Information Sciences with Applications, Vol. 3, 2024
https://doi.org/10.61356/j.iswa.2024.3314

Hybrid Attention-Enhanced Deep Learning for Accurate Hourly Energy Consumption Forecasting

Walid Abdullah [1,*], Ahmed Elmasry [1] and Ahmed Tolba [1]

[1] Department of Computer Science, Faculty of Computers and Informatics, Zagazig University, Zagazig 44519, Egypt; Emails: waleed@zu.edu.eg; a.elmasry24@fci.zu.edu.eg; a.tolba24@fci.zu.edu.eg..

* Correspondence: waleed@zu.edu.eg.

Abstract: Accurate forecasting of hourly energy consumption is essential for optimizing energy distribution, ensuring grid stability, and informing policy decisions. In this study, we propose a novel hybrid deep learning model that integrates attention mechanisms with long short-term memory for forecasting hourly electricity consumption. The model is trained and tested using a PJME_MW dataset, spanning from December 31, 2002, to January 2, 2018. The model was evaluated using a set of evaluation metrics R-squared score, Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Deviation (MAD), and Mean Absolute Error (MAE). The models are compared with established deep learning architectures such as ResNet, TCN, LSTM, RNN, and Attention CNN. The Comparative analysis demonstrates superior forecasting performance. The results showed that the proposed model outperformed all other models, it achieved the best accuracy with RMS, MAE, and R2 Score of 0.012, 0.007, and 0.992 respectively, which validates the effectiveness of our approach in enhancing prediction accuracy for energy consumption. The source code is publicly accessible at https://github.com/Hourly-Energy-Consumption.

Keywords: Deep Learning, LSTM, Attention, Energy Consumption, Usage Forecasting.

1. Introduction

Energy consumption forecasting is crucial for utilities, and grid operators, and efficiently allocates resources, manages demand-supply imbalances, and plans infrastructure upgrades [1]. Traditional methods like statistical models have limitations such as Limited Data Handling and Inability to Account for Seasonality and External Factors in addition to limitations in capturing the complex temporal patterns inherent in energy consumption data [2]. With the rise of artificial intelligence (AI) and deep learning, particularly Long Short-Term Memory (LSTM) networks, significant advancements have been made in time series forecasting [3], these techniques offer a powerful alternative for overcoming these limitations and achieving enhanced time series forecasting accuracy by enabling models to learn intricate dependencies and trends from sequential data.

Deep learning models, particularly recurrent neural networks (RNNs) and their variants like LSTM networks, possess a remarkable ability to learn intricate patterns from large datasets. This allows them to capture the non-linear relationships between various factors and energy consumption. these models are designed to process sequential data like hourly energy consumption measurements through effectively capturing the temporal dependencies within the data, and understanding how past consumption values influence future values [4, 5].

In recent years, attention mechanisms have further enhanced the capabilities of deep learning models by enabling them to selectively focus on relevant parts of the input data [6]. Originally popularized in natural language processing tasks, attention mechanisms have since been successfully applied to various domains, including computer vision and time series forecasting. In the context of energy consumption forecasting, attention mechanisms allow deep learning models such as LSTM to focus on specific parts of the input data by dynamically weighing the importance of different time steps that are most relevant for predicting the future value. This targeted focus leads to more accurate and informative predictions [7, 8].

In this work, we propose a novel hybrid deep learning model that integrates attention mechanisms with a recurrent deep neural network for forecasting hourly energy (specifically electricity) consumption. The model architecture combines the strengths of LSTM networks in capturing temporal dependencies with attention mechanisms' ability to highlight significant temporal features. By leveraging this hybrid approach, our model aims to enhance the accuracy and reliability of hourly energy consumption forecasts compared to traditional and other existing deep learning architectures. The model is trained and tested using a PJME_MW dataset and evaluated using the asset of evaluation metrics including R-squared score, Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Deviation (MAD), and Mean Absolute Error (MAE). The experimental results demonstrated that the proposed model achieved the best accuracy compared to other models, which demonstrates how well our method performs in increasing the accuracy of energy consumption predictions.

