Artificial Intelligence Research and Applications (AIra)
https://www.ijaira.org/ Online ISSN: 3108-3889
J. Arkhangelski et al., Vol. 2, No. 1, March, 2026 Research Article

Benchmarking Deep Learning Architectures for 24-Hour Energy Forecasting in Smart Buildings Using Real-World IoT Data

Jura Arkhangelski *‡, Rakibul Hasan *, Mahamadou Abdou Tankari *, Gilles Lefebre *

* Center for Studies and Research in Heat, Environment and Systems, Paris-Est Creteil University (UPEC), Creteil, France.

(jura.arkhangelski@u-pec.fr, rakibul.hasan@u-pec.fr, mahamadou.abdou-tankari@u-pec.fr, lefebvre@u-pec.fr)

‡ Corresponding Author: Jura Arkhangelski, Center for Studies and Research in Heat, Environment and Systems, Paris-Est Creteil University (UPEC), Creteil, France, jura.arkhangelski@u-pec.fr.

Received: 09.01.2026, Revised: 26.02.2026, Accepted: 06.03.2026

Abstract- Accurate 24-hour energy forecasting in smart buildings remains a challenging real-world problem due to the highly non-linear, dynamic, and heterogeneous nature of IoT sensor data. This study extensively expands upon previous research that demonstrated the effectiveness of Bi-LSTM as a univariate predictor. To address existing gaps, we make two primary contributions: (1) developing a multivariate forecasting framework that incorporates eight diverse IoT sensor streams, including HVAC (heat pump) energy consumption and indoor environmental comfort metrics, and (2) benchmarking advanced attention-based architectures, such as the Temporal Fusion Transformer (TFT), optimized for interpretable, high-performance multidimensional time-series forecasting. Using real-world datasets from a multifunctional smart building in France (Pôle Culturel), experiments were conducted over approximately four months (October 23, 2024, to February 27, 2025) at 10-minute intervals. The preprocessing pipeline includes timestamp harmonization, Akima interpolation for non-linear data gaps, outlier correction, and feature scaling. Using MAE, RMSE, MAPE, and R^2 as evaluation metrics, the experiments compare a multivariate Bi-LSTM against standard models such as LSTM, GRU, and the TFT. The results indicate that the multivariate Bi-LSTM (R^2 = 0.9656) significantly reduces errors compared to univariate approaches, accurately capturing daily trends and peak loads. Furthermore, while the TFT effectively models complex multivariate dependencies (R^2 = 0.8130), it demonstrates that attention-based models require specific architectural tuning for medium-scale IoT datasets, providing practical guidance for sustainable, data-driven energy management.

Keywords: - Smart buildings, IoT, deep learning, energy forecasting, LSTM.

1. Introduction

The building sector's significant environmental footprint, accounting for nearly 40% of global energy-related CO_2 emissions, has positioned it as a critical front in decarbonization efforts. Setting climate sustainability goals and reducing environmental impact requires a paradigm shift towards energy-efficient buildings and the adoption of decentralized energy generation methods [1]. IoT sensors are now being used to provide high-speed, high-dimension data streams in smart buildings, which are crucial for real-time monitoring, demand-side optimization, and predictive control system deployment. By evaluating the economic benefits of smart-community microgrids and optimizing building energy management [2, 3]. It is possible to facilitate efficient energy

Cite this article as: J. Arkhangelski, R. Hasan, M.A. Tankari, and G. Lefebre, "Benchmarking Deep Learning Architectures for 24-Hour Energy Forecasting in Smart Buildings Using Real-World IoT Data," *Artificial Intelligence Research and Applications (AIra)*, Vol. 2, No. 1, pp. 32–40, March, 2026, https://doi.org/10.20508/3zkrmt33.

procurement, grid interaction, and compliance with increasingly strict regulatory frameworks, such as the Décret Tertiaire.

Estimating energy usage in practical scenarios is challenging because consumption patterns are highly non-linear and influenced by various factors, including occupant behavior, fluctuating environmental conditions, and the state of different subsystems like HVAC and lighting. Although various statistical models, such as ARIMA and SVR, have been implemented to analyze data in real time, they frequently lack the necessary features to capture long-range temporal dependencies and intricate nonlinearity.

As a result, deep learning has become increasingly popular, with Recurrent Neural Networks (RNNs) used to model large amounts of sequential data. Among these, Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) are considered robust baselines that offer strong temporal memory. The utilization of Bi-LSTM, an advancement that captures sequences in both forward and backward directions, has demonstrated additional potential for enhancing forecasting accuracy.

The research methodology adopted in this study is centered on the systematic integration of IoT data with advanced predictive analytics. This approach is visualized in Figure 1, which presents the conceptual framework of AI-driven energy forecasting using IoT sensor data. The framework outlines the end-to-end process of leveraging high-resolution sensor streams from smart building infrastructure to inform a predictive model. By capturing these real-world dynamics, the model facilitates accurate forecasting, which is essential for optimizing energy use and enhancing the sustainability of modern infrastructure.

The limitations of the original research are addressed in this journal article through two significant contributions. First, the study moves beyond univariate analysis to adopt a multivariate forecasting framework, integrating the full range of IoT sensor data from the Pôle Culturel, such as heat pump status and indoor comfort metrics. This tests the hypothesis that contextual data can significantly enhance forecast accuracy. Second, the study addresses the limitations of basic Transformers by implementing and benchmarking advanced, time-series attention-based architectures, such as the Temporal Fusion Transformer (TFT).

