

Light-SustainFormer: A Lightweight Temporal Fusion Transformer-Based CNN–BiGRU Framework for Sustainable Smart Building Energy Forecasting

Arti Ranjan^{1,2*}, Ashish Arya³

¹ Department of Computer Science, Niels Brock Copenhagen Business College (NBCBC), Copenhagen, Denmark

² Department of Computer Science & Engineering, Indira Gandhi Delhi Technical University for Women (IGDTUW), New Delhi, India

³ Department of Computer Science & Engineering, Indian Institute of Information Technology (IIIT), Sonapat, Haryana, India

arti.ranja@gmail.com¹², ashishknew81@gmail.com³

*Corresponding author: Arti Ranjan (arti.ranja@gmail.com)

Abstract.

Energy consumption forecasting is central to the design of smart, sustainable, and cost-efficient buildings. Recent deep learning and Transformer-based models have improved forecasting accuracy, yet most such models prioritise predictive performance over computational cost and model size. Full Transformer and Temporal Fusion Transformer architectures capture long-range temporal patterns, but their large dimensions and multiple attention blocks require high computational cost. This paper introduces Light-SustainFormer, a lightweight deep learning framework for short-term energy consumption forecasting in smart buildings. We propose a Lite-Temporal Fusion Transformer attention module that streamlines the conventional TFT by using fewer attention heads and a single attention block with a lightweight feed-forward layer. Instead of applying an attention block directly to raw time-series data, the proposed model first passes the inputs through CNN layers to capture short-term energy fluctuations, followed by Bi-GRU layers to produce compact temporal representations. This CNN–BiGRU-based compression reduces computational cost without sacrificing the ability to model long-range energy usage patterns. Light-SustainFormer learns energy consumption behaviour across weather, occupancy, time of day, and seasonal variations. On experimental benchmarks, the Light-SustainFormer achieves an MAE of 0.118 kWh, an RMSE of 0.164 kWh, a MAPE of 4.82%, and an R^2 of 0.961. These results demonstrate that a lightweight, attention-based architecture can deliver strong forecasting accuracy at low computational cost, providing a practical solution for energy management in smart buildings.

Keywords: Smart Buildings, Energy Consumption Forecasting, Sustainable Energy Management, Light-SustainFormer, Temporal Fusion Transformer, CNN–BiGRU, Lightweight Attention.

1. Introduction

Buildings are among the major contributors to global energy consumption and carbon emissions. According to the International Energy Agency (IEA), building operations account for nearly 30% of global energy consumption and 26% of energy-related emissions [1]. Therefore, improving building energy efficiency is essential for sustainable urban development. Smart buildings address this challenge by using sensors, smart meters, occupancy monitoring, building management systems, and automated control techniques to collect real-time data and manage energy use intelligently. In this context, accurate energy consumption forecasting plays an important role in reducing energy waste, improving demand-side management, lowering operational costs, and supporting sustainable building operations. Building energy consumption is highly dynamic and depends on several factors, including historical electricity usage, weather conditions, occupancy behaviour, time of day, day of week, holidays, and seasonal variations. Traditional statistical models often struggle to capture these

nonlinear and temporal dependencies. Recent studies show that machine learning and deep learning models are more effective for building energy forecasting because they can learn complex patterns from historical and contextual data as per Figure 1.[2], [3].

