Explainable Dual-Attention Encoder–Decoder Model for Natural Gas Consumption Forecasting Using Algerian Hourly Data

Randa Ladlani[1][*], Samiha Ait Taleb [1,2] and Abderrazak Sebaa[1,2]

[1] Laboratoire LITAN, École supérieure en Sciences et Technologies de l'Informatique et du Numérique, RN 75, Amizour, 06300 Bejaia, Algeria
[2] LIMED Laboratory, Computer Science Department, University of Bejaia, 06000 Bejaia, Algeria
*ladlani@estin.dz, aittaleb@estin.dz, sebaa@estin.dz

Abstract. Natural gas consumption forecasting supports efficient energy management and resource planning in industrial environments. The proposed architecture combines an encoder–decoder structure with dual attention mechanisms — temporal and feature-level — alongside cyclical encodings and moving-average representations to improve forecasting accuracy and stability. Evaluation on Algerian hourly natural gas data yields Test MAE = 0.0255 and $R^2 = 0.9740$, outperforming classical machine learning and deep learning baselines by up to 38%. Cross-domain validation on the GEFCom2014 electricity load benchmark confirms generalizability ($R^2 = 0.9817$, 67% MAE reduction over XGBoost). SHAP analysis quantifies feature contributions, identifying historical consumption and temporal encodings as the dominant predictors, with meteorological variables providing secondary refinement.

Keywords: Natural Gas Consumption Forecasting, Dual Attention Mechanism, Encoder-Decoder Architecture, Explainable AI, SHAP.

1 Introduction

Natural gas consumption forecasting underpins operational decisions in energy management, including supply-demand balancing, infrastructure planning, and cost control for energy-intensive economies [1], [2]. Demand prediction remains difficult due to nonlinear interactions among consumption patterns, meteorological conditions, and socioeconomic variables; empirical studies confirm that excluding weather inputs degrades forecast accuracy, making meteorological integration a functional requirement rather than an optional enhancement [3], [4].

Encoder-decoder architectures and attention-based models have shown strong predictive capability for sequential energy data, capturing long-range temporal dependencies and dynamic feature interactions. These architectures, however, typically operate as black-box systems, restricting adoption in operational environments where model transparency must accompany predictive accuracy. The integration of explainable AI

© The Author(s) 2026
D. Agti et al. (eds.), *Proceedings of the International Conference on Artificial Intelligence Applications in Business Administration in MENA Region (ICAIABA 2026)*, Advances in Economics, Business and Management Research 393,
https://doi.org/10.2991/978-94-6239-711-8_32

methods has therefore become a technical requirement for deployment in regulated or operationally critical forecasting contexts.

Algeria constitutes a particularly relevant case on three grounds: it is one of Africa's principal natural gas producers and a primary supplier to European markets [10]; its consumption is shaped by climatological variability across three structurally distinct zones — the Mediterranean coastal strip, the semi-arid interior plateau, and the Saharan south — producing demand dynamics that differ systematically from those in temperate European markets; and no prior study has applied a deep learning architecture combining dual attention and post-hoc explainability to this setting, a gap documented in the Related Work section below.

This study proposes an enhanced encoder-decoder architecture incorporating dual attention mechanisms to capture both temporal dependencies and feature-level interactions. SHapley Additive exPlanations (SHAP) are applied to quantify feature contributions across forecasting horizons, providing interpretable analysis of model predictions. The proposed framework achieves improved predictive accuracy while satisfying the transparency requirements of operational forecasting contexts.

This paper makes the following contributions:

- Proposes an enhanced encoder-decoder forecasting architecture integrating dual attention mechanisms for improved short-term natural gas consumption prediction.
- Develops a feature engineering framework incorporating cyclical temporal encodings, moving averages, and meteorological variables with differential scaling.
- Applies SHAP-based interpretability to quantify temporal and feature-level contributions, providing domain-specific analysis of Algerian gas demand.
- Demonstrates superior predictive performance relative to baseline deep learning and machine learning models on real-world Algerian data, and validates cross-domain generalisability on the GEFCom2014 electricity load benchmark ($R^2 = 0.9817$, 67% MAE reduction over XGBoost).

