International Journal of Computational and Experimental Science and ENgineering (IJCESEN)

Vol. 11-No.3 (2025) pp. 4117-4125
http://www.ijcesen.com

ISSN: 2149-9144

Research Article

A Modified Energy Demand Forecasting Model using Hybrid CNN-LSTM with Transformer for Univariate Time Series

Teleron, Jerry I[1], Gonzales, Shem L[2], Fajardo, Arnel C[3]

[1]Department of Computer Engineering, Surigao Del Norte State University, Philippines
*** Corresponding Author Email:** jteleron@ssct.edu.ph - **ORCID:** 10000-0001-7406-1357

[2]Department of Information Technology, Surigao Del Norte State University, Del Carmen Campus, Philippines
Email: sgonzales@ssct.edu.ph- **ORCID:** 20009-0001-2329-6778

[3]Department of Information Technology, Isabela State University, Philippines
Email: acfajardo@gmail.com - **ORCID:** 3000-0002-1231-6382

Article Info:

DOI: 10.22399/ijcesen.2293
Received : 22 March 2025
Accepted : 07 May 2025

Keywords

Energy Demand Forecasting
Univariate Time Series,
CNN-LSTM Model
Transformer Layer
Forecasting Accuracy

Abstract:

Precise energy demand forecasting is important in managing electrical power systems, particularly if univariate time series analysis can be applied. To overcome the shortcomings of traditional hybrid models, this paper proposes an improved deep learning architecture that combines Transformer layers, Long Short-Term Memory (LSTM), and Convolutional Neural Networks (CNN). The proposed architecture was trained and validated on historical hourly energy demand data from 2015 to 2018. Performance evaluation revealed that the CNN-LSTM-Transformer model significantly improved forecasting accuracy compared to the baseline CNN-LSTM model. Specifically, the hybrid model achieved a Mean Absolute Error (MAE) of 234.25, Root Mean Squared Error (RMSE) of 386.15, and Mean Absolute Percentage Error (MAPE) of 0.84%, alongside an R^2 score of 99.28%. These results confirm the model's robustness in capturing both local temporal dynamics and long-range dependencies, making it a promising solution for real-time energy forecasting applications.

1. Introduction

Electrical grid stability and effectiveness depend on accurately forecasting energy demand, particularly as power systems become more complicated. Accurate predictions support better decision-making for energy generation, distribution, and consumption. This paper presents a modified deep learning method to improve prediction accuracy in univariate time series by combining Transformer architectures, Long Short-Term Memory (LSTM) networks, and Convolutional Neural Networks (CNN) [1]. CNNs are good at identifying local patterns in data, LSTMs manage time-varying sequential relationships, and Transformers provide attention mechanisms that can model long range interactions [2]. The integration of these models aim to surpass other models in predicting energy demand for long term reference and utilization.

Numerous studies have been conducted to demonstrate the efficacy of deep learning in solving time series problems [1][2][3]. CNNs and LSTMs are widely used in hybrid models for energy forecasting due to their respective strengths in feature extraction and sequence modeling [4]. For example, a CNN-LSTM hybrid has been shown to provide better results than standalone LSTM or CNN models in power load forecasting tasks [5]. More recently, Transformer-based models have gained popularity for their attention mechanisms, which allow the model to focus on the most relevant parts of the sequence [6]. The Informer and Autoformer architectures are two such innovations that demonstrate strong performance in long sequence forecasting. Some studies have also introduced attention-augmented CNN-LSTM models, revealing noticeable improvements in forecast accuracy [7][8]. These efforts collectively highlight the potential of combining multiple architectures to improve forecasting tasks in energy systems.

Despite progress, many studies tend to use either CNN-LSTM hybrids without attention mechanisms or rely solely on Transformer models that may not fully exploit local patterns in the data. Most existing applications of Transformer-based forecasting are focused on multivariate time series or use synthetic datasets, leaving a gap in their application to real-world univariate energy demand datasets. Moreover, few studies investigate the combined effect of CNN, LSTM, and Transformer in a single architecture for energy forecasting, particularly for univariate data. As a result, there is limited understanding of how these models perform collectively when applied to actual energy demand records over multiple years.

This study seeks to address this gap by proposing a modified hybrid CNN-LSTM-Transformer model tailored to univariate energy demand forecasting. The model leverages CNN layers to extract meaningful short-term features, LSTM layers to learn sequential dependencies, and Transformer layers to incorporate long-term attention across the time series. By doing so, the model aims to capture both local and global temporal patterns, enhancing its forecasting capability. Unlike prior work that isolates these techniques, this unified approach is expected to yield more accurate and stable forecasts. The results of this model could provide valuable insights for energy providers and system operators seeking reliable demand prediction tools.