The remainder of the paper is organized as follows: Section 2 discusses and reviews the most recent techniques and relevant research, in energy consumption forecasting. Section 3 covers the materials and methods utilized in this paper, including the DL models and dataset. Section 4 presents the experimental setup; Section 5 presents the experimental results and discussion. The paper's conclusion is presented in Section 6.

2. LiteratureReview

In this section, we review relevant literature on the utilization of various techniques, including traditional statistical models and advanced deep learning approaches, in the field of electricity consumption forecasting. By examining previous studies, we aim to identify existing gaps and highlight key findings, methodologies, and advancements that have been made. This review will provide a comprehensive overview of the state-of-the-art in energy consumption forecasting and set the context for our proposed hybrid model.

Traditional statistical models, such as autoregressive integrated moving averages (ARIMA) and exponential smoothing, have been widely used for time series forecasting. These models rely on historical data to predict future values by identifying patterns and correlations. In [9], ARIMA models were proposed to forecast future electricity consumption. the results demonstrated the models' efficiency and accuracy as well as their capacity to rival other methods for predicting power consumption based on the utilization of the Mean Absolute Percentage Error (MAPE) to evaluate the prediction's accuracy, but the prediction error was relatively high. It was 4.332%. Furthermore. The ARIMA models using different parameter sets are proposed by the authors in [10] to forecast power consumption. The three ARIMA models are examined for possible power consumption to provide the necessary degree of performance. The study used electricity consumption data in industries in Guangdong province in China, the results of the experiment demonstrate that the ARIMA (1,1,1) is a reliable predictor of power consumption with high precision and consistent predictions. Finaly ARIMA models, although effective for short-term forecasting, often struggle with capturing the nonlinearities and complexities inherent in electricity consumption data.

With the advent of machine learning, more sophisticated techniques began to emerge. Machine learning models such as Support Vector Machines (SVM), Random Forests, and K-nearest neighbors (KNN). and others have been employed to enhance forecasting accuracy by capturing more complex patterns than traditional models. This study [11] presents an SVR model with an immune algorithm (IA) to forecast the electric loads, its results are compared with other models such as the regression model, and ANN model, and the final findings indicate that indicate that the SVR model with IA (SVRIA) results in better forecasting performance than the other methods. SVR was also proposed with a fruit fly optimization algorithm for seasonal electricity consumption forecasting [12]. The results showed that the proposed hybrid technique has a high accuracy in electricity consumption forecasting applications. KNN and Random Forest also have been used for the same reason and their results were fairly acceptable [13, 14]. These machine learning methods, however, typically require extensive feature engineering and may not fully exploit the temporal dependencies present in the data.

In recent years, Deep learning has revolutionized time series forecasting by offering models that can automatically learn representations from raw data, thus eliminating the need for manual feature engineering. Among the most popular deep learning models are Recurrent Neural Networks (RNNs) and their variants, such as Long Short-Term Memory (LSTM) networks and Gated Recurrent Units (GRUs). These models are particularly adept at handling sequential data and capturing long-term dependencies. LSTM, GRU, and RNN models were tested in [15] to accurately determine the electrical load based on the current electrical loads of the electricity. company. The results showed that the GRU model performed the best, having the lowest error and the highest accuracy. However, other models also achieved very high accuracy in predicting future peaks of electricity consumption. Furthermore, Authors In [15], proposed a hybrid deep learning model that combines convolutional characteristics of neural networks with LSTM for household electrical energy consumption forecasting. This proposed convolutional LSTM (ConvLSTM) architecture performs better than alternative models with the lowest root mean square error, according to experimental results.

Recently, Attention mechanisms have been introduced to enhance the performance of neural networks by allowing them to focus selectively on important time steps in the input sequence. This selective focus helps in capturing intricate temporal dependencies that are crucial for accurate forecasting. In recent years, attention mechanisms have been successfully integrated with various deep learning architectures, including LSTMs and Convolutional Neural Networks (CNNs), to improve their predictive capabilities [7, 16-17]. The main contributions of this study are twofold. First, it introduces a novel hybrid architecture that combines LSTM and attention mechanisms for electricity consumption forecasting. Second, it provides a comprehensive performance comparison with other established deep learning models, such as ResNet, TCN, standard LSTM, RNN, and Attention CNN. The results indicate that the proposed model achieves superior accuracy and reliability, thereby advancing the state-of-the-art in this domain.