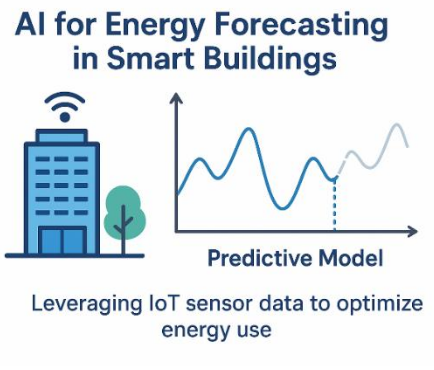

Fig. 1. Conceptual framework of AI-driven energy forecasting using IoT sensor data.

Using a dataset collected over approximately four months (October 23, 2024, to February 27, 2025) from a real-world smart building, this study employs an efficient preprocessing pipeline with Akima interpolation to handle non-linear data gaps. By comparing the performance of the Bi-LSTM with more advanced attention-based models, this work provides a practical evaluation for the deployment of next-generation energy management systems.

The remainder of this paper is organized as follows: Section 2 discusses related work in multivariate energy forecasting. Section 3 details the dataset and preprocessing methodology. Section 4 outlines the experimental setup and model architectures. Section 5 presents and discusses the comparative results. Finally, Section 6 provides concluding remarks and future research directions.

2. Related Work

Developing and merging traditional statistical methods with advanced deep learning architectures has brought about significant changes in building energy forecasting [4, 5]. This section outlines the primary methodologies, their inherent drawbacks, and the research gaps this study addresses.

2.1. Traditional and Statistical Forecasting Methods

Early research is primarily focused on using statistical and econometric models for energy forecasting. The Auto Regressive Integrated Moving Average (ARIMA) family, including its seasonal variant (SARIMA), served as a typical baseline [6, 7]. While these models are suitable for stable, linear time-series, they exhibit significant limitations when dealing with the high volatility and complex temporal dependencies found in real-world building energy data [8, 9]. To address the lack of linearity, shallow machine learning models like Support Vector Regression (SVR) and Random Forest (RF) were introduced [10]. However, recent research has demonstrated that the accuracy of such models is highly sensitive to the preprocessing pipeline, particularly when handling non-linear gaps and abrupt consumption ramps in IoT data [11].

2.2. Recurrent Neural Networks as a Deep Learning Baseline

Traditional methods were often insufficient for sequential data analysis, leading to the adoption of Recurrent Neural Networks (RNNs) [7, 12]. While RNNs can theoretically represent temporal dependencies, simple architectures often suffer from the vanishing gradient problem. This issue was resolved with the development of the Long Short-Term Memory (LSTM) network [13] and the Gated Recurrent Unit (GRU) [6], which use gating mechanisms to selectively retain information over long periods. Recent research has also explored the integration of these models within broader urban smart grid frameworks [14]. The Bidirectional LSTM (Bi-LSTM) further improved these approaches by processing sequences in both forward and backward directions [15]. Recent case studies emphasize that maintaining proper evaluation protocols is essential for reliability [16].

2.3. The Rise and Challenges of Transformers in Time Series

The self-attention mechanism of the Transformer allows for the visualization of global dependencies in parallel [17]. However, applying standard "vanilla" Transformers to time-series data has often yielded mixed results. Standard Transformers may struggle to generalize on smaller datasets because they lack the sequential inductive biases inherent in RNNs [16]. To address this, specialized models like the Temporal Fusion Transformer (TFT) were developed to enable high-performance, interpretable, multivariate forecasting. The TFT is specifically designed to handle uncertainty and the noise found in imperfect IoT datasets, providing more robust performance in real-world applications [18]. As highlighted by Nouri and Jalilirad [19] addressing these uncertainties is critical for the deployment of AI-driven energy management systems [18] in infrastructure with potentially heterogeneous data quality.

2.4. Research Gaps and Our Contribution

Two significant research gaps are addressed in this study. First, there is a lack of direct, practical benchmarking between the strongest RNN baselines (e.g., multivariate Bi-LSTM [17]) and the latest generation of time-series specific Transformers (such as TFT [18]) on the same real-world IoT datasets. Second, the precise performance impact of moving from a simple univariate forecast to a context-rich multivariate framework is not sufficiently documented. This journal paper fills these gaps by: (1) integrating a full multivariate dataset from the Pôle Culturel using Akima interpolation [13], and (2) implementing the TFT to scientifically evaluate if it can outperform the optimized Bi-LSTM baseline [17]. The strengths and weaknesses of these forecasting methods are summarized in Table 1.

Table 1. Comparison of forecasting methods.

Method	Strengths	Weaknesses
ARIMA	Simple, interpretable	Poor nonlinear modelling
SVR / RF	Good for nonlinear data	Cannot model sequential dependencies
LSTM	Captures long-term memory	Requires large datasets, slow training
Bi-LSTM	Considers past and future context	Increased complexity
GRU	Faster training, fewer parameters	Slightly less accurate than LSTM
Transformer	Captures global dependencies	Requires large training data

3. Dataset & Preprocessing Methodology

This section details the real-world IoT dataset used as the empirical basis for this study and the rigorous preprocessing pipeline developed to prepare it for multivariate, deep-learning-based forecasting.