2 Related Work

Statistical approaches constitute the foundational layer of natural gas demand forecasting. Moving average and exponential smoothing models provide computational efficiency but fail to capture nonlinear consumption dynamics [5], [6]. Autoregressive models such as ARMA, ARIMA, SARIMA, and SARIMAX extend this baseline by incorporating seasonal components, yielding measurable gains in monthly gas demand prediction [7], [9]. Gaussian Process Regression has been applied for probabilistic load prediction [10]. These models are structurally limited in representing nonlinear variable interactions, motivating the adoption of data-driven alternatives.

LSTM and GRU networks capture temporal dependencies in consumption sequences [11], [12]. Ensemble learning approaches have also been applied to natural gas consumption forecasting, demonstrating competitive performance through model combination strategies [24]. Encoder-decoder frameworks produce strong results in multi-step buildi.

building energy forecasting when combined with attention mechanisms [17]. Decomposition-based models reduce non-stationarity effects [13], [14], and clustering methods segment consumers by temporal and socioeconomic characteristics [15], [16].

Attention-based sequence-to-sequence architectures assign adaptive weights to historical time steps [17], [4]; Transformer-based models extend this to longer horizons. The majority of these architectures restrict attention to the temporal dimension and do not model feature-level interactions.

Explainable AI techniques have been systematically reviewed for energy and power systems, confirming their growing role in operational forecasting [21]. Among these, SHAP has been established as a model-agnostic framework for quantifying feature contributions in energy forecasting [18], [19]; a 2023 comparative study confirmed its effectiveness over Grad-CAM [18], and recent work confirms that distributional shift is a documented operational risk in single-year trained models [20]. The application of SHAP to multi-step architectures incorporating both temporal and feature-level attention remains largely unexplored.

The surveyed literature confirms that existing models do not jointly address temporal attention, feature-level attention, and interpretability within a single architecture. This work addresses this gap by proposing a dual-attention encoder-decoder with SHAP-based attribution, targeting both predictive accuracy and transparent feature contribution analysis.

3 Methodology

This section presents the proposed pipeline for short-term natural gas consumption forecasting, integrating data preprocessing, feature engineering, a dual-attention encoder–decoder architecture, and interpretability analysis.

3.1 Datasets and Preprocessing

Dataset 1 — Algerian Hourly Natural Gas (2014). The primary dataset consists of hourly natural gas consumption records for Algeria during 2014, covering residential and industrial sectors [10]. Meteorological covariates were sourced from NASA POWER. A sliding window strategy was applied, using 24 hours of historical observations as input to predict the subsequent 12 hours of consumption. Sequences were partitioned chronologically into training (70%), validation (15%), and test (15%) sets to prevent data leakage and preserve temporal ordering. Input features were standardized via z-score normalization; the target variable was scaled using Min–Max normalization, with all scaling parameters derived exclusively from the training partition. The dataset spans a single calendar year (8,760 observations), capturing daily, weekly, and seasonal patterns but excluding inter-annual variability; this constraint is addressed in Section 5.

Dataset 2 — GEFCom2014 Electricity Load. Cross-domain generalisability was assessed using the GEFCom2014 load forecasting benchmark [23], comprising hourly electricity load records for a US utility zone (Zone 1) spanning 2006–2010, with 41,377 observations and 25 anonymous weather variables (w1–w25) as covariates. The iden-

tical sliding window configuration, chronological split, and feature engineering pipeline defined for Dataset 1 were applied without modification, isolating architectural behavior as the variable of interest across energy carriers and geographic contexts.
Table 1 summarises the input features used in the model. As illustrated in Figure 1, the full forecasting pipeline covers data sourcing, cleaning, feature engineering, sequence creation, scaling, model training, hyperparameter tuning, evaluation, and SHAP-based interpretability analysis.

Table 1. Input features used in the forecasting model.