3. Material and Methodologies

In this section, the experimental procedures and methodology employed in this study to develop and evaluate the proposed hybrid attention-deep neural network model are detailed. The utilized dataset, preprocessing steps, and models' training procedures are also outlined.

3.1 Dataset

PJM Interconnection LLC (PJM) is a regional transmission organization (RTO) in the United States. It is part of the Eastern Interconnection grid operating an electric transmission system serving all or parts of Delaware, Illinois, Indiana, Kentucky, Maryland, Michigan, New Jersey, North Carolina, Ohio, Pennsylvania, Tennessee, Virginia, West Virginia, and the District of Columbia. PJME_MW data at an hourly frequency from December 31, 2002, to January 2, 2018 dataset collected from PJM's website and are in megawatts (MW) [18]. consumption data is presented in Figure 1 in addition, its distribution is shown in Table 1.

Table 1. PJME_MW dataset distribution.

Count	Mean	Std	Min	25%	50%	75%	Max
145366.0	32080.2	6464.0	14544.0	27573.0	31421.0	35650.0	62009.0

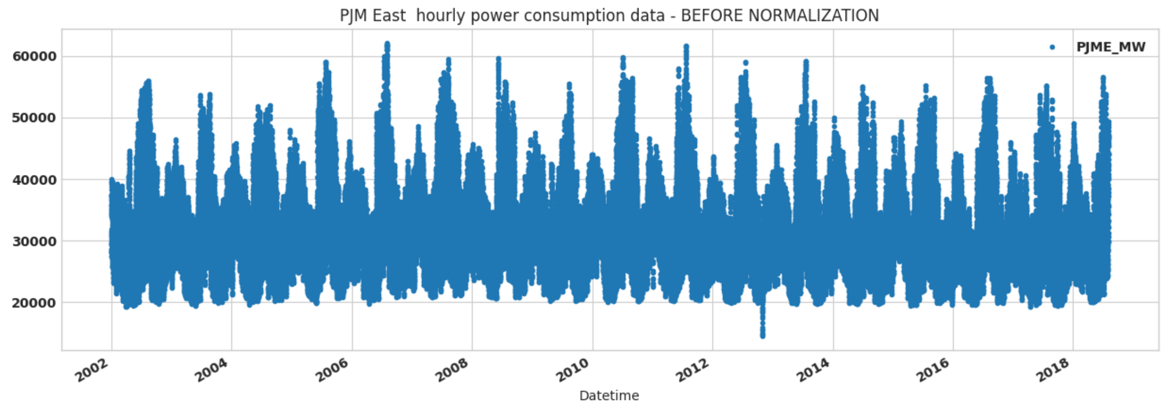

Figure 1. PJM hourly consumption data.

3.2 Data Preparation

Where initially, the process of data cleaning is executed to eliminate any discrepancies, null values, and extreme values that have an impact on the effectiveness of the model. Additionally, doing data normalization by Min Max Scalar is crucial to guarantee that all features are standardized to a comparable scale, hence avoiding the dominance of any specific feature during model training.

$$Xscaled = \frac{X - Xmin}{Xmax - Xmin} \quad (1)$$

Ultimately, the dataset is divided into two distinct sets: training, and testing. where using the split data 01-Jan-2016 where all data before mentioned date used in training data and all data after mentioned date used in testing data During training models, we used 0.1 from training data used in validation, this division allows the model to be trained on one piece of the data, evaluated on another section, and validated on a separate portion to examine its capacity to generalize.

3.3 Long-Short Term Memory

Long Short-Term Memory (LSTM) networks are a type of recurrent neural network (RNN) designed to model sequential data by capturing long-term dependencies and mitigating the vanishing gradient problem. LSTMs achieve this through a unique cell state structure and gating mechanisms, which regulate the flow of information. The three primary gates in an LSTM cell are:

- Input Gate: Controls the extent to which new information flows into the cell state.
- Forget Gate: Determines the amount of information from the previous cell state to be retained.
- Output Gate: Controls the output flow of cell information to the next hidden state.