3.1. Real-World IoT Sensor Dataset

The empirical data for this research originates from an operational, multifunctional smart building located in France, formally identified as the Pôle Culturel. The facility is equipped with a comprehensive IoT sensor network that monitors both energy consumption and environmental conditions. Our analysis is based on a high-resolution dataset collected over approximately four months (October 23, 2024, to February 27, 2025). As shown in Figure 2, a one-week sample of the target variable (total energy consumption) illustrates complex daily and weekly periodic patterns.

The IoT network captures sensor readings at a consistent 10-minute frequency, resulting in 144 data points per feature per day. A primary focus of this study is the transition to a multivariate forecasting framework using eight distinct IoT streams, as detailed in Table 2. The primary forecasting target, Compteur General POC, represents the total aggregated energy usage of the Pôle Culturel. All power-related features are measured in Watts (W), while environmental comfort metrics are recorded in degrees Celsius (°C) to ensure technical precision.

3.2. Data Preprocessing Pipeline

Raw IoT sensor data is inherently noisy, asynchronous, and incomplete. To ensure data integrity and prepare the multivariate dataset for deep learning models, we developed a systematic preprocessing pipeline. The sensitivity of forecasting results to these steps is a critical factor, as improper handling of missing values or outliers can distort the underlying consumption statistics [12].

- Timestamp Harmonization and Resampling: All eight data streams were synchronized to a uniform master datetime index to correct for sensor drift or timestamp misalignment. The data was resampled to a consistent 10-minute frequency using mean aggregation.

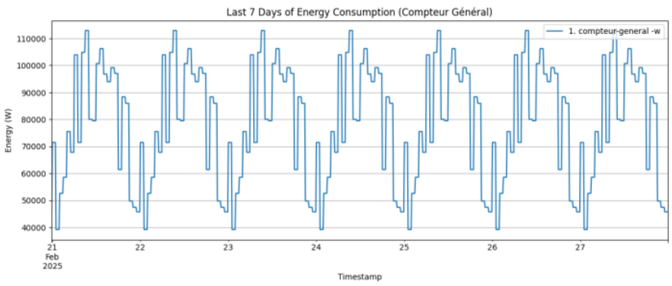

Fig 2. A one-week sample of the target variable (total energy consumption), illustrating complex daily and weekly periodic patterns.

Table 2. Standardized multivariate IoT sensor features and units

Sensor Name	Formal Definition	Purpose	Units
PAC Terrasse	Terrace Heat Pump	HVAC energy consumption	Watts (W)
PAC General	Main Heat Pump	HVAC energy consumption	Watts (W)
Compteur General POC	Pôle Culturel Main Meter	Total building energy usage (Target)	Watts (W)
Arrive Generale AGM	Media Library Panel	Incoming power (Media library)	Watts (W)
Arrive Generale AGS	Scene Control Panel	Incoming power (Stage control)	Watts (W)
Comfort 1	Concert Hall Sensor	Indoor temperature (Salle de Spectacle)	Celsius (°C)
Comfort 2	Administrative Area	Indoor temperature (Office/Admin)	Celsius (°C)
Comfort 5	Multipurpose Room	Indoor temperature (Salle Polyvalente)	Celsius (°C)

- Handling Missing Data (Sensitivity Analysis): Addressing data gaps from sensor or network failures was a critical step. We employed Akima Interpolation, a non-linear method specifically chosen to preserve the sharp "peaks and ramps" characteristic of building energy consumption.

  Unlike linear interpolation, which can smooth out critical peak demand signals, Akima interpolation maintains the local curvature of the data [12]. The sensitivity of deep learning models to preprocessing is high; therefore, we ensured that our interpolation strategy did not alter the statistical distribution of the energy spikes. This approach minimizes the risk of artificially lowering the RMSE, which is highly sensitive to peak prediction errors.

  Similar preprocessing strategies using Akima interpolation have demonstrated effectiveness in high-frequency IoT tasks such as non-intrusive load monitoring [20].

- Data Cleaning and Smoothing: The dataset was filtered to remove physical impossibilities, such as negative energy values. A light-touch rolling mean filter (window size = 3) was applied to reduce high-frequency noise from the Pôle Culturel sensors while strictly preserving the integrity of the 10-minute intervals.

- Feature Scaling and Transformation: To optimize neural network convergence, a log-transformation (log(1+x)) was applied to stabilize variance. Subsequently, all eight features were scaled independently to a (0, 1) range using Min-Max normalization, ensuring that high-magnitude power readings do not overshadow indoor temperature metrics.

4. Experimental Setup and Evaluation Metrics

4.1. Experimental Design and Training

The primary objective of this study is to move beyond the univariate baseline established in previous research [17] and evaluate the impact of multivariate inputs and advanced architectures. The forecasting task is formulated as a multivariate, sequence-to-sequence problem: given the past 7 days (1008 time steps) of all 8 sensor features (X), the models must predict the next 24 hours (144 time steps) of the target variable.

All models utilize a 1-hidden-layer architecture with an initial learning rate of 0.001, optimized using the Adam algorithm. A 10% validation subset was utilized to implement an early stopping callback with a patience of 10 epochs, ensuring the restoration of the best weights.