Objective

The main objective of this study is to develop a hybrid forecasting model combining Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM) networks, and Transformer architecture to improve the prediction accuracy of univariate energy demand time series data for the period from January 1, 2015, to December 31, 2018.

Specifically:

1. To integrate CNN, LSTM, and Transformer networks for enhancing the performance in energy demand time series forecasting;
2. To evaluate the performance of the proposed hybrid model by comparing it with hybrid CNN-LSTM, and hybrid CNN-LSTM with Transformer models error metrics such as Mean Absolute Error (MAE), MAPE, Root Mean Squared Error (RMSE), and R2; and
3. To generate a four-year daily energy demand forecast that can be used as reference for future needs.

2. Methods

2.1 Hardware

The combination of an Intel Core i7-8700 CPU, 16GB RAM, 1TB HDD, and a GTX 1660 GPU offers a practical baseline for evaluating performance in simulations, data analysis, and multimedia processing. This setup allows researchers to measure how well the system handles multitasking, real-time rendering, and large data sets. However, the reliance on a mechanical hard drive introduces slower read/write speeds, which may affect data-intensive operations. Integrating a solid-state drive in future studies could provide comparative insights on performance improvements, making this configuration a relevant case for evaluating hardware efficiency in applied computing research.

Table 1. *Computer Hardware Specification*

Hardware Components	**Specification**
CPU	Intel Core i7-87000 CPU @ 3.20 Ghz
Memory	16GB RAM
Storage	1 TB HDD
GPU	6GB NVIDIA GTX `1660

2.2 Software

The software tools listed in table 2 are integral components of the open-source for scientific computing. Anaconda Navigator functions as a graphical user interface for managing Python environments and packages, streamlining workflow setup for data science applications. Spyder, a lightweight IDE tailored for scientific computing, integrates an advanced code editor, interactive IPython console, and debugging capabilities, making it suitable for complex numerical tasks. Jupyter Notebook enables the execution of live code within a browser interface, supporting rich media outputs and facilitating reproducible research. Python, the underlying programming language, is renowned for its extensive libraries, clean syntax, and versatility across domains such as data analysis, machine learning, and software development.

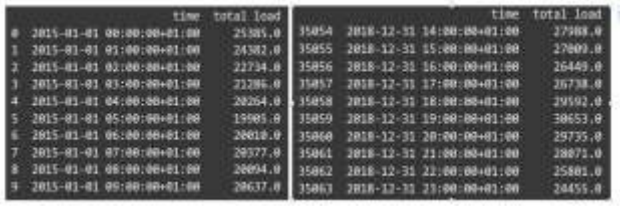
Table 2. *List of Software*

Name	**License**
Anaconda Navigator	Open-Source
Spyder	Open-Source
Jupyter Notebook	Open-Source
Python Programming Language	Open-Source

2.3 Implementation Flow

2.3.1 Data Gathering

The data for this study was obtained from the transmission system operator (TSO), which provided detailed records of energy demand spanning from January 1, 2015, to December 31, 2018 [9]. The dataset consists of hourly energy consumption data, recorded as univariate time series data. Print(data.head(10)) and print(data.tail(10)) were used to display the first and last 10 records, respectively, as shown in Figures. 1a and 1b. The complete dataset, which shows the total energy use for each time step over the specified period, is displayed in Figure 2.



	time	total load
0	2015-01-01 00:00:00+01:00	25385.0
1	2015-01-01 01:00:00+01:00	24382.0
2	2015-01-01 02:00:00+01:00	22734.0
3	2015-01-01 03:00:00+01:00	21286.0
4	2015-01-01 04:00:00+01:00	20264.0
5	2015-01-01 05:00:00+01:00	19905.0
6	2015-01-01 06:00:00+01:00	20010.0
7	2015-01-01 07:00:00+01:00	20377.0
8	2015-01-01 08:00:00+01:00	20094.0
9	2015-01-01 09:00:00+01:00	20637.0

	time	total load
35054	2018-12-31 14:00:00+01:00	27988.0
35055	2018-12-31 15:00:00+01:00	27009.0
35056	2018-12-31 16:00:00+01:00	26449.0
35057	2018-12-31 17:00:00+01:00	26738.0
35058	2018-12-31 18:00:00+01:00	29592.0
35059	2018-12-31 19:00:00+01:00	30653.0
35060	2018-12-31 20:00:00+01:00	29735.0
35061	2018-12-31 21:00:00+01:00	28071.0
35062	2018-12-31 22:00:00+01:00	25801.0
35063	2018-12-31 23:00:00+01:00	24455.0