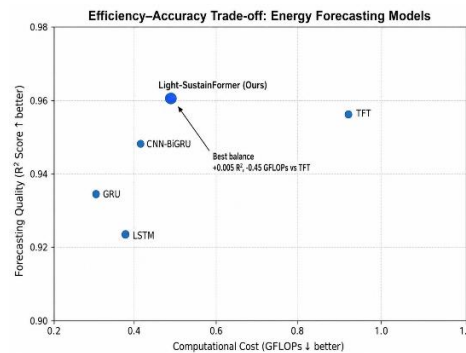


Figure 1: Efficiency–Accuracy Trade-off Analysis of Energy Forecasting Models

Deep learning models such as CNNs, LSTMs, GRUs, and hybrid CNN-LSTM architectures have been widely used for energy prediction [4]. CNN-based models can extract local temporal patterns, while recurrent models such as LSTM and GRU can learn sequential dependencies. However, CNNs may not fully capture long-term relationships, and recurrent models may become less effective for long forecasting horizons and multiple interacting variables. Transformer-based models have improved time-series forecasting through attention mechanisms that capture long-range dependencies. The Temporal Fusion Transformer (TFT) is particularly useful for multi-horizon forecasting because it combines recurrent processing, self-attention, variable selection, and interpretable temporal analysis [5], [6]. However, Transformer-based models are often computationally intensive, limiting their suitability for sustainable, resource-constrained smart building environments. To address this limitation, this paper proposes Light-SustainFormer, a lightweight CNN–Bi-GRU–Light-TFT attention framework for short-term energy consumption forecasting in smart buildings. The model uses CNN layers to extract local energy usage patterns, Bi-GRU layers to learn compact sequential representations, and a lightweight TFT-inspired attention module to capture long-range dependencies and important energy-related features. This design reduces computational complexity while maintaining accurate and interpretable energy forecasting. The major contributions of this paper are:

- A lightweight hybrid forecasting framework, Light-SustainFormer, is proposed for sustainable smart-building energy consumption forecasting.
- A novel Light-TFT Attention Module is introduced by modifying the conventional TFT attention structure using reduced attention heads, a smaller hidden dimension, a single attention block, and a lightweight feed-forward layer.
- CNN and Bi-GRU layers are used before the attention module to compress local and sequential energy patterns, reducing the computational burden of Transformer-based forecasting.
- The proposed model is evaluated using both forecasting accuracy metrics and computational efficiency metrics to demonstrate its suitability for sustainable smart-building deployment.

Unlike a full Temporal Fusion Transformer, the proposed Light-SustainFormer uses a compact Light-TFT Attention Module designed specifically for efficient smart-building energy forecasting. The module reduces computational overhead by using fewer attention heads, a smaller hidden dimension, a single attention block, and a lightweight feed-forward layer.

The remainder of this paper is organised as follows. Section 2 reviews related work on energy forecasting and Transformer-based models. Section 3 presents the proposed Light-SustainFormer methodology. Section 4 describes the training procedure and implementation setup. Section 5 discusses the experimental results, ablation study, and computational complexity analysis. Finally, Section 6 concludes the paper and outlines future research directions.

2. Literature Review

Energy consumption forecasting for buildings has been widely studied using statistical, machine learning, and deep learning techniques. Early energy consumption forecasting models used linear regression, autoregressive methods, and support vector regression techniques. These methods were simple and interpretable [7], [8]. These methods perform well under stable consumption patterns. Still, smart building energy consumption is affected by nonlinear dynamic factors such as ambient temperature, humidity, occupancy, HVAC operation, appliance usage, holidays, and seasonal variations [9], [10]. Conventional models struggle to capture the complex temporal and contextual relations in real-world energy consumption data. Machine learning models such as Random Forests, Extra Trees, gradient boosting, and XGBoost improved forecasting accuracy by learning nonlinear relationships between energy consumption and environmental variables [11], [12]. Feature selection and optimisation-based approaches have also been explored to reduce irrelevant variables and improve prediction accuracy [13]. However, most conventional machine learning models treat the input data as fixed feature vectors and do not explicitly capture sequential behaviour such as daily load cycles, weekly demand variations, peak-load patterns, and seasonal trends. Later, deep learning models were introduced to overcome some of these limitations. Recurrent Neural Networks, LSTM, and GRU are effective for learning temporal dependencies from energy time-series data [14], [15]. CNN-based models have also been applied because they can extract local temporal patterns and sudden changes caused by HVAC activation, appliance switching, lighting usage, and occupancy variation [16], [17]. Hybrid models such as CNN-LSTM, CNN-GRU, and CNN-BiLSTM further improve forecasting by combining local feature extraction with sequential learning [18], [19]. However, recurrent and hybrid models may still have limited ability to capture long-range temporal dependencies and identify the most important time steps in longer forecasting windows. Recently, attention-based and Transformer-based models have gained attention for time-series forecasting. The self-attention mechanism in Transformers can model relationships between distant time steps and improve long-range dependency learning [20]. The Temporal Fusion Transformer is particularly well-suited to multi-horizon forecasting because it combines variable selection, recurrent processing, gated residual networks, and interpretable attention mechanisms [21]. These properties make TFT useful for building energy forecasting, where energy consumption depends on historical usage, weather, occupancy, and known future temporal variables [22]. Recent studies have also used

interpretable TFT-based architectures with clustering and decomposition strategies for building energy forecasting [23].

