Black Sea Journal of Engineering and Science
doi: 10.34248/bsengineering.1799782

Open Access Journal
e-ISSN: 2619 – 8991

Research Article
Volume 9 - Issue 2: 872-879 / March 2026

TOWARDS SMARTER ENERGY PLANNING: DISTRICT-BASED NATURAL GAS CONSUMPTION FORECASTING IN ISTANBUL WITH BI-GRU

Hikmet CANLI[1*]

[1]*Istanbul Gedik University, Faculty of Engineering, Department of Software Engineering, 3400, Istanbul, Türkiye*

Abstract: With the increasing demand for energy, accurate forecasting of natural gas consumption in large cities has gained strategic importance. This paper compares the performance of various forecasting models using natural gas consumption data of districts in Istanbul, the most populous city in Türkiye. Using a time series analysis approach, the deep learning models Long Short Term Memory (LSTM), Gated Recurrent Unit (GRU), Bidirectional LSTM (BI-LSTM), Bidirectional GRU (BI-GRU) and Recurrent Neural Network (RNN) are compared with machine learning models Linear Regression (LR), Ridge, Support Vector Regression (SVR), Random Forest (RF) and Gradient Boosting algorithms. Model performance was evaluated with R^2 and Mean Absolute Percentage Error (MAPE) metrics. According to the results, the BI-GRU model provided the highest accuracy with a MAPE of 12.98% and an R^2 value of 0.8902. LSTM, BI-LSTM and RNN models also showed high success. In contrast, linear regression-based models performed poorly with low R^2 and high MAPE values. Among the machine learning methods, only the Random Forest model gave a strong result with an R^2 of 0.9063, while the MAPE value remained high. As a result, deep learning models, especially BI-GRU, better capture short-term consumption fluctuations and provide better forecasts compared to classical models.

Keywords: Energy consumption, Natural gas, Deep learning, Machine learning

***Corresponding author:** Istanbul Gedik University, Faculty of Engineering, Department of Software Engineering, 3400, Istanbul, Türkiye
E mail: Hikmet.canli@gedik.edu.tr (H. CANLI)
Hikmet CANLI https://orcid.org/0000-0003-3394-7113

Received: October 09, 2025
Accepted: February 25, 2026
Published: March 15, 2026

Cite as: Canlı, H. (2026). Towards smarter energy planning: district-based natural gas consumption forecasting in Istanbul with BI-GRU. *Black Sea Journal of Engineering and Science*, *9*(2), 872-879.

1. Introduction

Energy is one of the fundamental building blocks of a country's economic and social development. Today, per capita energy consumption is rising significantly with urbanization, industrialization and rising living standards (Stambouli, 2011; Demirel et al., 2017). Increasing energy demand on a global scale lead to the depletion of fossil fuel-based energy resources and supply-demand imbalances, which creates serious challenges in terms of energy supply security, sustainability and planning (Beyca et al., 2019; Azadeh et al., 2019; Haksevenler, 2025).

In this context, accurate and timely forecasting of energy consumption is critical for managing energy supply, planning infrastructure investments and formulating strategic energy policies (Kizilaslan and Karlik, 2008). Especially in metropolitan areas, modeling of different consumption types such as residential and commercial use necessitates the development of intelligent forecasting algorithms suitable for the dynamic nature of the system (Potočnik et al., 2007).

Although real-time energy consumption estimation for natural gas systems at the district level has not been comprehensively studied, there are a limited number of studies investigating real-time or near real-time energy demand estimation in relevant contexts. In particular, recurrent neural network-based approaches have been successfully applied to short-term electricity load forecasting and urban energy management systems, demonstrating their suitability for real-time use due to their ability to model temporal dependencies and seasonal patterns (Marino et al., 2016; Kong et al., 2019). These studies demonstrate that deep learning-based forecasting models can effectively support real-time decision-making processes in smart grid and smart city infrastructures (Zanella et al., 2014). However, most existing studies focus on aggregate electricity or energy consumption data, and real-time natural gas consumption prediction at the district level has been relatively understudied. This study contributes to the literature by addressing this gap through a district-based prediction framework.

Recent studies have further demonstrated the effectiveness of data-driven and machine learning–based approaches for energy consumption forecasting at different spatial and temporal scales. For instance, Karadağ and Sağtaş (2025) investigated artificial intelligence models for energy demand forecasting and reported that machine learning techniques can significantly improve prediction accuracy compared to traditional statistical methods. Similarly, Arslantürk and Şirin (2025) analyzed machine learning approaches for energy management applications and emphasized the importance of advanced learning algorithms in supporting efficient energy planning and decision-making processes. These findings reinforce the relevance of applying deep learning–based forecasting frameworks to urban energy systems and provide methodological support for the district-level natural gas consumption forecasting approach proposed in this study.