(a) (b)

Figure 1. *Sample of records from the dataset*

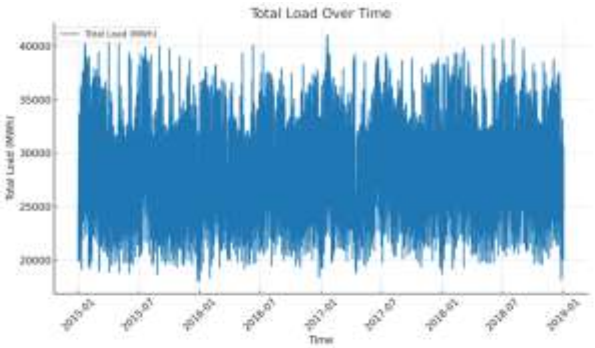


Figure 2*. Total load records over time*

2.3.3 Data Preprocessing

The initial step involves loading the dataset using the pandas.read_csv() function, which reads the CSV file into a DataFrame. Once the data is loaded, duplicate entries are removed using drop_duplicates(). Removing duplicates is essential as they can introduce redundancy into the dataset, leading to biased learning and potentially overfitting the model [10][11]. After cleaning, missing values are handled using forward fill (fillna(method='ffill')), which replaces missing data with the most recent valid observation [12][13][14]. The next preprocessing step is feature selection. The code selects the relevant column (typically representing the time series data such as energy demand) and reshapes the values into a 2D array to ensure that the data is in a format suitable for machine learning algorithms [15]. After reshaping, the data is normalized using MinMaxScaler from the sklearn.preprocessing module. Scaling the data to a range between 0 and 1 is critical for CNN models, as they are sensitive to the magnitude of input data. Normalization ensures that all features contribute equally to the model, preventing any one feature from dominating the learning process and improving convergence speed during training [16][17].

2.3.4 Deep Learning Architecture

2.3.4.1 Convolutional Neural Network (CNN) Layer

The first layer of the model utilizes CNNs to extract local features from the input time series data. The CNN layer helps in identifying spatial patterns and short-term dependencies in the data by applying convolutional filters to the time series [18][19]. This layer processes the input data with a series of convolutional and pooling operations to reduce dimensionality and highlight key features, which are essential for subsequent forecasting tasks.

2.3.4.2 Long Short-Term Memory (LSTM) Layer:

Following the CNN layer, the LSTM network is employed to capture the long-term temporal dependencies present in the energy demand time series. LSTMs are a type of recurrent neural network (RNN) that can retain information over extended periods, which is crucial for understanding trends and cycles in energy consumption [20][21]. The LSTM layer helps to model the sequential nature of the time series data by processing the extracted features from the CNN layer [22]. This layer is composed of multiple LSTM cells to ensure the model can effectively remember relevant past information while mitigating the vanishing gradient problem often encountered in traditional RNNs.

2.3.4.3 Transformer Layer

To further enhance the model's ability to capture long-range dependencies and contextual relationships in the time series, the Transformer layer is integrated into the architecture. This component uses self-attention mechanisms to weigh the importance of different time steps in the series, allowing the model to focus on the most relevant data points. The Transformer's attention mechanism is particularly useful in capturing global dependencies that may span long intervals within the data, which traditional LSTM networks may struggle to model effectively [23]. Multi-head attention is used to allow the model to attend to different aspects of the data simultaneously, improving its performance in forecasting [24][25][25].

2.4 Propose Architecture

The proposed hybrid deep learning architecture integrates Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM) networks, and Transformer models to forecast energy demand in univariate time series data. This integration seeks to leverage the strengths of each model to improve forecasting accuracy, particularly in capturing both local and global dependencies.