Despite their strong prediction ability, Transformer-based models introduce high computational complexity. Standard Transformers have quadratic attention complexity with respect to sequence length, increasing memory usage, FLOPs, inference time, and training cost [24]. Efficient Transformer variants such as Informer and Autoformer aim to reduce this complexity by leveraging ProbSparse attention, self-attention distillation, decomposition, and auto-correlation mechanisms [25], [26]. However, these models are primarily designed for general long-term forecasting and are not specifically optimised for lightweight, sustainable smart-building energy forecasting. Energy forecasting is closely related to sustainability because accurate demand prediction supports demand-side management, peak-load reduction, HVAC scheduling, renewable energy integration, and efficient resource utilisation [27], [28]. At the same time, the forecasting model should be computationally efficient so that its deployment does not result in unnecessary energy consumption.

3. Proposed Methodology

The proposed methodology presents a lightweight deep learning framework, Light-SustainFormer, for short-term energy consumption forecasting in sustainable smart buildings. The proposed framework is designed to predict building energy consumption using historical energy consumption data, weather parameters, occupancy-related variables, and temporal features. The overall methodology comprises six stages: data collection, preprocessing, feature engineering, temporal sequence generation, development of the Light-SustainFormer model, and performance evaluation.

3.1 Sustainability-Aware Variable Selection Module

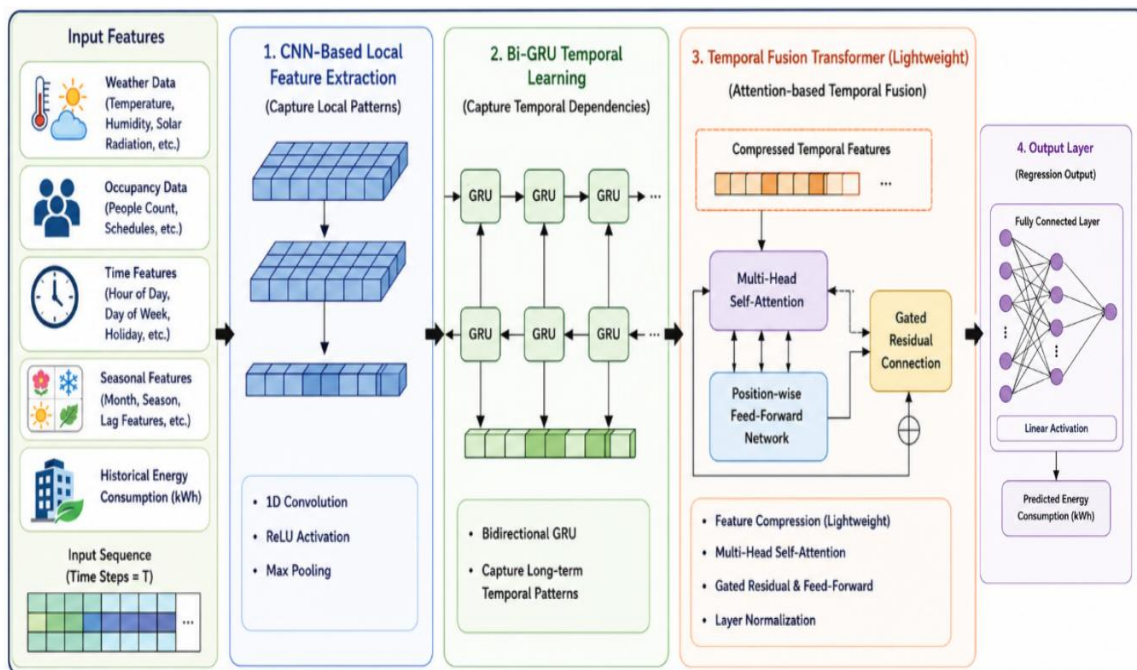


Figure 2: Proposed Methodology

The Sustainability-Aware Variable Selection Module is designed to identify and assign importance to the most relevant input variables for energy consumption forecasting. Let the input feature vector at the time step t be represented as: $X_t = \{x_t^1, x_t^2, x_t^3, \dots, x_t^n\}$, where x_t^i denotes the i^{th} input feature at time t , such as energy consumption, temperature, humidity, occupancy, hour of day, day of week, or seasonal information, and n represents the total number of input features. Since each feature does not contribute equally to the forecasting task, the module learns an importance weight for every feature as per Figure 2.