These gates enable LSTMs to effectively learn and remember important features over long sequences, making them suitable for time series forecasting tasks, including electricity consumption prediction.

3.4 Attention Mechanisms

Attention mechanisms have been introduced to enhance the capability of neural networks by allowing them to focus on relevant parts of the input sequence selectively. In the context of time series forecasting, attention mechanisms dynamically weigh the importance of different time steps, enabling the model to prioritize significant periods that impact future predictions. This selective focus helps the model to capture intricate temporal patterns and relationships more effectively than traditional RNNs.

The attention mechanism operates by computing a set of weights for each time step in the input sequence, which is then used to generate a weighted sum of the inputs. This weighted sum highlights the most relevant information for making accurate predictions. By assigning higher weights to critical periods, the model can prioritize information that significantly impacts future electricity consumption, thus improving prediction accuracy.

3.5 Proposed Model

The core innovation of this study is the development of a hybrid model that combines the strengths of LSTM networks with attention mechanisms to forecast hourly electricity consumption more accurately. The proposed model architecture includes:

- Input Layer: Receives the hourly electricity consumption data as sequential inputs.
- LSTM Layer: Composed of 64 units, this layer captures the underlying long-term dependencies and temporal patterns in the electricity consumption data.
- Attention Layer: Consisting of 32 units, this layer dynamically assigns weights to different time steps, allowing the model to focus on significant periods that influence future consumption.
- Output Layer: A dense layer with a linear activation function is used to generate the final electricity consumption forecast.

By integrating attention mechanisms with the LSTM layer, the proposed hybrid model effectively captures both long-term dependencies and critical short-term variations in the electricity consumption data. This integration enhances the model's ability to provide precise and reliable forecasts, addressing the limitations of traditional LSTM and other deep learning architectures. The architecture of the proposed hybrid is presented in Figure 2.

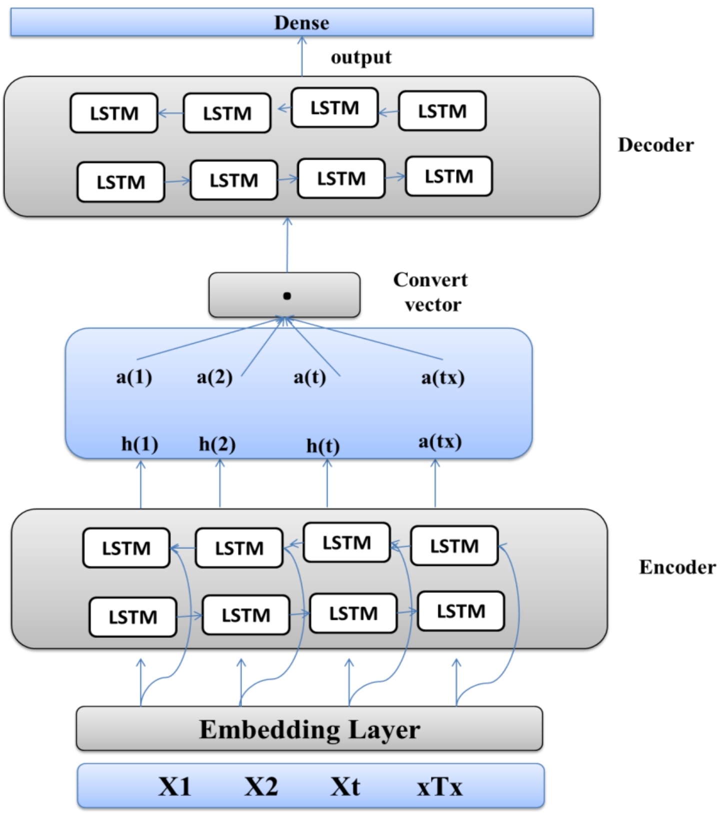

Figure 2. The architecture of the proposed hybrid attention-deep neural network model for hourly electricity consumption forecasting.

4. Experimental Setup

This section investigates the performance of the proposed model using a widely used dataset Hourly Energy Consumption. In addition, it is compared to different DL models, such as ResNet, TCN, LSTM, RNN, and the proposed model.