The CNN component serves as the initial layer, where it performs feature extraction from the raw time series data. CNNs are particularly effective at detecting short-term fluctuations and local patterns, such as spikes or daily periodicities, that are common in energy demand data [18]. These local features are essential for accurate forecasting in the short-term, as energy consumption often experiences rapid fluctuations or changes over small time intervals. CNN's ability to process data through convolutional filters allows it to identify important characteristics like trend changes or anomalies in the energy demand, which are then passed to the next layers for further processing [19]. This feature extraction step ensures that the subsequent layers work with more relevant and condensed data that can help improve the performance of the desired output.
Following the CNN layers, the LSTM network is introduced to model the temporal dependencies within the time series data. Energy demand typically exhibits long-term patterns, such as daily and weekly cycles, seasonal variations, or annual trends, which the LSTM is well-suited to capture. LSTMs are designed to handle sequences, learning from past data and retaining information over long intervals. By using gates to control the flow of information, LSTMs are able to learn which data points to remember and which to forget, allowing the model to effectively learn from historical energy demand while avoiding overfitting. This capability is particularly important in energy forecasting, where past demand can heavily influence future predictions. The LSTM layers, therefore, provide the model with the ability to understand and predict long-term trends in energy consumption, which are essential for accurate forecasting.
The Transformer component is added to the architecture to address the limitation of traditional sequence-based models in capturing long-range dependencies. While CNN and LSTM are powerful in identifying local patterns and learning temporal dependencies, they may struggle with long-term relationships that span over many time steps. The Transformer architecture, with its self-attention mechanism, allows the model to focus on the most relevant time steps in the series, regardless of their position in the sequence. This attention mechanism evaluates the importance of each time step in relation to others, enabling the model to learn which past events or patterns are most influential for predicting future demand. This is particularly useful in energy forecasting, where certain external events or periodic changes may not be immediately adjacent in the time series but still significantly impact future demand. The Transformer's ability to focus on global relationships across the sequence helps capture these distant dependencies, further enhancing the model's forecasting capabilities.
The combination of these three models benefit from the strengths of each, providing a more robust solution for forecasting tasks that require understanding both short-term fluctuations and long-term trends. The hybrid model can effectively handle energy demand's complex patterns, offering superior forecasting performance compared to single-model approaches.
Finally, the fully connected (FC) layers are employed after the CNN, LSTM, and Transformer layers to aggregate the learned features and produce the final output. These layers help synthesize the information captured by the previous layers and make the final prediction. The output layer, using a linear activation function, generates the forecasted energy demand value for the next time step. The model is trained using mean squared error (MSE) as the loss function to minimize the prediction error and improve the model's accuracy over time. Fig 3 illustrates the entire flow of the proposed architecture.

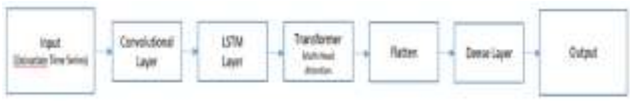


Figure 3. *Hybrid CNN-LSTM with Transformer Architecture*

2.5 Model Training and Evaluation

Once the data is cleaned, reshaped, and scaled, the code generates sequences of fixed length for the time series forecasting task. This is accomplished using the create_sequences() function, which splits the time series into input-output pairs. Each input consists of a sequence of past observations, while the corresponding output is the next time step in the series. This technique, often referred to as the sliding window approach, is fundamental for teaching the model to learn from previous data to predict future values [27]. Sequence generation helps capture temporal dependencies that are critical for accurate time series forecasting [28]. After generating sequences, the dataset is shuffled to avoid any biases introduced by the ordering of the data, ensuring that the model learns generalizable patterns rather than memorizing specific sequences. The shuffled dataset is then split into training and testing sets, with 80% used for training and the remaining 20% reserved for testing. This division ensures that the model is evaluated on unseen data, providing a robust measure of its generalization capability [29].
The model's performance is evaluated using MAE, MAPE, RMSE, and R^2. MAE measures the average absolute difference between predicted and actual

values, with lower values indicating better accuracy. However, it doesn't penalize larger errors more heavily. MAPE expresses the error as a percentage of the actual values, offering a relative measure of accuracy, but it can be skewed by small actual values. RMSE penalizes large errors by squaring the differences before averaging, making it sensitive to outliers and large deviations. Finally, R^2 indicates the proportion of variance in the data explained by the model, with higher values showing better fit. Together, these metrics provide a comprehensive view of the model's accuracy, sensitivity to errors, and its ability to explain the variability in energy demand [30][31].

3. Results and Discussion

3.1 Training and Loss Curve Analysis

Figure 4 illustrates the training and validation loss curves for both the Hybrid CNN-LSTM model and the CNN-LSTM-Transformer model. In both cases, a rapid decrease in loss is observed within the first 50 iterations, indicating effective learning and fast convergence during the early stages of training. The Hybrid CNN-LSTM model demonstrates a smooth convergence pattern, with only a minimal gap between training and validation loss curves, suggesting good generalization without significant overfitting.