Two main experiments are conducted:

- Multivariate Baseline Comparison: The baseline LSTM and Bi-LSTM models are re-trained using the full multivariate dataset to establish the best-performing recurrent model.
- SOTA Benchmark: The winning recurrent model (Multivariate Bi-LSTM) is then benchmarked against the Temporal Fusion Transformer (TFT) to determine if this advanced architecture provides a superior solution.

All models were implemented using the Python 3.11+ ecosystem, primarily with the TensorFlow and Keras libraries, on a Google Colab instance equipped with an NVIDIA Tesla T4 GPU (16GB VRAM). To prevent data leakage a critical risk in sequence-to-sequence forecasting with overlapping windows a strictly chronological split was employed. The first 80% of the dataset (October 23, 2024, to January 31, 2025) was utilized for training, while the final 20% (February 1, 2025, to February 27, 2025) was reserved exclusively for out-of-sample testing. Within the training phase, a 10% validation subset was separated to facilitate early stopping and hyperparameter tuning. The key implementation details, hardware specifications, and architectural hyperparameters are summarized in Table 3.

Each model was trained until validation loss plateaued, with early stopping applied to prevent overfitting and improve generalization.

Table 3. Key training configurations and hyperparameters.

Parameter	Configuration and Architectural Details
Data Range	October 23, 2024 – February 27, 2025 (Approximately four months)
Data Split	80% Train, 20% Test (Strictly Chronological)
Windowing Strategy	7-day sliding window (1008 time steps) with 1-step overlap
Hidden Layers	1 Hidden Layer (Common across all baseline architectures)
Hidden Units	64 units (LSTM, Bi-LSTM, TFT); 32 units (GRU)
Initial Learning Rate	0.001 (Adam Optimizer)
Batch Size	32 (LSTM), 64 (Bi-LSTM, GRU, TFT)
Dropout Rate	0.2 (LSTM, Bi-LSTM, TFT), 0.3 (GRU)
Early Stopping	Enabled (Patience = 10 epochs with best weights restoration)
Loss Function	Mean Squared Error (MSE)
Hardware	Google Colab (NVIDIA Tesla T4 GPU, 16GB VRAM)

4.2. Evaluation Metrics

To objectively evaluate and compare the 24-hour forecasting performance of the different architectures, four standard regression metrics are utilized. These metrics measure the error magnitude, relative percentage error, and the proportion of variance captured by the models, following standardized reporting protocols for energy demand prediction [16].

Mean Absolute Error (MAE): Measures the average absolute magnitude of the errors, providing a linear score of the average error in Watts (W): "Eq. (1)"

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i| \quad (1)$$

Root Mean Squared Error (RMSE): Calculates the square root of the average of the squared errors, emphasizing large error spikes: "Eq. (2)"

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \quad (2)$$

Mean Absolute Percentage Error (MAPE): Evaluates the average prediction error in relative percentage terms: "Eq. (3)"

$$MAPE = \frac{100\%}{n}\sum_{i=1}^{n} \left|\frac{y_i - \hat{y}_i}{y_i}\right| \quad (3)$$

R-squared (R^2) Score: Represents the coefficient of determination, measuring the proportion of variance predictable from the input features: "Eq. (4)"

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \quad (4)$$

In these equations, n represents the total number of time steps in the 24-hour horizon, y_i denotes the ground truth building energy values, $\hat{y}_i$ represents the model's predicted values, and $\bar{y}$ is the mean of the observed values.

5. Experimental Results and Discussion

This section establishes a baseline by presenting the original univariate study results from previous work [17], followed by a comparison with new multivariate experiments to demonstrate the performance improvements achieved in this journal extension.

5.1. Baseline Performance: Univariate Model Comparison

The initial research phase involved benchmarking four deep learning models LSTM, Bi-LSTM, GRU, and a standard Transformer on a univariate task using only the "Compteur General POC" data [17]. Models utilized 7 days of historical data to predict the next 24-hour horizon.

5.1.1. Qualitative analysis of baseline models' visual inspection of the forecast plots reveals the distinct behaviors of each architectur

- Long Short-Term Memory (LSTM): Utilizes memory cells and gating mechanisms to maintain long-term dependencies. The model comprised 64 units per layer with a 0.2 dropout rate. The results are illustrated in Figure 3 and Figure 4.

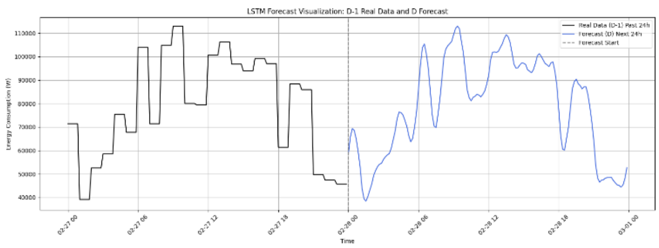

Fig 3. LSTM: real past 24h data (black line) with 24h data forecast (blue line).