4.1 Experimental Environment Setup

All experiments in this paper were conducted on the Kaggle environment with Nvidia Tesla P100 GPU and 30 GB of RAM using Python Version 3.10.0 and Keres API Version 3.0. The Adam optimizer [19] with a learning rate of 0.001 was used to train the weights of implemented and proposed models for 100 epochs and using early stopping with patience 10 to prevent overfitting [20].

4.2 Evaluation Metrics

We have used 5 metrics to evaluate the models' effectiveness. Before using all these metrics, we must know about the residual error, i.e., $(\boldsymbol{y}-\hat{\boldsymbol{y}})$. Here, y and indicate the real values and $\hat{\boldsymbol{y}}$ indicate predicted values. The performance indicators used to evaluate the performance of those models are:

- **Mean squared error**

For calculating the MSE, take the real value, subtract the predicted value, and square that difference. Repeat that for all samples. Then, sum all of those squared values and divide by the number of samples.

$$\mathrm{MSE}=\frac{\Sigma(y_i-p_i)^2}{n} \tag{2}$$

- **Root mean square Error**

The Root Mean Squared Error (RMSE) is the most common performance indicator for regression models. It measures the average difference between predicted values and real values. The lower the RMSE, the better the model and its predictions.

$$\mathrm{RMSE}=\sqrt{\frac{\sum_{i=1}^{N}.(x_i-\hat{x}_i)^2}{N}} \tag{3}$$

Where $x_i$real values and $\hat{x}_i$predicted values, N number of samples, $\sum_{i=1}^{N}$. Summation from I to N

- **Mean absolute error**

The mean absolute error (MAE) is defined as the average variance between the real and predicted values

$$MAE=\frac{1}{n}\sum_{i=1}^{n}\ |y_1-\hat{y}_i|^2 \tag{4}$$

Where y_i predicted value, p_i the real value of I, n number of samples

- **R2 Score:**

The regression coefficient determines the best possible score is 1.0, and less than that is less efficient and gets worse

$$R^2=1-\frac{SS_{RES}}{SS_{TOT}}=1-\frac{\Sigma\mathrm{i}(y_i-\hat{y}_i)^2}{\Sigma\mathrm{i}(y_i-\overline{y})^2} \tag{5}$$

Where y_i real values and $\hat{y}_i$ predicted values.

- **Median absolute error**

The measure of variability in a univariate sample of quantitative data is effectively captured by the median absolute error (MedAE). The value of it might range from 0 to infinity. Hence, a decrease in the value corresponds to an increase in the accuracy of the model.

$$\mathrm{MedAE}(y, \hat{y}) = \mathrm{median}(|y_1 - \hat{y}_1|, \dots \dots \dots |y_n - \hat{y}_n|) \tag{6}$$

5. Results and Discussion

5.1 Experimental Results

The performance of the proposed hybrid attention-deep neural network model is compared against several established deep learning models, including ResNet, Temporal Convolutional Network (TCN), LSTM, Recurrent Neural Network (RNN), and Attention CNN in terms of a set of evaluation metrics. and the results showed that the proposed model achieved the best accuracy and outperformed all other models, which validates the effectiveness of our approach in enhancing prediction accuracy for energy consumption. The comparative results are summarized in Table 2. Figure 3 represents the MSE values of the proposed model across each epoch in the training. While

Table 2. Performance comparison of models for hourly electricity consumption forecasting.

Model	R2 Score	MSE	RMSE	MAD	MAE
ResNet	0.981	0.000	0.019	0.009	0.012
TCN	0.991	0.000	0.012	0.004	0.006
LSTM	0.991	0.000	0.013	0.005	0.007
RNN	0.971	0.001	0.023	0.014	0.017
Attention CNN	0.965	0.015	0.025	0.008	0.011
Attention DNN	0.992	0.000	0.012	0.004	0.007

To further understand the training process and performance of the proposed hybrid model, the Mean Squared Error (MSE) values across each epoch during the training phase are plotted in Figure 3. Further the accuracy of the proposed model's predictions is evaluated by comparing the predicted electricity consumption values against the actual (real) values. Figure 4 presents this comparison, highlighting the model's ability to closely follow the real consumption patterns.