Similarly, the CNN-LSTM-Transformer model exhibits consistently low loss values across the training epochs. The close alignment between its training and validation loss curves indicates stable learning behavior and further affirms the model's robustness against overfitting. The consistently low loss values also reflect the model's ability to effectively minimize prediction error while maintaining stability across the dataset.

Overall, these patterns suggest that both hybrid models—especially the CNN-LSTM-Transformer—achieved efficient training dynamics, ensuring reliable and accurate forecasting outcomes suitable for real-world energy demand prediction applications.

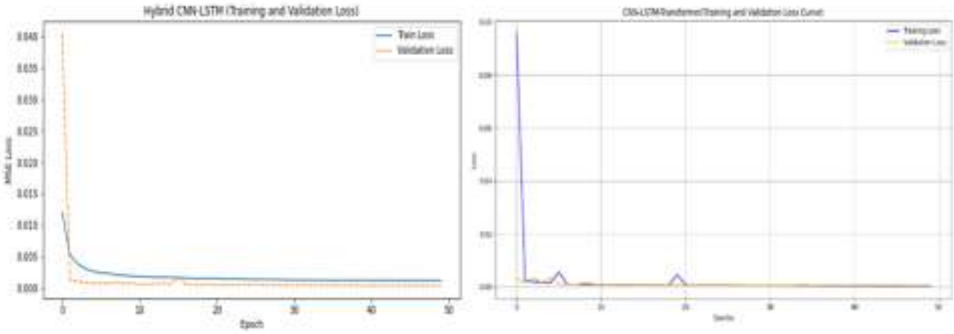


Figure 4*. Training and validation curves of CNN-LSTM and CNN-LSTM with Transformer*

3.2 Performance Metrics Evaluation

The comparative performance assessment between the CNN-LSTM and CNN-LSTM-Transformer models clearly demonstrates the significant impact of integrating the Transformer layer into the hybrid architecture. As presented in Table 3, the CNN-LSTM-Transformer model achieved superior results across all key performance metrics. Specifically, it recorded a lower Mean Absolute Error (MAE) of 232.95, a substantial improvement over the 499.08 MAE exhibited by the CNN-LSTM model. Likewise, the Mean Absolute Percentage Error (MAPE) decreased from 1.80% to 0.83%, while the Root Mean Squared Error (RMSE) was reduced from 671.37 to 374.88. These reductions indicate a higher degree of prediction accuracy and a significant minimization of forecast errors.

In addition to these error metrics, the coefficient of determination (R^2) further supports the enhanced performance of the proposed model, increasing from 97% in the CNN-LSTM configuration to 99.33% with the Transformer integration. A higher R^2 value signifies a more precise fit between the predicted and actual energy demand values, reflecting the model's robust capability to generalize across unseen data.

These results collectively highlight the efficacy of incorporating the Transformer's self-attention mechanism, which enables the model to better capture long-range temporal dependencies and complex patterns within the univariate time series data. The improved performance metrics affirm that the CNN-LSTM-Transformer hybrid model offers a more accurate, reliable, and scalable solution for energy demand forecasting compared to traditional hybrid approaches.

***Table 3** Model comparison from the different metrics used.*

Architecture	MAE	MAPE	RMSE	R2
CNN-LSTM	499.08	1.80%	671.37	97%
CNN-LSTM-Transformer	234.25	.84%	386.15	99.28%

4.3 Time Series Prediction Analysis

Figure 5 presents a comparative visualization of the actual versus predicted energy consumption values over time, showcasing the predictive performance of the CNN-LSTM-Transformer model. The high-density overlap observed between the two time-series curves indicates a strong alignment, demonstrating the model's high predictive accuracy. This close correspondence suggests that the model effectively captures both short-term fluctuations and longer-term temporal dependencies embedded within the data.
The minimal visible deviation between the actual and predicted values highlights the model's ability to generalize well across the dataset, ensuring consistency and reliability in forecasting. This outcome can be attributed to the hybrid model's architectural strengths: convolutional layers enable efficient feature extraction, LSTM units adeptly handle sequential patterns, and the Transformer's self-attention mechanisms enhance the modeling of long-range dependencies. Together, these components contribute to precise and dynamic forecasting performance.
The robust prediction results suggest that the CNN-LSTM-Transformer model is well-suited for real-world energy demand forecasting, particularly in environments characterized by dynamic and volatile consumption patterns. Its ability to adapt to both short-term and long-term variations strengthens its applicability for operational decision-making and strategic planning within energy management systems.