Category	Features
Meteorological	Temperature, Wet Bulb Temperature, Wind Speed, Humidity
Temporal	Hour, Day of Week, Day of Month, Month
Cyclical Enc.	HourSin, DayOfWeekSin, DayOfMonthSin, MonthSin
Trend	Moving averages
Calendar	Holiday indicator

3.2 Proposed Forecasting Architecture

The proposed framework adopts an encoder–decoder sequence learning architecture designed to model nonlinear temporal dependencies and feature interactions in natural gas consumption data, integrating stacked Bidirectional Long Short-Term Memory (BiLSTM) layers with dual attention mechanisms.
The dual-attention framework consists of two complementary mechanisms. Temporal attention refers to a multi-head self-attention module in the encoder that assigns adaptive importance weights across historical time steps, enabling selective focus on the most informative periods in the input sequence. Feature-level attention, implemented as context attention in the decoder, learns differential importance weights across input feature dimensions at each decoding step, suppressing less informative features dynamically rather than treating all inputs uniformly. Together, these mechanisms identify both when historical patterns are relevant and which features contribute most to future consumption.
The encoder extracts temporal representations from historical input sequences using stacked BiLSTM layers. Bidirectional processing captures forward and backward temporal dependencies, improving the model's ability to represent long-term consumption dynamics. BiLSTM was selected to exploit symmetric consumption patterns such as weekday-weekend transitions within the 24-hour lookback window; the backward pass operates exclusively on historical data and does not access future values, ensuring strict causality during inference. Given an input sequence $X = \{x_1, x_2, \ldots, x_T\}$, where T denotes the observation window, the BiLSTM encoder produces hidden representations by combining forward and backward states:

$$h_t = [\overrightarrow{h_t}; \overleftarrow{h_t}]$$

The encoded sequence is refined using a temporal multi-head self-attention mechanism; the resulting attention outputs are combined with BiLSTM representations to produce contextual embeddings summarizing temporal and feature-level information. The decoder generates multi-step forecasts using an LSTM-based sequential prediction strategy initialized with the encoder's final hidden and cell states. Teacher forcing and dropout regularization are applied during training to improve convergence stability and reduce overfitting. The decoder output passes through a fully connected linear layer mapping latent representations to final consumption predictions.