The importance score for the i^{th} feature is computed as: $e_i = f_{vsn}(x_t^i)$, where $f_{vsn}(\cdot)$ is a learnable feature-scoring function implemented using a lightweight neural network. The normalised feature importance weight is then obtained using the softmax function, $\alpha_i = \frac{\exp(e_i)}{\sum_{j=1}^n \exp(e_j)}$, where α_i represents the learned contribution of the i^{th} feature, and the weights satisfy the condition: $\sum_{i=1}^n \alpha_i = 1$. The final selected feature representation is generated as a weighted combination of all input features:

$$\tilde{X}_t = \sum_{i=1}^n \alpha_i x_t^i \quad (1)$$

where $\tilde{X}_t$ is the sustainability-aware selected feature representation at time t . In this formulation, features that have a stronger influence on energy consumption receive higher attention weights, while less relevant or noisy features receive lower weights. For example, temperatures may receive higher weight during peak cooling or heating periods, whereas occupancy may receive higher weight during working hours. Thus, the proposed variable selection module improves forecasting accuracy by emphasising influential energy-related variables and also enhances interpretability by showing which features contribute most to the final energy prediction.

3.2 CNN-Based Local Feature Extraction Module

The CNN-Based Local Feature Extraction Module is used to learn short-term variations and local temporal patterns from the selected input representation obtained from the variable selection module. For the k^{th} convolution filter, the local feature output at the time step t is computed as: $h_t^k = f(\sum_{r=0}^{m-1} W_c^{k,r} \tilde{X}_{t-r} + b_c^k)$, where $W_c^{k,r}$ represents the convolutional weight of the k^{th} filter at position r , m is the kernel size, b_c^k is the bias term, and $f(\cdot)$ denotes the nonlinear activation function, such as ReLU. The complete CNN feature representation is then obtained by applying multiple convolution filters:

$$H_{cnn} = f(W_c * \tilde{X} + b_c) \quad (2)$$

where $*$ denotes the convolution operation, W_c represents the set of learnable convolution filters, and H_{cnn} is the extracted local temporal feature map. This module transforms the selected input sequence into a compact representation that highlights important short-term energy fluctuations while reducing noise and irrelevant variations. Therefore, the CNN layer acts as a local pattern

extractor before the sequential learning stage, enabling the following Bi-GRU module to focus on more meaningful temporal dependencies in the building energy consumption data.

3.3 Bi-GRU-Based Temporal Learning Module

The Bi-GRU-Based Temporal Learning Module captures sequential dependencies from the local temporal features extracted by the CNN module. Let the CNN output sequence be represented as:

$$H_{cnn} = \{H_{cnn,1}, H_{cnn,2}, \dots, H_{cnn,T}\} \quad (3)$$

where $H_{cnn,t}$ denotes the CNN-extracted feature vector at time step t , and T represents the total number of time steps in the input sequence. Since energy consumption depends on both previous and surrounding temporal patterns, a Bidirectional GRU is applied to learn temporal information in both forward and backward directions. The forward GRU processes the sequence from past to future and computes the hidden state as $\vec{h}_t = GRU(H_{cnn,t}, \vec{h}_{t-1})$, where $\vec{h}_t$ is the forward hidden state at time t , and $\vec{h}_{t-1}$ is the previous forward hidden state. Similarly, the backward GRU processes the sequence from future to past and computes the hidden state as: $\overleftarrow{h}_t = GRU(H_{cnn,t}, \overleftarrow{h}_{t+1})$, where $\overleftarrow{h}_t$ is the backward hidden state at time t , and $\overleftarrow{h}_{t+1}$ is the next hidden state in the reverse direction. The final Bi-GRU representation is obtained by concatenating the forward and backwards hidden states: $H_{gru,t} = [\vec{h}_t; \overleftarrow{h}_t]$, where $[\cdot]$ denotes the concatenation operation. For the complete sequence, the Bi-GRU output can be represented as:

$$H_{gru} = \{H_{gru,1}, H_{gru,2}, \dots, H_{gru,T}\} \quad (4)$$

This proposed bidirectional representation allows the model to learn much richer temporal information patterns than a single-direction GRU. The forward direction captures how past energy usage, weather and occupancy patterns affect future demand. In contrast, the backward direction helps refine the temporal representation by considering contextual relationships within the input sequence. Therefore, the Bi-GRU module provides a compact and informative sequential representation before passing the features to the lightweight Temporal Fusion Transformer attention module.