This study focuses on accurately predicting natural gas consumption in the districts of Istanbul, the most densely populated city in Türkiye. By combining two datasets based on different customer types, a comprehensive analysis is performed with time series natural gas consumption data. In this context, five different deep learning models (LSTM, GRU, BI-LSTM, BI-GRU, RNN) and five machine learning models (Linear Regression, Ridge, SVR, Random Forest, Gradient Boosting) were configured and model performances were compared.

As shown in Table 2, the highest accuracy was achieved with the BI-GRU model (12.98% MAPE and 0.8902 R^2). Other deep learning models such as BI-LSTM and LSTM also showed high performance. On the other hand, linear regression models showed poor predictive power with low R^2 and high MAPE values, while only the Random Forest model provided high accuracy with an R^2 value of 0.9063, but produced unstable predictions with a high MAPE value of 56.17%.

The main contributions of this study can be summarized as follows:

a) By combining two different datasets of natural gas consumption in Istanbul, a comprehensive dataset suitable for time series analysis has been created.

b) A comprehensive model comparison is performed by applying different deep learning and machine learning models for natural gas consumption.

c) Forecasting performance is evaluated using statistical criteria such as MAPE and R^2, and the most appropriate forecasting model is proposed.

This study compares both classical and advanced deep learning methods on a large dataset covering all 39 districts of Istanbul and created through the integration of different user classes, achieving high spatial resolution prediction performance. In the remainder of the study, the dataset and models used are introduced in detail, model comparisons are made in the light of the findings, and in the last section, suggestions for future studies are presented in line with the findings.

2. Materials and Methods

In this section, the data, methods and implementation process used in the study will be explained in detail. In line with the purpose of the study, information about the preparation of the dataset, analysis methods and the characteristics of the algorithms used will be given.

2.1. Natural Gas Dataset: Monthly Natural Gas Consumption in Istanbul by District

Istanbul Metropolitan Municipality (IBB) open data portal is a free portal that provides access to data published by municipalities and environmental organizations. It provides free access to many datasets under the headings of Economy, Disaster Management, Energy, Life, Governance, People and Environment. This data can be accessed through different methods such as xml, .xlsx, .csv and api. In addition, this data has Istanbul Metropolitan Municipality Open Data License and consists of public sector information licensed under the 4.0 International (CC BY 4.0) license.

Table 1. Natural gas dataset

Feature	Description	Type
id	Sequence Number	Numeric
Year	Measured Year	Numeric
Month	Measured Month	Numeric
District	Consumption Region	Categorical
Class of Use	Combi boiler, stove, shohpen, etc.	Categorical
User Class	Residential, Commercial, Industrial Facilities, etc.	Categorical
Consumption	Total Consumption m3	Numeric

The natural gas consumption data used in this study were obtained from the IBB Open Data Portal, which provides openly licensed and publicly available datasets collected by authorized municipal institutions. According to the documentation provided by the IBB Open Data Portal, consumption data are collected at the district level and obtained from official billing and monitoring systems operated by municipal service providers. The portal ensures data reliability and consistency for scientific research by applying standard data validation and anonymization procedures before publishing data publicly. The Natural Gas Dataset is contains the monthly natural gas consumption of 39 districts in Istanbul from January 2015 to April 2023. It consists of 3901 records in total and its properties are detailed in Table 1.

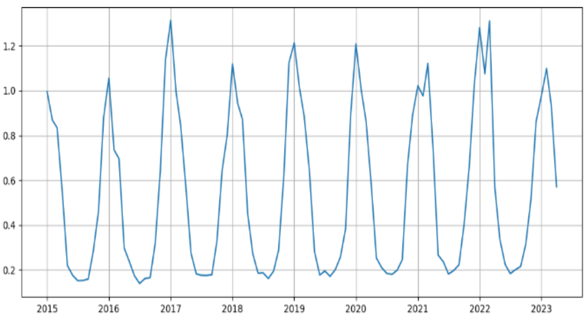

Figure 1. Total natural gas consumption (m^3) by time.

In order to better analyze and understand the dataset, data on the number of users and consumption amount by natural gas usage class for the districts in the same date range, which are also available on IBB open data portal, were added to the dataset.

As part of data preprocessing, missing values were completed using the forward fill method, and consumption values were scaled to the 0–1 range using Min–Max normalization.