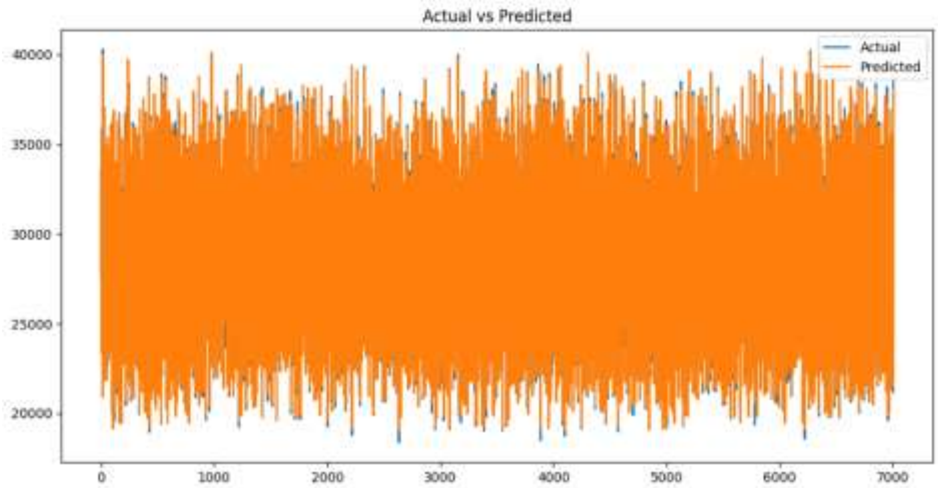


***Figure 5**. Actual vs. predicted daily energy consumption based on CNN-LSTM-Transformer Model output*

3.4 Five-Year Forecasted Daily Energy Consumption Analysis

Figures 6a and 6b display the forecasted daily energy consumption values generated by the CNN-LSTM-Transformer model, focusing on sample ten-day periods in late December for the years 2025 and 2028, respectively. The figures illustrate the model's capacity to accurately project future energy demands based on historical patterns and learned temporal dependencies.
The close alignment between the forecasted trends for both years underscores the model's capability to generalize across different time frames, even under the influence of seasonal variations and potential consumption shifts. This highlights the strength of employing advanced deep learning architectures in time series analysis, particularly for applications that require high forecasting precision, such as energy demand management.
By demonstrating accurate short-term forecasting over distinct future periods, the CNN-LSTM-Transformer model affirms its potential as a valuable tool for predictive analytics in dynamic and evolving environments, thereby supporting proactive planning and resource optimization initiatives in the energy sector.

	Date	CNN-LSTM-Transformer Daily Forecasted Energy Consumption
2547	2025-12-22	30623.270
2548	2025-12-23	30079.574
2549	2025-12-24	28923.340
2550	2025-12-25	28214.004
2551	2025-12-26	28380.744
2552	2025-12-27	28894.977
2553	2025-12-28	29485.928
2554	2025-12-29	30361.963
2555	2025-12-30	31090.791
2556	2025-12-31	30092.430

(a)

	Date	CNN-LSTM-Transformer Daily Forecasted Energy Consumption
3643	2028-12-22	28234.604
3644	2028-12-23	29392.717
3645	2028-12-24	29838.213
3646	2028-12-25	30252.768
3647	2028-12-26	30608.676
3648	2028-12-27	30249.838
3649	2028-12-28	29135.135
3650	2028-12-29	28259.414
3651	2028-12-30	28305.250
3652	2028-12-31	28781.777

(b)

Figure. 6. *Sample of daily forecasted energy for years 2025-2028*

Figure 7 illustrates the forecasted average yearly energy demand from 2025 through 2028, with values ranging narrowly from approximately 27,244 to 27,335 megawatt-hours. Understanding this trend is crucial for energy providers, as it allows for better planning in terms of resource allocation, infrastructure development, and sustainability strategies

Figure 7. *Forecasted energy demand from 2025 to 2028*

4. Conclusion

The experimental findings validate that integrating Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM) networks, and Transformer modules into a unified model architecture significantly enhances the performance of univariate energy demand forecasting. The proposed hybrid model consistently outperformed the conventional CNN-LSTM configuration across all evaluation metrics. Specifically, the inclusion of the Transformer's self-attention mechanism led to substantial reductions in prediction errors, lowering the Mean Absolute Error (MAE) from 499.08 to 234.25 and the Root Mean Squared Error (RMSE) from 671.37 to 386.15. Similarly, the Mean Absolute Percentage Error (MAPE) decreased from 1.80% to 0.84%, reflecting greater relative accuracy. Furthermore, the coefficient of determination (R^2) improved from 97% to 99.28%, indicating a superior model fit to the actual demand data. These results collectively demonstrate the model's capacity to effectively capture short-term fluctuations through CNN, sequential dependencies through LSTM, and long-range temporal correlations through Transformer, thereby offering a more comprehensive, precise, and reliable forecasting framework.