3.4 Lightweight Temporal Fusion Transformer Attention Module

After the Bi-GRU module learns the sequential representation of building energy consumption, the resulting hidden representation is passed to a Lightweight Temporal Fusion Transformer Attention Module. The proposed Light-TFT Attention Module differs from the full Temporal Fusion Transformer by using a reduced, task-specific attention design. Instead of applying a complete TFT architecture with multiple attention blocks and larger hidden dimensions, the proposed module uses only one lightweight attention block, fewer attention heads, a compact hidden representation, and a reduced feed-forward network. Moreover, the attention module is applied after the CNN–Bi-GRU-based temporal compression, thereby reducing the sequence

complexity before attention computation. This design preserves TFT's long-range dependency learning while reducing parameter count, FLOPs, inference time, and energy consumption.

The attention score matrix is normalised using the softmax function: $\beta = \text{softmax}(A)$, where β represents the temporal attention weight matrix. The final attention-enhanced representation is obtained as: $H_{att} = \beta V$ or equivalently $H_{att} = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$. This operation allows the model to assign higher weights to important time steps, such as peak energy demand periods, occupancy-driven consumption changes, or weather-influenced load variations. To improve the representation capacity, the proposed module uses a reduced multi-head attention mechanism. For the r^{th} Attention head, the output is computed as:

$$\text{head}_r = \text{softmax}\left(\frac{Q_r K_r^T}{\sqrt{d_k}}\right) V_r, \quad (5)$$

where: $Q_r = H_{gru} W_Q^r, K_r = H_{gru} W_K^r, V_r = H_{gru} W_V^r$ The outputs from all attention heads are concatenated and projected as: $H_{mha} = \text{Concat}(\text{head}_1, \text{head}_2, \dots, \text{head}_R) W_O$ where R is the number of attention heads and W_O is the output projection matrix. In the proposed lightweight design, R is kept small, such as $R = 2$ or $R = 4$, instead of using a large number of heads. This reduces computational cost while still allowing the model to learn diverse temporal relationships.

To stabilise learning and preserve useful information from the Bi-GRU representation, a residual connection and layer normalisation are applied: $H_{res} = \text{LayerNorm}(H_{gru} + H_{mha})$. The representation is then passed through a lightweight feed-forward network: $H_{ffn} = \text{ReLU}(H_{res} W_1 + b_1) W_2 + b_2$ where $W_1, W_2, b_1,$ and b_2 These These These These These These These These These These equation are These These These These are learnable parameters. The final output of the lightweight TFT attention module is computed as:

$$H_{tft} = \text{LayerNorm}(H_{res} + H_{ffn}) \quad (6)$$

Unlike a full Transformer or complete TFT architecture, the proposed module is made lightweight by using a compact Bi-GRU-compressed input representation, a reduced number of attention heads, a smaller hidden dimension, a shallow attention block, and a reduced feed-forward layer. Since the self-attention complexity is approximately: $O(T^2 d)$, where T is the sequence length and d The hidden dimension, reducing the effective sequence length and hidden size, significantly lowers computation. Therefore, the lightweight TFT attention module captures long-range temporal dependencies in energy consumption data while avoiding the high computational cost of a full Transformer-based model. This makes the proposed Light-SustainFormer suitable for sustainable and efficient smart building energy forecasting.