Figure 1 shows the total amount of natural gas usage of all districts by time in m3. Since 2023 data is available until April, the remaining data of the year is not included in the graph.

Figure 2 shows the annual natural gas consumption of the 6 districts with the highest consumption in m3. The district with the highest consumption is Esenyurt. The reason why this district consumes more than the others is the high number of meters. Again, consumption data in other districts vary in direct proportion to the number of meters.

Figures 3 and 4 show the amount of natural gas used by type of use and user class. It is observed that the amount of natural gas used in residential buildings is much higher than other classes.

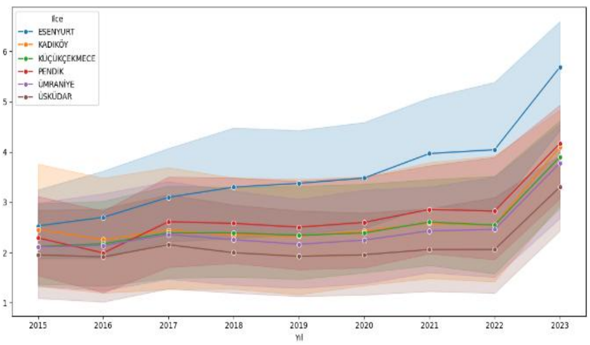

Figure 2. Annual natural gas consumption (m³) in the 6 most consuming districts.

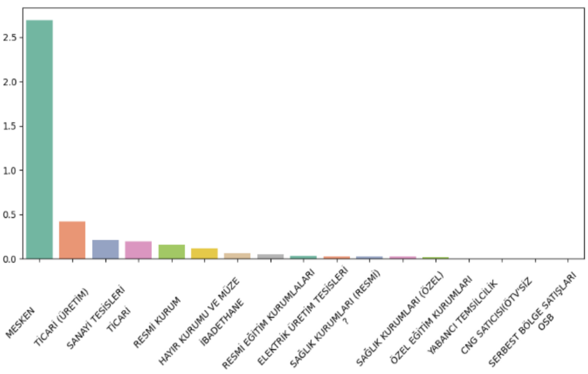

Figure 3. Districts total natural gas consumption (m³) by user class.

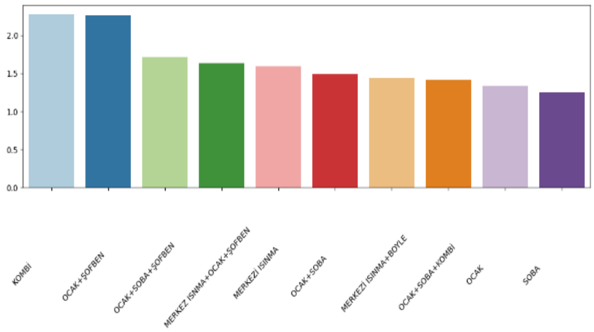

Figure 4. Top 10 most consumed (m³) uses.

2.2. Model Selection

The prediction models used in this study were selected to provide a comprehensive comparison between classical machine learning methods and deep learning-based time series models. Linear regression and tree-based methods were included as baseline models due to their widespread use and interpretability in energy forecasting studies. Recurrent neural network-based models (RNN, LSTM, BI-LSTM, GRU, and BI-GRU) were selected due to their proven ability to capture temporal dependencies and seasonal patterns in sequential energy consumption data. The inclusion of both unidirectional and bidirectional architectures allows for a fair assessment of the impact of bidirectional learning on prediction performance.

2.3. Recurrent Neural Networks (RNN)

Recurrent Neural Networks (RNNs) are types of artificial neural networks designed to work with sequential data. RNNs are used in sequential data-driven tasks such as natural language processing, time series analysis, and speech recognition by retaining information from previous time steps through a "hidden state" (Goodfellow et al., 2016). Unlike classical neural networks, RNNs share the same weights at different time steps and can thus model time-dependent dependencies (LeCun et al., 2015). However, due to problems such as the vanishing gradient in moving information in long sequences, advanced variants such as Gated Recurrent Unit (GRU) and Long Short-Term Memory (LSTM) have been developed (Hochreiter and Schmidhuber, 1997).

2.4. Long Short-Term Memory (LSTM) and BI-LSTM

Long Short-Term Memory (LSTM) is an RNN architecture developed to overcome the vanishing gradient problem of classical RNNs in learning long-term dependencies (Hochreiter and Schmidhuber, 1997). Thanks to its "forget", "enter" and "exit" gates, LSTM can keep information in memory for a long time and delete it when necessary. This makes it widely used in areas such as language modeling, speech recognition and sentiment analysis. Bi-LSTM trains an LSTM network in both forward (from future to past) and backward (from past to future) directions, using both past and future context information at each time step (Schuster and Paliwal, 1997; Başarslan, 2025). This structure can model context better than unidirectional LSTMs and provides higher accuracy, especially in natural language processing tasks (Graves and Schmidhuber 2005).