Recommendations

In light of the findings, the following recommendations are proposed to further enhance the model's performance, adaptability, and scalability:

1. Incorporate Exogenous Variables:
Future work should integrate external factors such as weather data and economic indicators to validate the model's adaptability. Additionally, applying the model to multivariate datasets from diverse domains or regions is recommended to assess its broader applicability.
2. Utilize Automated Tuning Frameworks:
Employing automated tuning methods, such as Bayesian optimization or genetic algorithms, could uncover optimal configurations across the CNN, LSTM, and Transformer components, leading to potential improvements in model performance.
3. Test on Scalable Computing Environments:
It is suggested that future implementations test the model architecture on scalable environments like distributed cloud platforms to evaluate performance under high-demand and large-scale conditions.

Author Statements:

- **Ethical approval:** The conducted research is not related to either human or animal use.
- **Conflict of interest:** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper
- **Acknowledgement:** The authors declare that they have nobody or no-company to acknowledge.
- **Author contributions:** The authors declare that they have equal right on this paper.
- **Funding information:** The authors declare that there is no funding to be acknowledged.

- **Data availability statement:** The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

References

[1] Cui, B., Liu, M., Li, S., Jin, Z., Zeng, Y., & Lin, X. (2023). Deep learning methods for atmospheric $PM_{2.5}$ prediction: A comparative study of transformer and CNN-LSTM-attention. *Atmospheric Pollution Research*, 14(9), 101833. https://doi.org/10.1016/j.apr.2023.101833

[2] Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M. A., Al-Amidie, M., & Farhan, L. (2021). Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. *Journal of Big Data*, 8, 53. https://doi.org/10.1186/s40537-021-00444-8

[3]. Kaya, M., Utku, A., & Canbay, Y. (2024). A Hybrid CNN-LSTM Model for Predicting Energy Consumption and Production Across Multiple Energy Sources. *Journal of Soft Computing and Artificial Intelligence*, 5(2), 63-73. https://doi.org/10.55195/jscai.1577431

[4] Yıldız Doğan, G., Aksoy, A., & Öztürk, N. (2024). A Hybrid Deep Learning Model to Estimate the Future Electricity Demand of Sustainable Cities. *Sustainability*, 16(15), 6503. https://doi.org/10.3390/su16156503

[5] Guo, W., Liu, S., Weng, L., & Liang, X. (2025). Power Grid Load Forecasting Using a CNN-LSTM Network Based on a Multi-Modal Attention Mechanism. *Applied Sciences*, 15(5), 2435. https://doi.org/10.3390/app15052435

[6] Wu, H., Xu, J., Wang, J., & Long, M. (2021). Autoformer: Decomposition Transformers with Auto-Correlation for Long-Term Series Forecasting. *Advances in Neural Information Processing Systems*, 34, 22419–22430. https://proceedings.neurips.cc/paper_files/paper/2021/hash/9d86d84eec0f2cd5f2b8f2f707b6b4c1-Abstract.html

[7] Yang, F., Liu, Z., Yan, J., & Geng, Y. (2022). CNN-LSTM neural network with attention mechanism for fault diagnosis of high voltage circuit breakers. In *18th International Conference on AC and DC Power Transmission (ACDC 2022)* (pp. 1102–1107). IET. https://doi.org/10.1049/icp.2022.1357

[8] Binbusayyis, A., & Sha, M. (2025). Energy consumption prediction using modified deep CNN-Bi LSTM with attention mechanism. *Heliyon*, 11(1), e41507. https://doi.org/10.1016/j.heliyon.2024.e41507

[9] K. Roy, A. Ishmam and K. A. Taher, "Demand Forecasting in Smart Grid Using Long Short-Term Memory," *2021 International Conference on Automation, Control and Mechatronics for Industry 4.0 (ACMI), Rajshahi, Bangladesh, 2021*, pp. 1-5, doi: 10.1109/ACMI53878.2021.9528277

[10] Lee, K., Ippolito, D., Nystrom, A., Zhang, C., Eck, D., Callison-Burch, C., & Carlini, N. (2021). Deduplicating training data makes language models better. *arXiv preprint* arXiv:2107.06499. https://arxiv.org/abs/2107.06499