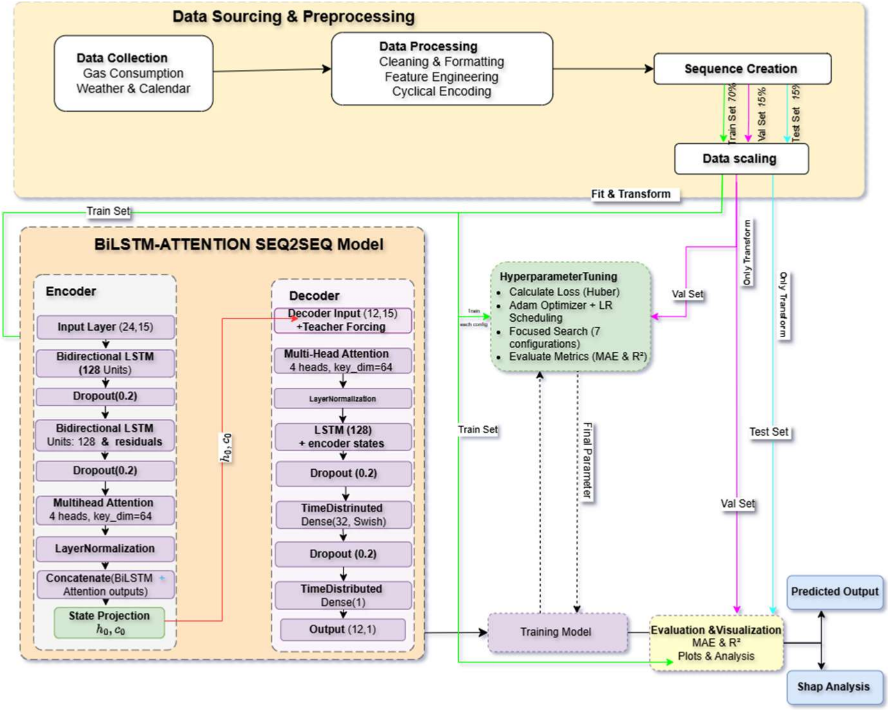


Fig. 1. Proposed Forecasting Architecture

3.3 Training and Evaluation Setup

The model was trained using the Adam optimizer with an initial learning rate of 0.001 and exponential decay scheduling. Huber loss was applied for robustness against consumption spikes. Regularization includes dropout, L2 weight regularization, and early stopping. Performance was evaluated using Mean Absolute Error (MAE) and the coefficient of determination (R^2):

$$\mathrm{MAE} = {}^{1}/_{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|\,, R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2} \qquad (1)$$

3.4 Hyperparameter Optimization

Hyperparameter optimization was conducted using random search over the space defined in Table 2. Increasing LSTM units to 128, switching the optimizer to Adam, and reducing regularization coefficients produced the final configuration that balances generalization and model expressiveness.

Table 2. Hyperparameter search space and optimized configuration.

Parameter	Search Range	Original	Optimized
Activation	tanh, relu, elu, leaky relu	leaky relu	leaky relu
Dropout rate	0.0–0.3	0.3	0.2
Optimizer	Adam, SGD, RMSprop	SGD	Adam
LSTM Units	64, 128, 256	16	128
Batch Size	16–128	64	64
Learning rate	0.0001–0.01	0.001	0.001
Attention Heads	2, 3, 4	3	4
Key Dimensions	16, 32, 64	64	64
Dense Units	16, 32, 64	32	32
L2 Regularization	0.0001–0.01	0.01	0.001

3.5 Implementation and Interpretability Setup

The framework was implemented in TensorFlow 2.17.0 with the Keras API under Python 3.11.7, executed on a standard laptop (Intel CPU, 16 GB RAM, Windows 10). Training used mini-batch gradient descent with a batch size of 64, a maximum of 80 epochs, ReduceLROnPlateau scheduling, and early stopping with patience set to 10. The model converged at epoch 54 in approximately 9.5 minutes, confirming that the architecture trains within resource-constrained hardware without specialized accelerators.

SHAP values were computed using KernelExplainer with 100 randomly selected background training samples, producing stable feature contribution estimates within standard laptop memory constraints. The implementation code is publicly available at the project repository. [1] the raw Algerian consumption records were provided by Dr. Oussama Laib under a restricted arrangement and cannot be redistributed; the original source is cited in [10]. Meteorological variables are freely accessible via the NASA POWER portal at https://power.larc.nasa.gov/.

[1]https://github.com/Randaladlani99/Explainable-Dual-Attention-BiLSTM-E coder-Decoder-for-Natural-Gas-Consumption-Forecasting/tree/main

4 Results and Discussion

On the Algerian dataset, the proposed model achieved Train MAE = 0.0159 (R^2 = 0.9915), Validation MAE = 0.0204 (R^2 = 0.9715), and Test MAE = 0.0255 (R^2 = 0.9740). The narrow gap between training and test metrics confirms that dropout, L2 regularization, and early stopping prevented overfitting. Figure 2 shows the training

1

and validation curves; the smooth, monotonic decline of both MAE and loss confirms stable convergence under the adaptive learning rate schedule.

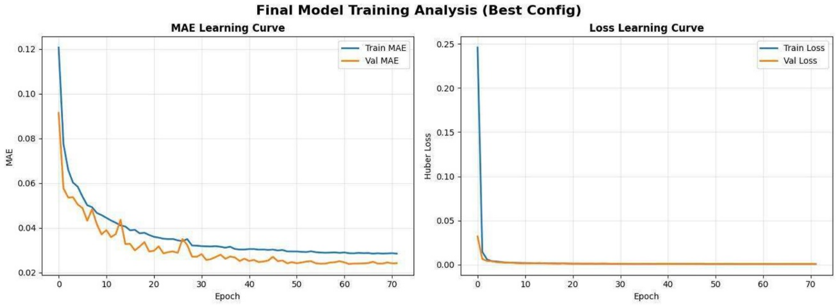


Fig 2. Training and validation loss and MAE curves on the Algerian dataset.