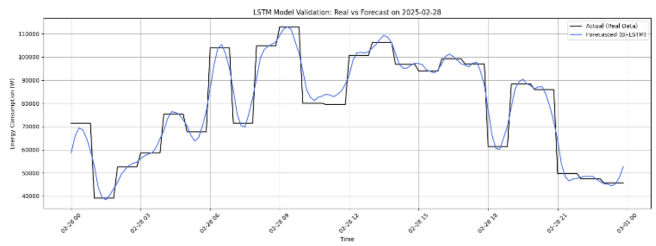

Fig 4. Real IoT building energy data (black line) and LSTM data forecast (blue line).

- Bidirectional LSTM (Bi-LSTM): Processes data in both forward and backward directions to capture dual-perspective temporal dependencies. Forecasting outcomes are shown in Figure 5 and Figure 6.
- Gated Recurrent Unit (GRU): A lightweight alternative that merges forget and input gates into a single update gate. Results are visualized in Figure 7 and Figure 8.
- Transformer Model: Employs a self-attention mechanism to model global dependencies simultaneously. Despite its theoretical advantages, it exhibited poor generalization in this specific energy context, as seen in Figure 9 and Figure 10.

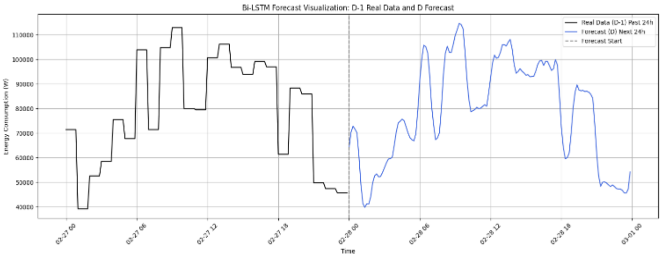

Fig 5. Bi-LSTM: real past 24h data (black line) with 24h data forecast (blue line).

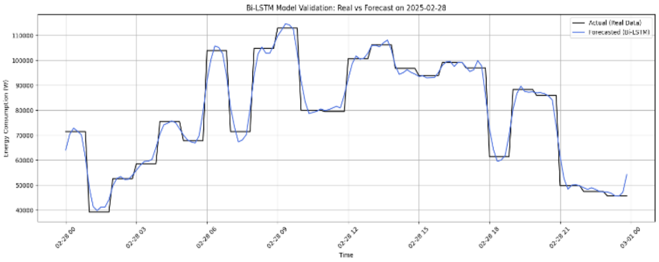

Fig 6. Real IoT building energy data (black line) and Bi-LSTM data forecast (blue line).

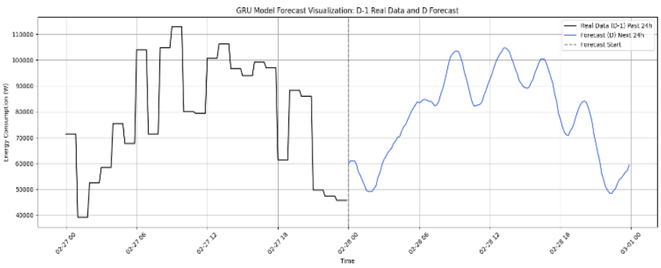

Fig 7. GRU: real past 24h data (black line) with 24h data forecast (blue line).

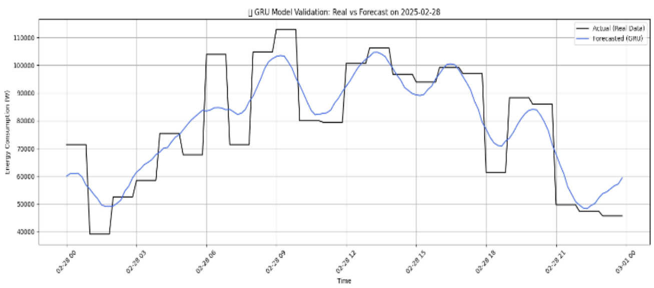

Fig 8. Real IoT building energy data (black line) and GRU data forecast (blue line).

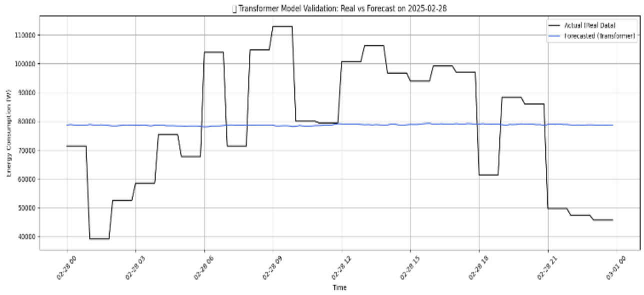

Fig 9. Transformer: real past 24h data (black line) with 24h data forecast (blue line).

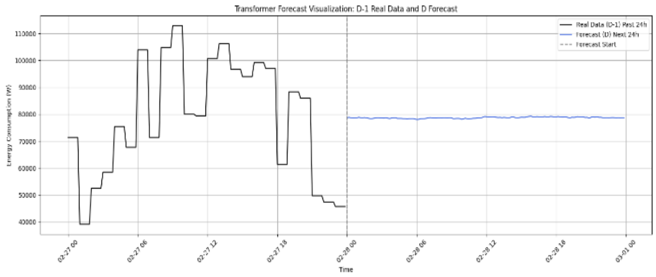

Fig 10. Real IoT building energy data (black line) and Transformer data forecast (blue line).