The input gate controls how much new information enters the cell. It is given in equation 1.

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \tag{1}$$

The Forget gate determines how much of the previous cell state is forgotten. It is given in equation 2.

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f) \tag{2}$$

The output gate determines how much of the cell state is output. It is given in equation 3.

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o) \tag{3}$$

$$C_t = f_t C_{t-1} + i_t \tanh(W_c x_t + U_c h_{t-1} + b_c) \tag{4}$$

$$h_t = o_t \tanh(C_t) \tag{5}$$

equations 4 and 5 give the cell state and hidden state information.

2.5. Gated Recurrent Unit (GRU) and BI-GRU

The GRU is an alternative to the RNN architecture proposed aims to learn long-term dependencies similar to the LSTM, but with a simpler structure. GRU regulates the flow of information through “update” and “reset” gates that control the cell state and is therefore computationally more efficient as it contains fewer parameters. BI-GRU is a bidirectional version of the classical GRU that allows it to receive information from both the past and the future. This structure is particularly successful in tasks where context information is important, such as language modeling, speech recognition and biomedical data analysis (Schuster and Paliwal, 1997; Yang et al., 2016). By looking at the data in forward and backward directions, BI-GRU can extract more information from the context, which improves accuracy, especially in sequence classification tasks.

Update gate (equation 6);

$$z_t = \sigma(W_z x_t + U_z h_{t-1} + b_z) \tag{6}$$

Reset gate (equation 7);

$$r_t = \sigma(W_r x_t + U_r h_{t-1} + b_r) \tag{7}$$

Updating the hidden state (equation 8);

$$h_t = (1 - z_t)(h_{t-1}) + z_t \tanh(W_h x_t + U_h(r_t h_{t-1}) + b_h) \tag{8}$$

2.6. Model Parameters and Training Process

In this study, a comprehensive modeling process was conducted using both deep learning-based and traditional machine learning algorithms to predict natural gas consumption at the district level based on time series data. On the deep learning side, Simple RNN, LSTM, GRU, BI-LSTM, and BI-GRU architectures were trained and evaluated to capture both short-term fluctuations and long-term temporal dependencies in consumption patterns.

Consumption data covering the period 2015-2023 was used as input for all deep learning models. Each model was structured with a recurrent layer containing 192 hidden neurons, followed by a dropout layer with a rate of 0.35 and a single-neuron output layer. These hyperparameter values were determined based on pre-training experiments and practices commonly adopted in the time series forecasting literature, aiming to balance model capacity and generalization performance while avoiding excessive model complexity.

The hyperbolic tangent (tanh) activation function was used in all recurrent layers. The training process was limited to a maximum of 60 epochs, and early stopping was applied using the EarlyStopping mechanism to reduce the risk of overfitting and increase the robustness of the model.

A Grid Search approach was used to further improve model performance and determine the optimal hyperparameter configurations. Within this framework, fundamental parameters such as learning rate, dropout rate, epoch count, and optimization strategies were systematically examined. The best-performing configurations were selected based on validation set performance, thereby improving the models' generalization ability.

Instead of aggressive, model-specific hyperparameter optimization, consistent parameter ranges were applied across architectures where appropriate. This strategy enables a fair and transparent comparison, allowing observed performance differences to be attributed primarily to the natural characteristics of the model architectures rather than extensive parameter tuning.

Additionally, to prevent data leakage and preserve the temporal structure of the time series, the dataset was divided chronologically rather than randomly. The data was split into training, validation, and test sets in an 80%-10%-10% ratio; the oldest observations were used for training, the next portion for validation, and the newest observations for testing. This time-ordered splitting strategy ensures that future information is not used during model training and provides a realistic assessment of prediction performance.

Table 2 summarizes the ranges evaluated by the Grid Search process and the optimal values obtained.