Table 3 reports test-set performance across all baselines on the Algerian dataset. The proposed model records the lowest error and highest R^2 among all evaluated architectures, reducing MAE by 38% relative to XGBoost and by 78% relative to standard LSTM. The performance advantage over unidirectional baselines is attributable to the backward temporal pass, which captures consumption symmetry across weekday-weekend transitions that a forward-only pass cannot represent within the 24-hour look-back window. Prediction error increases at a sub-linear rate beyond forecasting step 6, and temperature features account for a disproportionate share of error reduction at longer horizons.

Table 3. Test set performance on the Algerian dataset: proposed model vs. baseline approaches

Metric	Prop.	XGBoost	Grad.Boost	Ridge	RF	Mov.Avg	Naive	GRU	Seasonal	Seq2Seq
Test MAE	**0.0255**	0.0411	0.0415	0.0454	0.0537	0.0731	0.0734	0.0987	0.1042	0.1077
Test R^2	**0.9740**	0.9200	0.9191	0.8985	0.8692	0.7594	0.7791	0.6555	0.4881	0.5839

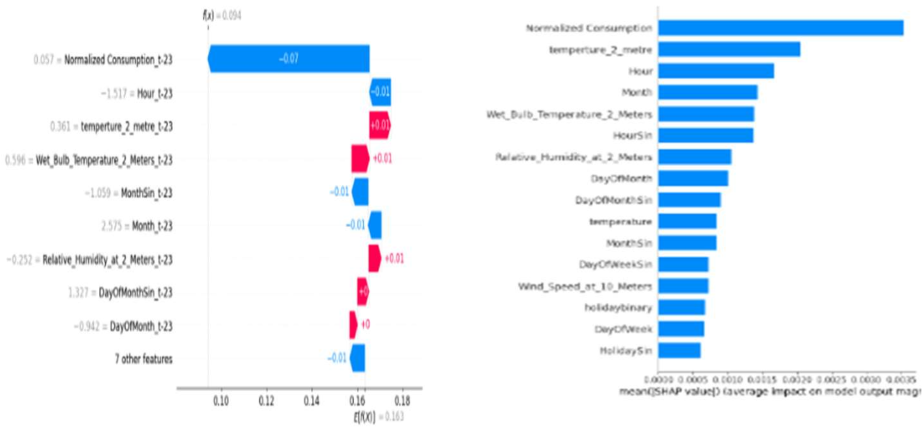
Cross-Domain Generalisability. To determine whether accuracy gains reflect genuine architectural merit rather than dataset-specific tuning, the proposed model was evaluated on the GEFCom2014 electricity load benchmark using the identical hyperparameter configuration and preprocessing pipeline, without domain-specific adaptation. Table 4 reports test-set performance across all evaluated models. The proposed architecture achieves MAE = 3.25, RMSE = 7.26, R^2 = 0.9817, and MAPE = 2.03%, ranking first across all four metrics. The MAE reduction over XGBoost reaches 67% on GEFCom2014 versus 38% on the Algerian dataset, indicating that dual attention produces stronger relative gains on larger, multi-year benchmarks where long-range temporal patterns are more pronounced. The Seq2Seq Basic encoder-decoder (MAE = 4.33, R^2 = 0.9585) trails the proposed model by 25% in MAE despite sharing the encoder-decoder structure, isolating the dual attention components as the source of the performance differential.

Taken together, the two evaluations establish that the proposed architecture generalises across energy carriers (natural gas vs. electricity), geographies (Algeria vs. United States), and dataset scales (8,760 vs. 41,377 records).

Table 4. Test set performance on GEFCom2014 (electricity load benchmark): proposed model vs. baseline approaches .