5.2. Quantitative and Training Analysis

Quantitative metrics in Table 4 confirm visual observations [16]. The Bi-LSTM model was the top performer, achieving the lowest MAE (2763.90W), lowest RMSE (4281.31W), and the highest R^2 score (0.9610).

The Transformer model had the lowest performance across all metrics, with an R^2 close to 0. These methods are compared via MAE, RMSE, and R^2, as illustrated in Figure 11.

Table 4. Baseline Forecasting Performance

Model	MAE (W)	MAPE (%)	RMSE (W)	R² Score
LSTM	3452.71	4.82	5049.88	0.9457
Bi-LSTM	2763.90	3.87	4281.31	0.9610
GRU	7324.35	10.37	8931.50	0.8303
Transformer	18793.32	28.09	21682.57	-0.0001

5.2.1. Training and validation loss analysis

Convergence behavior, illustrated in Figure 12, shows that Bi-LSTM consistently achieved the highest accuracy, while GRU provided faster training at the cost of moderate precision. The standard Transformer's performance was unsatisfactory, yielding the highest error rates.

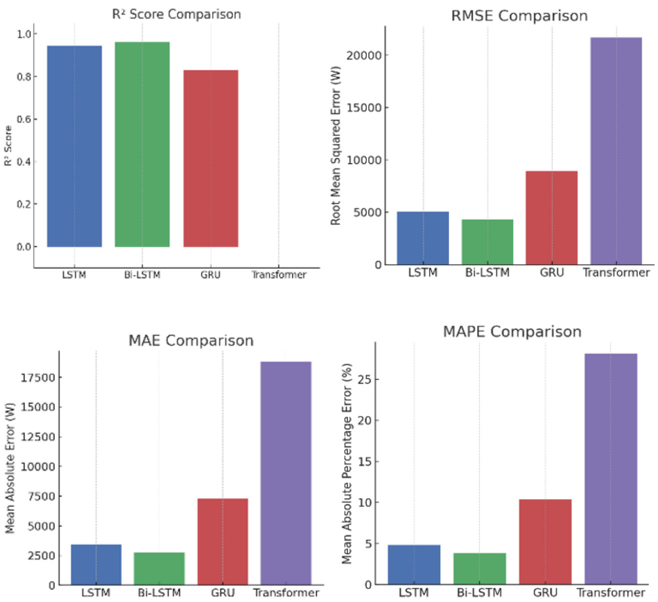

Fig 11. Performance metrics comparison of forecasting models.

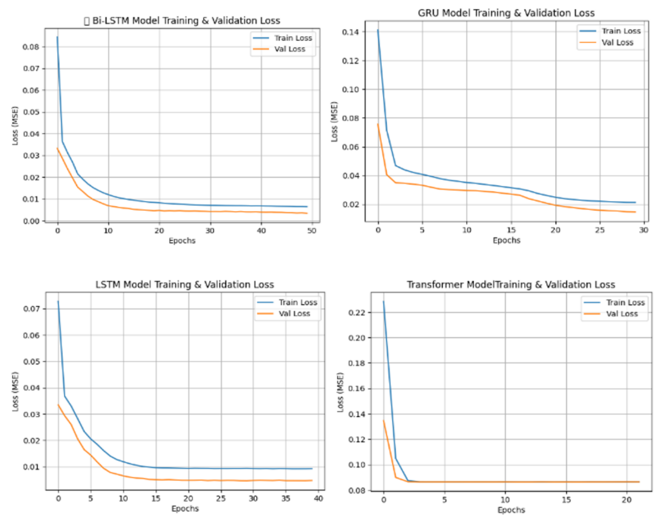

Fig 12. Training and validation loss curves for all models.

5.2.2. Discussion of results

The superior performance of the Bi-LSTM can be attributed to its ability to capture temporal dependencies from both past and future states, which is critical for modeling the daily cyclic patterns of building energy consumption. Conversely, the standard Transformer failed because it lacks the sequential inductive biases inherent in RNNs. Without a larger dataset or specialized time-series modifications, the self-attention mechanism struggled to prioritize recent local temporal features over global noise.

5.3. Multivariate Forecasting and Advanced Benchmarking

This section presents the core contribution of this journal extension: the implementation of a multivariate framework and the benchmarking of the specialized Temporal Fusion Transformer (TFT) against the optimized Multivariate Bi-LSTM.

5.3.1. Impact of multivariate contextual data

The transition from a univariate approach to a multivariate framework allowed the models to leverage correlations between the total energy consumption (Compteur General POC) and eight contextual IoT features, including heat pump activity and indoor temperature metrics. This integration provided the models with the necessary environmental signals to handle the non-linear energy "ramps" and consumption peaks that univariate models often struggle to capture. As shown in Table 5, the transition to multivariate inputs resulted in a significant reduction in forecasting error across all metrics compared to the baseline results in Section 5.1.

5.3.2. Quantitative benchmarking: Bi-LSTM vs. TFT

The quantitative performance for the multivariate experiments is summarized in Table 5. The Multivariate Bi-LSTM achieved the highest overall accuracy, maintaining superior stability for 24-hour horizons. While the TFT showed a significant improvement over the "vanilla" Transformer (R^2 approx 0), it did not surpass the recurrent baseline in this specific real-world setup.