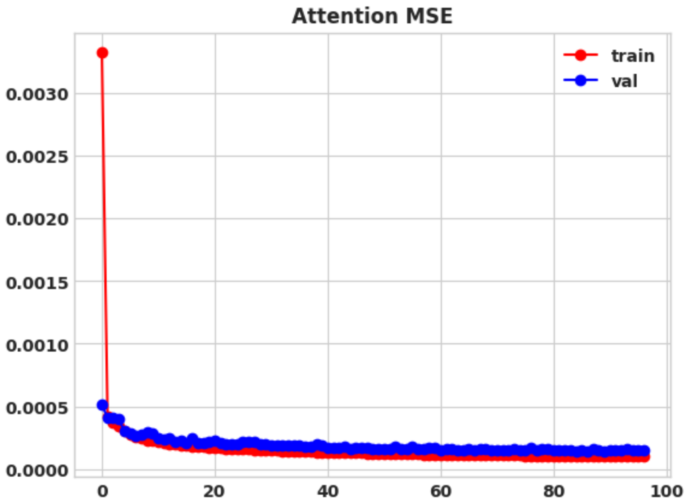

Figure 3. Illustrates the decreasing trend of the MSE values across epochs.

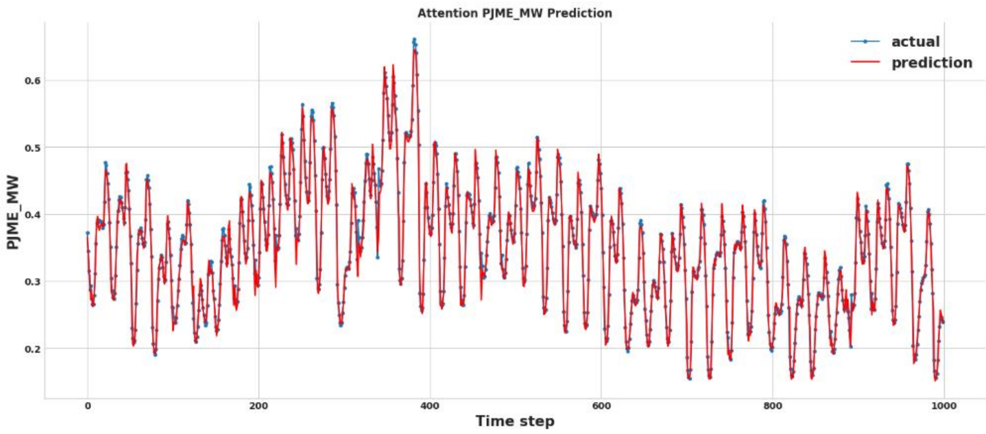

Figure 4. The predicted values are generated by the proposed hybrid model alongside the actual consumption values.

5.2 Discussion

The results demonstrate that the proposed hybrid attention-deep neural network model significantly outperforms the other models in several key metrics. Specifically, the Attention DNN achieved the highest RMS, MAE, and R2 Score of 0.012, 0.007, and 0.992 respectively, indicating the best fit to the observed data among all models evaluated. This high R^2 score suggests that the hybrid model captures the variability in hourly electricity consumption more accurately than the other models. In addition, the lower MSE and RMSE values indicate that the model's predictions are closer to the actual consumption values, reflecting its superior performance.

Comparatively, the ResNet, TCN, and traditional LSTM models also perform well but do not match the accuracy levels achieved by the proposed hybrid model. The RNN and Attention CNN models, while still valuable, exhibit higher error rates and lower R^2 scores, indicating a reduced ability to capture the intricate patterns in the hourly electricity consumption data.

6. Conclusion

This study presented a hybrid model combining LSTM networks and attention mechanisms to improve hourly electricity consumption forecasting. The proposed model outperformed traditional deep learning models such as ResNet, TCN, LSTM, RNN, and Attention CNN across key metrics.