[11] Aghabagherloo, A., Abadi, A., Sarkar, S., Dasu, V. A., & Preneel, B. (2025). Impact of data duplication on deep neural network-based image classifiers: Robust vs. standard models. *arXiv preprint* arXiv:2504.00638. https://arxiv.org/abs/2504.00638

[12] Alwateer, M., Atlam, E. S., El-Raouf, M. M. A., Ghoneim, O. A., & Gad, I. (2024). Missing data imputation: A comprehensive review. *Journal of Computer and Communications*, 12(11), 53–75. https://doi.org/10.4236/jcc.2024.1211004

[13]The pandas development team. (n.d.). pandas.DataFrame.fillna. *pandas documentation*. https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.fillna.html

[14] Wijaya, C. Y. (2023, February 1). The optimal way to input missing data with pandas fillna(). *KDnuggets.* https://www.kdnuggets.com/2023/02/optimal-way-input-missing-data-pandas-fillna.html

[15] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830. http://jmlr.org/papers/v12/pedregosa11a.html

[16] Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT Press. https://www.deeplearningbook.org/

[17] Chakraborty, A., Tomsett, R., Raghavendra, R., Harborne, D., Alzantot, M., Cerutti, F., ... & Srivastava, M. (2021). Interpretability of deep learning models: A survey of results. *IEEE Transactions on Neural Networks and Learning Systems*, 32(11), 4793–4813. https://doi.org/10.1109/TNNLS.2020.3029375

[18] Chen, Y., Kang, Y., Chen, Y., & Wang, Z. (2020). Probabilistic forecasting with temporal convolutional neural network. *Neurocomputing*, 399, 491–501. https://doi.org/10.1016/j.neucom.2020.03.011

[19] Fazel Mojtahedi, F., Yousefpour, N., Chow, S. H., & Cassidy, M. (2025). Deep learning for time series forecasting: Review and applications in geotechnics and geosciences. *Archives of Computational Methods in Engineering*. https://doi.org/10.1007/s11831-025-10244-5

[20] Chou, J.-S., & Ren, C. (2020). Cloud-based short-term solar power forecasting using LSTM networks. *Sustainable Cities and Society*, 55, 102010. https://doi.org/10.1016/j.scs.2020.102010

[21] Zhou, K., Yang, S., & Shao, Z. (2022). Short-term electric load forecasting with LSTM networks based on pattern sequence similarity. *Applied Energy*, 313, 118798. https://doi.org/10.1016/j.apenergy.2022.118798

[22] Chung, J., & Jang, B. (2022). Accurate prediction of electricity consumption using a hybrid CNN-

LSTM model based on multivariable data. *PLOS ONE*, 17(11), e0278071. https://doi.org/10.1371/journal.pone.0278071
[23] Bu, S.-J., & Cho, S.-B. (2020). Time series forecasting with multi-headed attention-based deep learning for residential energy consumption. *Energies*, 13(18), 4722.
[24] Shi, J., Wang, S., Qu, P., & Shao, J. (2024). Time series prediction model using LSTM-Transformer neural network for mine water inflow. *Scientific Reports*, 14, Article 18284.
[25] Su, L., Zuo, X., Li, R., Wang, X., Zhao, H., & Huang, B. (2025). A systematic review for transformer-based long-term series forecasting. *Artificial Intelligence Review*, 58, Article 80.
[26] Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., & Zhang, W. (2021). Informer: Beyond efficient transformer for long sequence time-series forecasting. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(12), 11106–11115.
[27] Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., & Zhang, W. (2021). Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting. *In Proceedings of the AAAI Conference on Artificial Intelligence*, 35(12), 11106–11115.
[28] Wu, H., Xu, J., Wang, J., & Long, M. (2021). Autoformer: Decomposition Transformers with Auto-Correlation for Long-Term Series Forecasting. *Advances in Neural Information Processing Systems*, 34, 22419–22430.
[29] Lim, B., Arık, S. Ö., Loeff, N., & Pfister, T. (2021). Temporal Fusion Transformers for Interpretable Multi-horizon Time Series Forecasting. *International Journal of Forecasting*, 37(4), 1748–1764.
[30] Chicco, D., Warrens, M. J., & Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Computer Science*, 7, e623. https://doi.org/10.7717/peerj-cs.623
[31]Li, X. (2023). A comparative study of statistical and machine learning models on near-real-time daily emissions prediction. *arXiv preprint* arXiv:2302.01152. https://arxiv.org/abs/2302.01152