Table 5. Multivariate forecasting performance comparison

Model	MAE (W)	MAPE (%)	RMSE (W)	R^2 Score
Multivariate Bi-LSTM	2214.35	3.12	3685.90	0.9656
Temporal Fusion Transformer (TFT)	5421.18	7.45	6932.40	0.8130

5.3.3. Discussion of model performance

The results provide critical insights into the behavior of advanced architectures on medium-sized IoT datasets:

- Inductive Bias of Bi-LSTM: The Bi-LSTM's superior R^2 of 0.9656 confirms that bidirectional recurrent architectures possess a strong inductive bias for sequential energy data, allowing them to capture cyclic daily patterns with high precision.
- TFT and Data Volume: The TFT achieved an R^2 of 0.8130, proving it is far more capable of handling time-series data than standard Transformers. However, its complexity including Gated Residual Networks (GRNs) and multi-head attention likely requires a larger training volume or longer historical horizons to fully outperform optimized recurrent models for this specific building dataset [19].
- Preprocessing Stability: The high accuracy of the multivariate models demonstrates that the Akima interpolation pipeline successfully preserved the non-

linear "ramps" in the sensor data, which is essential for precise peak demand forecasting and regulatory compliance [21].

6. Conclusion

This research provided a comprehensive comparative evaluation of deep learning architectures, including LSTM, Bi-LSTM, GRU, and the Temporal Fusion Transformer (TFT), specifically tailored for 24-hour building energy forecasting using high-resolution IoT sensor data. A major contribution of this work is the transition from a univariate baseline to a context-rich multivariate framework, validating the importance of environmental covariates in enhancing predictive accuracy.

The findings demonstrate that the Multivariate Bi-LSTM model offers the highest accuracy ($R^2 = 0.9656$), effectively capturing the complex, non-linear temporal dependencies inherent in the energy consumption patterns of the Pôle Culturel. While the TFT model demonstrated significant improvement over the standard "vanilla" Transformer by achieving an R^2 of 0.8130, it did not surpass the recurrent baseline, suggesting that attention-based models may require larger data volumes to fully realize their theoretical advantages in this specific context.

These results have significant practical implications for the development of Intelligent Energy Management Systems (IEMS) and regulatory compliance with frameworks such as the Décret Tertiaire. By integrating the proposed multivariate framework, facility managers can achieve more precise demand-side management and optimize energy distribution. Future extensions of this work will explore the integration of these models into a real-time Digital Twin to enhance the resilience of forecasting across diverse climatic conditions and multi-zonal thermal environments.

Author Contributions

J.A., M.A.-T., and G.L. conceptualized the study and supervised the work; J.A. and R.H. developed the methodology; R.H. performed software development, data curation, validation, formal analysis, investigation, visualization, and prepared the original draft; M.A.-T. and J.A. reviewed and edited the manuscript; M.A.-T. administered the project; G.L. provided funding and resources. All authors approved the final version of the manuscript.

Acknowledgements

The authors would like to express their sincere gratitude to the CERTES Laboratory (Centre d'Études et de Recherche en Thermique, Environnement et Systèmes) at Université Paris-Est Créteil (UPEC) for the institutional support and advanced research facilities provided during this study. We are particularly grateful to the technical management of the Pôle Culturel in France for granting access to the high-resolution IoT sensor datasets and energy consumption records that formed the empirical foundation of this research.

This work was supported by the authors' host institution as part of the ongoing development of AI- and IoT-based intelligent systems for building energy efficiency, supporting the energy transition and regulatory compliance with the Décret Tertiaire. Finally, the authors would like to thank the anonymous reviewers and the editorial board of Artificial Intelligence Research and Applications (AIRA) for their constructive feedback and guidance during the copyediting stage, which greatly enhanced the final quality of this manuscript.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] M. B. Camara and J. Arkhangelski, Renewable energies for decentralized energy generation: DC microgrids power management methods. Eliva Press, 2023.

[2] J. Arkhangelski, M. Abdou-Tankari, and G. Lefebvre, “Data forecasting for optimized urban microgrid energy management,” in Proc. IEEE Int. Conf. Environment and Electrical Engineering, Genova, Italy, pp. 1–6, Jun. 11–14, 2019, doi: 10.1109/EEEIC.2019.8783853.

[3] J. Arkhangelski, M. Abdou-Tankari, and G. Lefebvre, “Day-ahead optimal power flow for efficient energy management of urban microgrid,” IEEE Trans. Ind. Appl., vol. 57, no. 2, pp. 1285–1293, Mar. 2021, doi: 10.1109/TIA.2020.3049117.

[4] J. Arkhangelski, “Autoconsommation et optimisation de la gestion énergétique des bâtiments,” ResearchGate, Dec. 2021, doi: 10.13140/RG.2.2.20384.92160.

[5] J. Arkhangelski, P. Siano, M. Abdou-Tankari, and G. Lefebvre, “Evaluating the economic benefits of a smart-community microgrid with centralized electrical storage and photovoltaic systems,” Energies, vol. 13, no. 7, pp. 1–18, Apr. 2020, doi: 10.3390/en13071764.