Metric	Proposed	XGBoost	Grad.Boost	Ridge	RF	GRU	Seq2Seq	LSTM	Attn-LSTM	BiLSTM
MAE	**3.25**	9.88	10.23	12.27	10.93	9.96	4.33	10.94	10.53	10.62
RMSE	**7.26**	15.22	15.42	17.77	16.22	14.63	10.94	15.84	15.40	15.54
R^2	**0.9817**	0.9196	0.9174	0.8904	0.9086	0.9257	0.9585	0.9129	0.9177	0.9161
MAPE	**2.03**	5.83	6.11	7.68	6.53	6.22	2.71	6.67	6.47	6.56

SHAP Interpretability Analysis. SHAP was applied to quantify feature contributions at local and global levels. Figure 3 presents the SHAP summary plot and global feature importance ranking



(a) SHAP summary plot (b) Global feature importance

Fig. 3. SHAP summary plot and global feature importance ranking.

Normalized consumption ranks first because recent demand directly governs short-term forecasts. Cyclical temporal encodings (Hour, MonthSin, DayOfMonthSin) capture daily and seasonal periodicity. Temperature and wet bulb temperature rank among the leading meteorological predictors of heating demand. Humidity and wind speed exert weaker effects: humidity's influence is partially absorbed by the temperature features already present, and wind speed's contribution to building heat loss is secondary to the thermal-mass dynamics captured by temperature at hourly resolution. Calendar and holiday indicators contribute minimally because Algerian gas demand follows routine daily and weekly cycles rather than isolated calendar events.Quarterly R^2 values vary across seasons (Q1: 0.9656, Q2: 0.6260, Q3: 0.9169, Q4: 0.9590), confirming that temporal drift occurs within a single calendar year. Prediction at extreme demand levels (>0.9 normalized) shows measurable degradation (MAE = 0.0434, $R^2 = -10.73$), confirming out-of-distribution vulnerability. These findings motivate the operational safeguards outlined in Section 5.

5 Conclusion and Future Work

The proposed dual-attention encoder-decoder achieves a 38% MAE reduction over XGBoost and a 78% reduction over standard LSTM on the Algerian natural gas dataset (Test MAE = 0.0255, $R^2 = 0.9740$). Cross-domain evaluation on GEFCom2014 extends this result to $R^2 = 0.9817$ and a 67% MAE reduction over XGBoost using the identical architecture and pipeline, with stronger relative gains on the larger multi-year benchmark confirming that dual attention extracts proportionally more information from longer records. SHAP analysis identifies consumption history and temporal encodings as the dominant predictors; the full framework converged in 9.5 minutes on a standard laptop, and SHAP attributions provide decision-level transparency for trust-critical scheduling in grid operations and energy planning.The primary scope limitation is the use of a single year of Algerian training data. The 35% seasonal R^2 variation and degradation at extreme demand levels confirm that temporal drift is present within the available data. Three operational safeguards are proposed for deployment: (i) seasonal retraining every three months on a rolling window; (ii) automated drift detection using ADWIN or Page-Hinkley tests to trigger re-calibration when distributional shift is confirmed; and (iii) uncertainty quantification for extreme-demand forecasts to flag predictions outside the training distribution. Future work will extend the framework to multi-year Algerian datasets as records become accessible, incorporate additional socioeconomic indicators, and evaluate the architecture on additional international benchmarks spanning diverse consumption profiles and climate zones.

Acknowledgments. The authors thank Dr. Oussama Laib for providing the Algerian natural gas consumption dataset used in this study. The contributions of the co-authors, whose guidance and expertise were instrumental in shaping this work, are sincerely appreciated.