Table 2. Grid search hyperparameter optimization results

Parameters	Range	Best Value
Learning rate	0.001–0.01	0.005
Dropout	0.2-0.5	0.35
Epoch	30-100	60

Forecasting performance was evaluated using Mean Absolute Percentage Error (MAPE) and the coefficient of determination (R^2). MAPE was selected due to its interpretability and widespread use in energy consumption forecasting, as it expresses prediction error in relative percentage terms. R^2 was included to assess the model’s ability to explain the variance in consumption patterns. Although MAPE is known to be sensitive during low-consumption periods, its use in combination with R^2 provides a balanced evaluation by capturing both point-wise prediction accuracy and overall trend consistency (Armstrong and Collopy, 1992; Chicco and Jurman, 2020).

$$MAPE = \frac{100}{n} \sum_{t-1}^{n} \left| \frac{y_t - \hat{Y}}{y_t} \right| \quad (9)$$

$$R^2 = 1 - \frac{\sum_{t-1}^{n} (y_t - \hat{Y}_t)^2}{\sum_{t}^{n} (y_t - \bar{Y})^2} \quad (10)$$

Thanks to the metrics given in Equations 9 and 10, not only the accuracy but also the statistical explanatory power of the models were comparatively analyzed. In addition to the deep learning models, five different traditional machine learning methods (Linear Regression, Ridge Regression, Random Forest, Gradient Boosting and Support Vector Regression) were also trained and evaluated on the same dataset. Although the details of these models are not included in detail, they are included in the comparative overall performance analysis. Thus, the relative performance of both classical and deep learning-based methods in time series based forecasting problems is presented in a holistic manner.

3. Results

In this study, five deep learning models (LSTM, GRU, RNN, BI-LSTM and BI-GRU) and five classical machine learning models (Linear Regression, Ridge, SVR, Random Forest, Gradient Boosting) were trained and tested for time series based natural gas consumption forecasting. Performance evaluation is based on MAPE (%) and R^2 metrics.

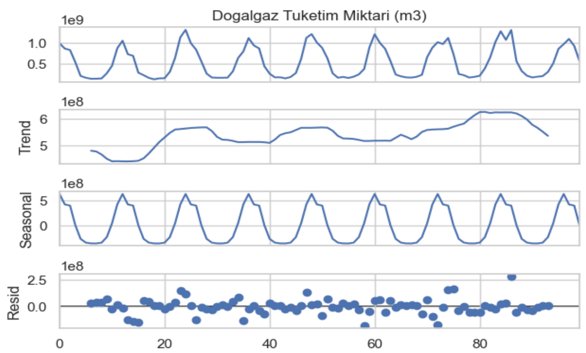

Figure 5. Decomposition into seasons.

Figure 5 presents a decomposition of natural gas consumption into its seasonal components. It is observed that consumption increases significantly in the winter months and decreases in the summer months, indicating that seasonality should be taken into account in the modeling process.

In addition to deep learning–based recurrent architectures, several classical machine learning models, including Random Forest, were included as baseline methods to provide a broader performance comparison. These models were not the primary focus of the study but were used to benchmark the proposed BI-GRU model against commonly applied traditional approaches in energy consumption forecasting. For consistency, the classical models were trained using standard parameter settings commonly reported in the literature.

Table 3. Model performance comparison

Model	Type	R^2	MAPE (%)
LSTM	DL	0.8814	19.60
RNN	DL	0.8525	17.03
GRU	DL	0.8172	24.45
BI-LSTM	DL	0.8840	19.38
BI-GRU	DL	0.8902	12.98
Liner Regression	ML	0.3401	141.13
Ridge	ML	0.3375	148.83
Random Forest	ML	0.9063	56.17
Gradient Boosting	ML	0.8614	119.83
SVR	ML	0.7498	176.35

When Table 3 and Figure 6-7 are analyzed, deep learning models showed a significantly superior prediction performance compared to classical machine learning models. In particular, the BI-GRU model achieved the lowest error rate with a MAPE value of 12.98% and stood out as the most successful model by explaining 89% of the variance in the data with an R^2 value of 0.8902. The RNN model has the second lowest error rate with a MAPE value of 17.03%. The LSTM and BI-LSTM models performed strongly with similar R^2 (≈0.88) and MAPE (≈19.5%) values.

On the other hand, classical machine learning algorithms exhibited lower accuracy. Linear Regression, Ridge and SVR models performed poorly with both very low R^2 (in the range 0.34-0.75) and high MAPE (141-176%). The Random Forest model produced a high R^2 value of 0.9063 but made highly inaccurate predictions with a MAPE of 56.17%. This suggests that the model suffers from instability in individual predictions when capturing patterns in the data. This indicates that the model was able to capture the overall variance and long-term trend in the data, but produced large relative errors during certain periods, particularly in months with low consumption levels. This behavior demonstrates sensitivity to outliers and limited robustness in point estimates. This result emphasizes that a high R^2 value alone does not imply superior forecasting performance and that complementary error measures such as MAPE are necessary for reliable forecasting.