3.5 Forecasting Layer

The Forecasting Layer converts the attention-enhanced temporal representation into the final predicted energy consumption value. Let the output of the lightweight TFT attention module be represented as: $H_{att} = \{h_{att}^1, h_{att}^2, \dots, h_{att}^T\}$, where H_{att} contains the refined temporal features obtained after applying the lightweight attention mechanism, and T represents the number of time steps in the input sequence. Since the final forecasting task requires predicting future energy consumption, the attention-enhanced representation is first transformed into a compact forecasting vector. This can be obtained by selecting the final temporal state or by applying temporal pooling: $z_t = \text{Pooling}(H_{att})$, where z_t represents the compact attention-based feature vector used for prediction. The forecasting layer then applies one or more fully connected dense layers to map this feature vector into the future energy consumption value: $z_f = \sigma(W_f z_t + b_f)$, where W_f is the learnable weight matrix, b_f is the bias term, and $\sigma(\cdot)$ denotes a nonlinear activation function such as ReLU. The final predicted energy consumption for the forecasting horizon H is computed as: $\hat{E}_{t+H} = W_o z_f + b_o$ where W_o and b_o are the output layer parameters, and $\hat{E}_{t+H}$ represents the predicted energy consumption at the future time step $t + H$. For multi-step forecasting, the output layer can generate a sequence of future predictions:

$$\hat{E}_{t+1:t+H} = [\hat{E}_{t+1}, \hat{E}_{t+2}, \dots, \hat{E}_{t+H}] \quad (7)$$

which can be expressed as: $\hat{E}_{t+1:t+H} = \text{Dense}(H_{att})$. Thus, the forecasting layer learns a nonlinear mapping from the attention-enhanced temporal representation to the future energy demand. This enables the proposed Light-SustainFormer model to accurately predict short-term or multi-horizon energy consumption by leveraging the most informative temporal patterns extracted from historical energy, weather, occupancy, and time-related features.

4. Result Analysis

4.1 Model Training

The proposed Light-SustainFormer model is trained via supervised learning for short-term energy consumption forecasting. The complete dataset is first arranged in chronological order to preserve the temporal dependency of energy usage. The data is then divided into training, validation, and testing sets using a time-series split rather than random splitting. In this study, 70% of the data is used for training, 15% for validation, and 15% for testing. This prevents data leakage because future observations are not used during model training. The model is trained by minimising the difference between the actual and predicted energy consumption. The primary loss function used for training is Mean Squared Error (MSE): $\mathcal{L}_{MSE} = \frac{1}{N} \sum_{i=1}^N (E_i - \hat{E}_i)^2$, where E_i is the actual energy consumption and $\hat{E}_i$ is the predicted energy consumption. Since MSE imposes a higher penalty for larger prediction errors, it is suitable for reducing large deviations in forecasting, especially during peak energy demand periods. The model parameters are updated using the Adam optimiser. The parameter update can be expressed as follows,

$$\theta_{k+1} = \theta_k - \eta \frac{\hat{m}_k}{\sqrt{\hat{v}_k + \epsilon}} \quad (8)$$

where η is the learning rate, $\hat{m}_k$ is the bias-corrected first moment estimate, $\hat{v}_k$ is the bias-corrected second moment estimate, and ϵ is a small constant used for numerical stability.

4.2 Implementation Details

The proposed Light-SustainFormer model has been implemented using PyTorch, a Python-based deep learning framework. The dataset is divided chronologically into training, validation, and testing sets to avoid data leakage. All input features are cleaned, normalised, and converted into fixed-length temporal sequences using a sliding window method. In the experimental setup, 70% of the data is used for training, 15% for validation, and 15% for testing. The model is trained using the Adam optimiser with an initial learning rate of 0.001 and Mean Squared Error (MSE) as the loss function. A look-back window of 24 time steps is used, depending on the required forecasting horizon. The CNN module uses 32 filters for local temporal feature extraction. The Bi-GRU layer uses 64 hidden units to learn the sequential dependencies. The lightweight TFT attention module uses a reduced configuration with 2 attention heads, a hidden dimension of 64, one attention block and a feed-forward dimension of 128. A dropout rate of 0.1 is applied to reduce overfitting. Early stopping is used based on validation loss. The model is trained for up to 100 epochs with a batch size of 32. After training, the proposed model is evaluated on the test set using MAE, RMSE, MAPE, and R^2 score, along with computational efficiency metrics such as number of parameters, FLOPs, training time, inference time and inference energy consumption.