References

1. Liu, J. et al.: Natural gas consumption forecasting: A discussion on forecasting history and future challenges. J. Natural Gas Sci. Eng. 90, 103930 (2021).
2. Soldo, B.: Forecasting natural gas consumption. Applied Energy 92, 26–37 (2011).
3. Panek, W., Włodek, T.: Natural Gas Consumption Forecasting Based on Variability of External Meteorological Factors Using Machine Learning Algorithms. Energies 15(1), 348 (2022).
4. Liu, J. et al.: An Enhancement Method Based on LSTM Neural Network for Short-Term Natural Gas Consumption Forecasting. Energies 16(3), 1295 (2023).
5. Akpinar, M., Yumusak, N.: Day-ahead natural gas forecasting using nonseasonal exponential smoothing methods. In: Proc. IEEE EEEIC (2017).
6. Yongmei, Z.: Forecasting the yield of natural gas in China. Energy Sources B 12(3), 207–210 (2017).
7. Hošovský, A. et al.: Comparative study of week-ahead forecasting of daily gas consumption using regression ARMA/SARMA and wavelet neural network models. J. Building Engineering 34, 101955 (2020).
8. Villacorta Cardoso, C.A., Lima Cruz, G.: Forecasting Natural Gas Consumption using ARIMA Models and ANNs. IEEE Latin America Trans. 14(5), 2233–2238 (2016).
9. Manigandan, P. et al.: Forecasting Natural Gas Production and Consumption in United States — Evidence from SARIMA and SARIMAX Models. Energies 14(19), 6021 (2021).
10. Laib, O. et al.: A Gaussian Process Regression for Natural Gas Consumption Prediction Based on Time Series Data. In: Proc. ICIF (2018).
11. Peng, S. et al.: Daily natural gas load forecasting based on LSTM, local mean decomposition, and wavelet threshold denoising. J. Natural Gas Sci. Eng. 95, 104175 (2021).
12. Manowska, A. et al.: Forecasting of Natural Gas Consumption in Poland Based on ARIMA-LSTM Hybrid Model. Energies 14(24), 8597 (2021).
13. Zhang, L. et al.: Multi-Step Ahead Natural Gas Consumption Forecasting Based on a Hybrid Model. Energies 15(19), 7437 (2022).
14. Jiang, S. et al.: Predicting monthly natural gas consumption in the USA using a hybrid two-layer decomposition model. Environ. Sci. Pollut. Res. 30(14), 40799–40824 (2023).
15. Fernandes, M. et al.: Segmentation of Residential Gas Consumers Using Clustering Analysis. Energies 10(12), 2047 (2017).
16. Dogan, O.: Segmentation Analysis of Companies' Natural Gas Consumption by Soft Clustering. In: Proc. ISKE (2020).
17. Gao, F. et al.: Forecasting residential electricity consumption using a hybrid machine learning model with online search data. Applied Energy 300, 117393 (2021).
18. Van Zyl, C. et al.: Harnessing eXplainable AI for feature selection in time series energy forecasting: A comparative analysis of Grad-CAM and SHAP. Applied Energy 353, 122079 (2023).
19. Liu, C., Yang, W.: Transformer fault diagnosis using machine learning: a method combining SHAP feature selection and intelligent optimization of LGBM. Energy Informatics 8(1) (2025).
20. Ye, L. et al.: Concept drift adaptation strategies for energy consumption forecasting: a survey. Energy and AI 16, 100361 (2024). https://doi.org/10.1016/j.egyai.2024.100361.
21. Machlev, R. et al.: eXplainable Artificial Intelligence (XAI) techniques for energy and power systems. Energy and AI 9, 100169 (2023). https://doi.org/10.1016/j.egyai.2022.100169.

22. Li, C. et al.: Transformer-based model for industrial natural gas demand forecasting. Applied Energy 357, 122513 (2024). https://doi.org/10.1016/j.apenergy.2024.122513.
23. Hong, T. et al.: Probabilistic energy forecasting: Global Energy Forecasting Competition 2014 and beyond. International Journal of Forecasting 32(3), 896–913 (2016). https://doi.org/10.1016/j.ijforecast.2016.02.001.
24. Ait Taleb, S., Sebaa, A., Bouzera, R., Aksouh, A., Ladlani, R., Hadjout, D.: Natural Gas Consumption Forecasting Based on Weighted Ensemble Learning. In: Aissani, D., Barkaoui, K., Roche, M. (eds) Research in Computer Science. CARI 2024. Communications in Computer and Information Science, vol. 2462. Springer, Cham (2025). https://doi.org/10.1007/978-3-031-88226-5_1

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.