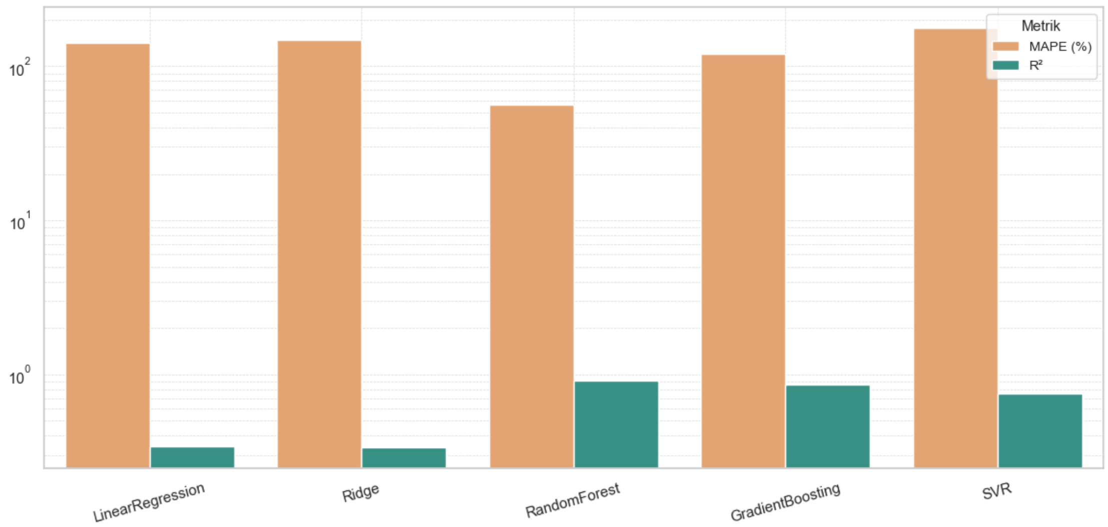

Figure 6. Performance comparison of machine learning models.

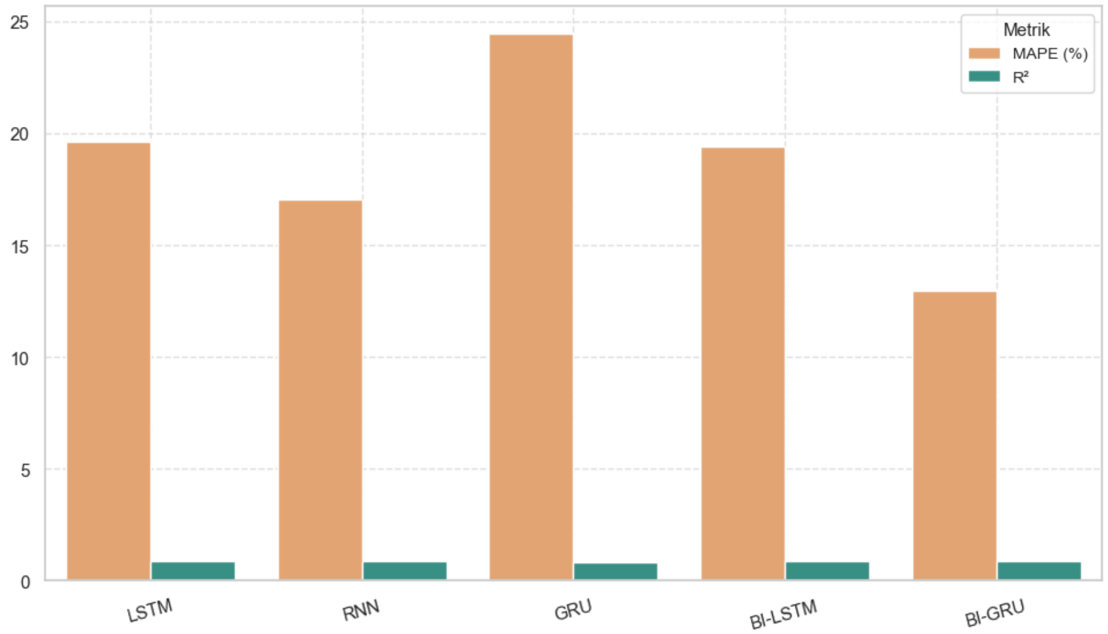

Figure 7. Performance comparison of deep learning models.