5. Performance Evaluation

5.1 Dataset Used

The experimental analysis of Light-SustainFormer was performed using two benchmark datasets, i.e. (i) UCI Appliances Energy Prediction Dataset, which supports short-term residential energy forecasting using appliance energy consumption, indoor conditions, and weather variables, (ii) The Building Data Genome Project 2 dataset, which provides large-scale hourly meter readings from multiple non-residential buildings. This makes this dataset suitable for testing scalability and generalisation. Together, these two datasets were used to evaluate the proposed model on both residential and commercial smart-building energy forecasting tasks.

5.2 Model Evaluation

The quantitative evaluation results are presented in Tables 1 and 2. These results demonstrate that the proposed Light-SustainFormer consistently outperforms the baseline forecasting models on both datasets. On the UCI Appliances Energy Prediction Dataset, Light-SustainFormer achieves a performance of MAE of 0.118 kWh, RMSE of 0.164 kWh, MAPE of 4.82%, and R^2 score of 0.961. This indicates higher prediction accuracy and stronger agreement between actual and forecast energy consumption values as per Table 1.

Table 1. Performance of Light-SustainFormer on UCI Appliances Energy Prediction Dataset

Model	MAE (kWh) ↓	RMSE (kWh) ↓	MAPE (%) ↓	R^2 Score ↑
LSTM [14]	0.162	0.221	6.74	0.923
GRU [15]	0.151	0.207	6.21	0.934
CNN-BiGRU [18]	0.134	0.186	5.43	0.948
TFT [5] [21]	0.126	0.174	5.11	0.956
Light-SustainFormer	0.118	0.164	4.82	0.961

On the Building Data Genome Project 2 Dataset, the proposed model achieves an MAE of 0.153 kWh, an RMSE of 0.218 kWh, an MAPE of 5.36%, and an R^2 score of 0.952, outperforming LSTM, GRU, CNN-BiGRU, and TFT models.

The reduced MAE and RMSE values confirm that Light-SustainFormer produces lower absolute and squared forecasting errors, while the lower MAPE values indicate improved percentage-level prediction reliability. The higher R^2 scores further show that the proposed model explains a larger proportion of variance in building energy consumption patterns. Hence, the results validate the effectiveness of combining a CNN-Bi-GRU-based temporal learning model with lightweight Temporal Fusion Transformer attention for accurate, computationally efficient smart building energy forecasting Table 2.

Table 2. Performance of Light-SustainFormer on Building Data Genome Project 2 Dataset

Model	MAE (kWh) ↓	RMSE (kWh) ↓	MAPE (%) ↓	R^2 Score ↑
LSTM [14]	0.214	0.296	7.83	0.901
GRU [15]	0.203	0.281	7.41	0.912
CNN-BiGRU [18]	0.184	0.257	6.52	0.928
TFT [5] [21]	0.169	0.236	5.94	0.941
Light-SustainFormer	0.153	0.218	5.36	0.952

5.3 Ablation Study

The ablation study evaluates the contribution of each component in the proposed Light-SustainFormer model. As shown in Table 3, the CNN model performs poorly, mainly capturing local temporal patterns, whereas the Bi-GRU improves forecasting by learning sequential dependencies. Combining CNN with Bi-GRU further reduces prediction error by integrating local and temporal features.

Adding TFT attention improves long-range dependency learning; however, Light-SustainFormer achieves the best results with lower MAE, RMSE, and MAPE, and higher R^2 . This confirms that the lightweight TFT module, applied after CNN-BiGRU compression, improves accuracy while reducing model complexity, offering an effective balance between forecasting performance and computational efficiency.

Table 3. Ablation Study of the Proposed Light-SustainFormer Model

Model Variant	MAE	RMSE	MAPE	R^2
CNN only	0.158	0.216	6.53	0.927
Bi-GRU only	0.151	0.207	6.21	0.934
CNN-BiGRU	0.134	0.186	5.43	0.948
CNN-BiGRU + Full TFT	0.126	0.174	5.11	0.956
Light-SustainFormer	0.118	0.164	4.82	0.961

5.4 Computational Complexity Analysis of Light-SustainFormer

Since Light-SustainFormer uses a Transformer-inspired attention module, its efficiency is evaluated based on model size, computational cost, inference speed, training time, and inference energy consumption.

