adaboost and PNN series connection." *Mathematical Problems in Engineering*, Vol. 10, No. 3, pp. 310-319, 2020.
- [28] Jiang, Weiwei. "Deep learning based short-term load forecasting incorporating calendar and weather information." *Internet Technology Letters*, Vol. 5, No. 4, p. 383, 2022.
- [29] Gómez, Andrea Molina, Kateryna Morozovska, Tor Laneryd, and Patrik Hilber. "Optimal sizing of the wind farm and wind farm transformer using MILP and dynamic transformer rating." *International Journal of Electrical Power & Energy Systems*, Vol. 136, p. 107645, 2022.
- [30] Rodriguez-Rivero, Jacob, Javier Ramirez, Francisco Jesús Martínez-Murcia, Fermín Segovia, Andrés Ortiz, D. Salas, Diego Castillo-Barnes et al. "Granger causality-based information fusion applied to electrical measurements from power transformers." *Information Fusion*, Vol. 57, pp. 59-70, 2020.
- [31] Rutskov, Alexey L., Viktor L. Burkovsky, and Evgeny V. Sidorenko. "Optimization of electric power systems using fuzzy neural network algorithms." *Journal of mechanics of continua and mathematical sciences*, Vol. 1, p. 264, 2020.
- [32] Haupt, Sue Ellen, Mayte Garcia Casado, Michael Davidson, Jan Dobschinski, Pengwei Du, Matthias Lange, Timothy Miller et al. "The use of probabilistic forecasts: Applying them in theory and practice." *IEEE Power and Energy Magazine*, Vol. 17, No. 6, pp. 46-57, 2019.