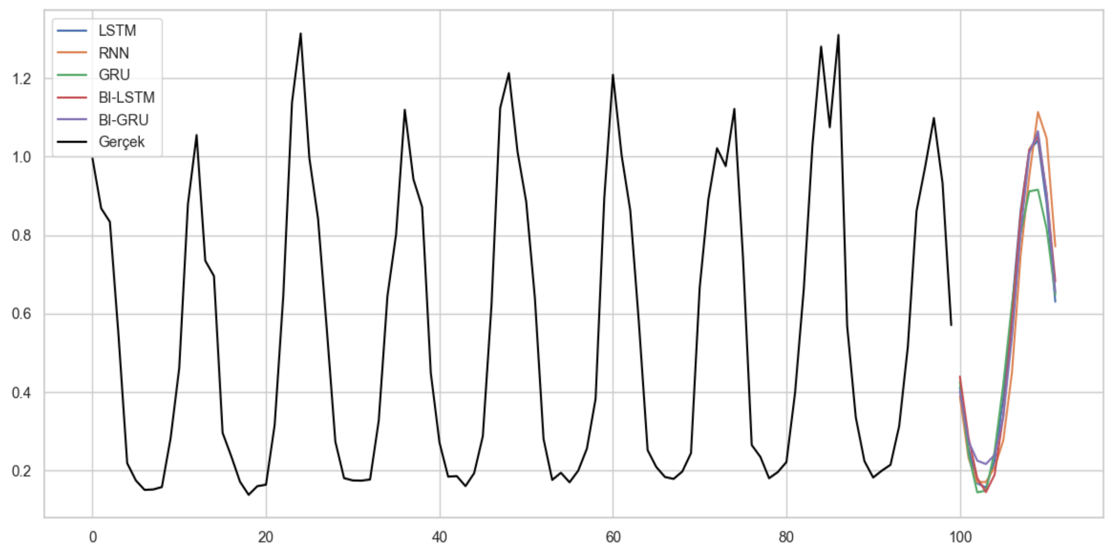

Figure 8. 12-Month forecast of deep learning models.

When examining Figure 8, BI-GRU and BI-LSTM successfully modeled consumption increases in winter months and decreases in summer months by more accurately capturing the trend of actual values. BI-GRU, in particular, was able to capture sudden consumption fluctuations (e.g., peaks due to severe weather conditions) with lower error. Models such as GRU and RNN also showed reasonable fit; however, LSTM-based models produced higher deviations in some months. Bidirectional structures (BI-GRU), by accounting not only for past information but also for future dependencies, better represented both trends and short-term anomalies within the time series.

4. Discussion and Conclusion

In this study, the forecasting performances of both deep learning and classical machine learning algorithms are compared using district-based monthly natural gas consumption data in Istanbul. Deep learning methods are much more successful than classical models in learning patterns in time-dependent data. In particular, the BI-GRU model stood out with its low error rate and high explanatory power, and was better able to represent short-term changes.

Among the classical machine learning methods, only the Random Forest model stood out with its high R^2 value, but its predictive reliability was poor due to its high MAPE value. This suggests that the model is too sensitive to the training data rather than its generalization capacity.

In conclusion, deep learning models - especially those with bidirectional structure (BI-GRU, BI-LSTM) - are more successful in capturing seasonality and short-term variations in time series data. Therefore, the use of deep learning methods in energy consumption forecasting in regions with complex consumption profiles such as metropolitan areas will provide decision makers with more reliable and realistic forecasts.

In addition to prediction accuracy, the BI-GRU model also offers practical advantages in terms of real-time applicability. Compared to more complex deep learning architectures, GRU-based models are known to require fewer parameters and lower computational costs, making them suitable for real-time prediction applications. The bidirectional structure further enhances the model's ability to capture temporal dependencies and short-term fluctuations in consumption patterns. Therefore, the proposed BI-GRU-based framework can be effectively integrated into smart city infrastructures, IoT-based energy management systems, and decision support platforms for urban energy planning. Such real-time integration can support proactive energy distribution, early detection of demand spikes, and more efficient infrastructure planning.

Author Contributions

The percentages of the author' contributions are presented below. The author reviewed and approved the final version of the manuscript.

	H.C.
C	100
D	100
S	100
DCP	100
DAI	100
L	100
W	100
CR	100
SR	100
PM	100
FA	100

C= concept, D= design, S= supervision, DCP= data collection and/or processing, DAI= data analysis and/or interpretation, L= literature search, W= writing, CR= critical review, SR= submission and revision, PM= project management, FA= funding acquisition.

Conflict of Interest

The author declared that there is no conflict of interest.

Ethical Consideration

Ethics committee approval was not required for this study because of there was no study on animals or humans.