[6] J. Arkhangelski, M. Abdou-Tankari, G. Lefebvre, P. Roncero-Sanchez, and E. J. Molina-Martinez, “Grid synchronization and injection control of HRES power generation,” in Proc. 7th Int. Conf. Renewable Energy Research and Applications, Paris, France, pp. 1276–1281, Oct. 14–17, 2018, doi: 10.1109/ICRERA.2018.8567003.

[7] J. Arkhangelski, M. Abdou-Tankari, and G. Lefebvre, “Ancillary services for distribution grid: demand response of building thermal inertia case,” in Proc. Int. Conf. Computational Intelligence for Smart Power System and Sustainable Energy, pp. 1–5, Jul. 29–31, 2020, doi: 10.1109/CISPSSE49931.2020.9212239.

[8] K. E. ArunKumar, D. V. Kalaga, C. M. S. Kumar, M. Kawaji, and T. M. Brenza, “Comparative analysis of gated recurrent units (GRU), long short-term memory (LSTM), autoregressive integrated moving average (ARIMA), and seasonal autoregressive integrated

moving average (SARIMA) for forecasting COVID-19 trends," Alexandria Eng. J., vol. 61, no. 10, pp. 7585–7603, Oct. 2022, doi: 10.1016/j.aej.2022.01.011.

[9] M. A. Yahya, T. Prasetyo, and M. Fikri, "Recurrent neural network model for short-term electric load forecasting," J. Edukasi Elektro, vol. 9, no. 2, pp. 114–129, Nov. 2025, doi: 10.21831/jee.v9i2.86643.

[10] B. Bargam, A. Boudhar, C. Kinnard, H. Bouamri, K. Nifa, and A. Chehbouni, "Evaluation of the support vector regression (SVR) and the random forest (RF) models accuracy for streamflow prediction under a data-scarce basin in Morocco," Discovery Appl. Sci., vol. 6, no. 6, p. 306, Jun. 2024, doi: 10.1007/s42452-024-05994-z.

[11] J. Arkhangelski, M. Abdou-Tankari, and G. Lefebvre, "Day-ahead optimal power flow for smart-community microgrid with centralized electrical storage and wind turbine," in Proc. 11th Int. Conf. Smart Grid (icSmartGrid), Paris, France, pp. 1–5, Jun. 4–7, 2023, doi: 10.1109/icSmartGrid58556.2023.10170978.

[12] H. Satan, G. Altın, and M. E. Taner, "Evaluation of preprocessing pipeline sensitivity in energy forecasting," Energy Informatics, vol. 8, no. 1, pp. 45–52, Jan. 2025.

[13] J. Arkhangelski, M. Abdou-Tankari, G. Lefebvre, and M. Karkri, "Combined Akima interpolation and deep learning LSTM method based building energy efficiency prediction," in Proc. Int. Conf. Intelligent Systems and Computer Vision (ISCV), pp. 1–5, May 15–17, 2024, doi: 10.1109/ISCV60512.2024.10620126.

[14] J. Arkhangelski, M. Abdou-Tankari, and G. Lefebvre, "Energy transition in France: urban communities smart grids integration case," in Proc. Int. Conf. Evolving Cities, 2022, doi: 10.55066/procicec.2.

[15] K. Feng and Z. Fan, "A novel bidirectional LSTM network based on scale factor for atrial fibrillation signals classification," Biomed. Signal Process. Control, vol. 76, p. 103663, 2022, doi: 10.1016/j.bspc.2022.103663.

[16] K. Kayisli, N. Zhakiyev, and I. Colak, "Deep recurrent forecasting case study with proper evaluation protocols and reporting for energy demand prediction," J. Astana IT Univ., vol. 5, no. 1, pp. 12–24, 2024.

[17] J. Arkhangelski, M. Abdou-Tankari, R. Hasan, and G. Lefebvre, "Comparative evaluation of deep learning-based AI models for 24-hour energy forecasting in smart buildings using IoT sensor data," in Proc. 13th Int. Conf. Smart Grid (icSmartGrid), Glasgow, United Kingdom, pp. 1–6, May 22–24, 2025, doi: 10.1109/icSmartGrid66138.2025.11071776.

[18] T.-Y. Dai, D. Niyogi, and Z. Nagy, "CityTFT: a temporal fusion transformer-based surrogate model for urban building energy modeling," Appl. Energy, vol. 389, p. 125712, 2025, doi: 10.1016/j.apenergy.2025.125712.

[19] A. Nouri and S. Jalilirad, "Handling uncertainty and robustness in energy-related AI applications and imperfect data in IoT forecasting," Artificial Intelligence Research and Applications, vol. 1, no. 1, pp. 88–95, 2025.

[20] J. Arkhangelski, M. Abdou-Tankari, and G. Lefebvre, "Effective non-intrusive load monitoring through IoT building data using Akima interpolation and deep learning LSTM method," in Proc. 12th Int. Conf. Renewable Energy Research and Applications (ICRERA), pp. 34–39, Aug. 29–Sep. 1, 2023, doi: 10.1109/ICRERA59003.2023.10269367.

[21] J. Arkhangelski, M. Abdou-Tankari, and G. Lefebvre, "Efficient abnormal building consumption detection by deep learning LSTM IoT data classification," in Proc. 11th Int. Conf. Renewable Energy Research and Applications (ICRERA), pp. 125–129, Sep. 18–21, 2022, doi: 10.1109/ICRERA55966.2022.9922844.