Specifically, the hybrid model achieved the highest R^2 score of 0.992 and the lowest RMSE of 0.012, indicating superior accuracy and reliability. The results demonstrate that integrating attention mechanisms with LSTM significantly enhances the model's ability to capture long-term dependencies and focus on critical time steps. This improved forecasting accuracy has important implications for optimizing resource allocation, ensuring grid stability, and supporting sustainable energy management. Future work can explore adding contextual features like weather and economic data and testing the model in real-time applications to further enhance its predictive capabilities and generalizability.

Author Contributions

All authors contributed equally to this work.

Funding

This research was conducted without external funding support.

Ethical approval

This article does not contain any studies with human participants or animals performed by any of the authors.

Conflicts of Interest

The author declares that there is no conflict of interest in the research

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Data Availability Statement

Not applicable

References

[1] Ahmad, T. and D. Zhang, A critical review of comparative global historical energy consumption and future demand: The story told so far. Energy Reports, 2020. 6: p. 1973-1991.

[2] Vivas, E., H. Allende-Cid, and R. Salas, A systematic review of statistical and machine learning methods for electrical power forecasting with reported mape score. Entropy, 2020. 22(12): p. 1412.

[3] Hochreiter, S. and J. Schmidhuber, Long short-term memory. Neural computation, 1997. 9(8): p. 1735-1780.

[4] Hewamalage, H., C. Bergmeir, and K. Bandara, Recurrent neural networks for time series forecasting: Current status and future directions. International Journal of Forecasting, 2021. 37(1): p. 388-427.

[5] Abdullah, W. and A. Salah, A novel hybrid deep learning model for price prediction. International Journal of Electrical and Computer Engineering (IJECE), 2023. 13(3): p. 3420-3431.

[6] Vaswani, A., et al., Attention is all you need. Advances in neural information processing systems, 2017. 30.

[7] Abbasimehr, H. and R. Paki, Improving time series forecasting using LSTM and attention models. Journal of Ambient Intelligence and Humanized Computing, 2022. 13(1): p. 673-691.

[8] Ning, Y., H. Kazemi, and P. Tahmasebi, A comparative machine learning study for time series oil production forecasting: ARIMA, LSTM, and Prophet. Computers & Geosciences, 2022. 164: p. 105126.

[9] Elsaraiti, M., et al. Time series analysis of electricity consumption forecasting using ARIMA model. in 2021 IEEE Green technologies conference (GreenTech). 2021. IEEE.

[10] Mahia, F., et al. Forecasting electricity consumption using ARIMA model. in 2019 International Conference on Sustainable Technologies for Industry 4.0 (STI). 2019. IEEE.

[11] Hong, W.-C., Electric load forecasting by support vector model. Applied Mathematical Modelling, 2009. 33(5): p. 2444-2454.

[12] Cao, G. and L. Wu, Support vector regression with fruit fly optimization algorithm for seasonal electricity consumption forecasting. Energy, 2016. 115: p. 734-745.

[13] Fernandez-Jimenez, L.A., et al., Short-term power forecasting system for photovoltaic plants. Renewable Energy, 2012. 44: p. 311-317.
[14] Lizondo, D.F., et al., Análisis de variables temporales para la predicción del consumo eléctrico. 2015.
[15] Abumohsen, M., A.Y. Owda, and M. Owda, Electrical load forecasting using LSTM, GRU, and RNN algorithms. Energies, 2023. 16(5): p. 2283.
[16] Wan, A., et al., Short-term power load forecasting for combined heat and power using CNN-LSTM enhanced by attention mechanism. Energy, 2023. 282: p. 128274.
[17] Yuan, Y., et al., Attention mechanism-based transfer learning model for day-ahead energy demand forecasting of shopping mall buildings. Energy, 2023. 270: p. 126878.
[18] PJM, hourly power consumption data. PJM's website. https://www.kaggle.com/datasets/robikscube/hourly-energy-consumption.
[19] Jais, I.K.M., A.R. Ismail, and S.Q. Nisa, Adam optimization algorithm for wide and deep neural network. Knowl. Eng. Data Sci., 2019. 2(1): p. 41-46.
[20] Prechelt, L., Early stopping-but when?, in Neural Networks: Tricks of the trade. 2002, Springer. p. 55-69.

Received: 05 Feb 2024, **Revised:** 24 May 2024,

Accepted: 23 Jun 2024, **Available online:** 26 Jun 2024.