Acknowledgements

I would like to thank the Istanbul Metropolitan Municipality for providing the vital data for this research.

References

Armstrong, J. S., & Collopy, F. (1992). Error measures for generalizing about forecasting methods: Empirical comparisons. *International Journal of Forecasting*, *8*(1), 69–80.

Arslantürk, V., & Şirin, B. T. (2025). Machine learning approaches for energy management in public buildings. *Black Sea Journal of Engineering and Science*, *8*(5), 1415–1428.

Azadeh, A., Zarrin, M., Rahdar Beik, H., & Aliheidari Bioki, T. (2015). A neuro-fuzzy algorithm for improved gas consumption forecasting with economic, environmental and IT/IS indicators. *Journal of Petroleum Science and Engineering*, *133*, 716–739. https://doi.org/10.1016/j.petrol.2015.07.002

Başarslan, M. S. (2025). M-C&M-BL: A novel classification model for brain tumor classification: Multi-CNN and multi-BiLSTM. *The Journal of Supercomputing*, *81*, Article 502. https://doi.org/10.1007/s11227-025-06964-x

Beyca, O. F., Cayir Ervural, B., Tatoglu, E., Ozuyar, P. G., & Zaim, S. (2019). Using machine learning tools for forecasting natural gas consumption in the province of Istanbul. *Energy Economics*, *80*, 937–949. https://doi.org/10.1016/j.eneco.2019.03.006

Chicco, D., & Jurman, G. (2020). The advantages of the

Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics*, *21*(1), Article 6.

Demirel, Ö. F., Zaim, S., Çalışkan, A., & Özuyar, P. (2012). Forecasting natural gas consumption in İstanbul using neural networks and multivariate time series methods. *Turkish Journal of Electrical Engineering and Computer Sciences*, *20*(5), Article 4. https://doi.org/10.3906/elk-1101-1029

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.

Graves, A., & Schmidhuber, J. (2005). Framewise phoneme classification with bidirectional LSTM networks. *Proceedings of the IEEE International Joint Conference on Neural Networks (IJCNN)*, 2047–2052.

Haksevenler, G. (2025). Towards carbon neutral cities: An insight for Istanbul, Türkiye. *Sustainable Development*, *33*(1), 717–732. https://doi.org/10.1002/sd.3154

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1735–1780.

Karadağ, İ., & Sağtaş, K. (2025). Energy consumption forecasting with artificial intelligence models. *Black Sea Journal of Engineering and Science*, *8*(6), 1780–1793.

Kizilaslan, R., & Karlik, B. (2008). Comparison neural networks models for short term forecasting of natural gas consumption in Istanbul. *Proceedings of the 1st International Conference on Applied Digital Information and Web Technologies (ICADIWT)*, 448–453. https://doi.org/10.1109/ICADIWT.2008.4664390

Kong, W., Dong, Z. Y., Jia, Y., Hill, D. J., Xu, Y., & Zhang, Y. (2019). Short-term residential load forecasting based on LSTM recurrent neural network. *IEEE Transactions on Smart Grid*, *10*(1), 841–851. https://doi.org/10.1109/TSG.2017.2753802

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444.

Marino, D. L., Amarasinghe, K., & Manic, M. (2016). Building energy load forecasting using deep neural networks. *Proceedings of the IEEE IECON 42nd Annual Conference of the Industrial Electronics Society*, 7046–7051. https://doi.org/10.1109/IECON.2016.7793413

Potočnik, P., Govekar, E., & Grabec, I. (2007). *Short-term natural gas consumption forecasting*. Proceedings of Applied Simulation and Modelling, Palma De Mallorca, Spain.

Schuster, M., & Paliwal, K. K. (1997). Bidirectional recurrent neural networks. *IEEE Transactions on Signal Processing*, *45*(11), 2673–2681.

Stambouli, A. B. (2011). Fuel cells: The expectations for an environmental-friendly and sustainable source of energy. *Renewable and Sustainable Energy Reviews*, *15*(9), 4507–4520. https://doi.org/10.1016/j.rser.2011.07.100

Yang, Z., Yang, D., Dyer, C., He, X., Smola, A., & Hovy, E. (2016). *Hierarchical attention networks for document classification*. Proceedings of the North American Chapter of the Association for Computational Linguistics (NAACL).

Zanella, A., Bui, N., Castellani, A., Vangelista, L., & Zorzi, M. (2014). Internet of Things for Smart Cities. *IEEE Internet of Things Journal*, *1*(1), 22–32. https://doi.org/10.1109/JIOT.2014.2306328