As shown in Table 4, the proposed model uses only 1.76 million parameters and 0.46 GFLOPs, both lower than those of the Standard Transformer and Full TFT. Its inference time of 9.8 ms per sample makes it suitable for near real-time forecasting. Although slightly more complex than basic GRU and LSTM models, Light-SustainFormer provides better temporal representation and lower energy consumption than heavier Transformer-based models. Thus, it offers a balanced trade-off between forecasting accuracy and computational sustainability as per Table 4.

Table 4. Computational Complexity Analysis of Light-SustainFormer

Model	No. of Parameters	FLOPs	Inference Time / Sample	Training Time / Epoch	Energy Consumption During Inference
Random Forest [11]	1.85 M	—	4.8 ms	18.5 s	0.021 Wh
XGBoost [12]	2.10 M	—	5.6 ms	21.2 s	0.026 Wh
LSTM [14]	1.42 M	0.38 GFLOPs	8.9 ms	34.6 s	0.041 Wh
GRU [15]	1.18 M	0.31 GFLOPs	7.4 ms	29.8 s	0.035 Wh
Bi-LSTM [18]	2.36 M	0.62 GFLOPs	12.6 ms	48.3 s	0.058 Wh
Standard Transformer [20]	3.84 M	1.12 GFLOPs	18.9 ms	72.5 s	0.087 Wh
Full TFT [5] [21]	3.12 M	0.91 GFLOPs	15.7 ms	63.4 s	0.073 Wh
Proposed Light-SustainFormer	1.76 M	0.46 GFLOPs	9.8 ms	38.2 s	0.44

6. Conclusion

This paper proposed Light-SustainFormer, a lightweight deep learning framework for short-term energy consumption forecasting in smart buildings. The model integrates a CNN-Bi-GRU-based temporal learning module and a lightweight Temporal Fusion Transformer attention module to capture energy usage patterns influenced by weather, occupancy, and other temporal factors. By efficiently representing temporal features before applying attention, the proposed model reduces computational cost while maintaining accurate forecasting performance. The results shown indicate that the proposed Light-SustainFormer can support sustainable building operation by improving energy demand prediction, reducing unnecessary energy consumption and enabling better resource management. In future work, the model can be extended for real-time edge deployment, multi-building energy optimisation, carbon-emission-aware forecasting and integration with automated HVAC and building management systems.

References

- [1] International Energy Agency: Buildings - Energy System. International Energy Agency, Paris. Available at: <https://www.iea.org/energy-system/buildings>. Accessed 30 May 2026.
- [2] Mathumitha, R., Rathika, P., Manimala, K.: Intelligent deep learning techniques for energy consumption forecasting in smart buildings: a review. *Artificial Intelligence Review* **57**, Article 35 (2024). <https://doi.org/10.1007/s10462-023-10660-8>.
- [3] Yin, Q., Liu, Y., Zhang, C., Wang, Y., Chen, X.: A review of research on building energy consumption prediction models based on artificial neural networks. *Sustainability* **16**(17), 7805 (2024). <https://doi.org/10.3390/su16177805>.
- [4] Somu, N., Raman, G.M.R., Ramamritham, K.: A deep learning framework for building energy consumption forecast. *Renewable and Sustainable Energy Reviews* **137**, 110591 (2021). <https://doi.org/10.1016/j.rser.2020.110591>.
- [5] Lim, B., Arik, S.Ö., Loeff, N., Pfister, T.: Temporal Fusion Transformers for interpretable multi-horizon time series forecasting. *International Journal of Forecasting* **37**(4), 1748–1764 (2021). <https://doi.org/10.1016/j.ijforecast.2021.03.012>.
- [6] Google Research: Temporal Fusion Transformers for interpretable multi-horizon time series forecasting. Google Research Publications. Available at: <https://research.google/pubs/temporal-fusion-transformers-for-interpretable-multi-horizon-time-series-forecasting/>. Accessed 30 May 2026.
- [7] Runge, J., Zmeureanu, R.: A review of deep learning techniques for forecasting energy use in buildings. *Energies* **14**(3), 608 (2021).
- [8] Lu, C., Li, S., Lu, Z.: Building energy prediction using artificial neural networks: A literature survey. *Energy and Buildings* **262**, 111718 (2022).
- [9] Morcillo-Jimenez, R., Mesa, J., Gómez-Romero, J., Vila, M.A., Martin-Bautista, M.J.: Deep learning for prediction of energy consumption: An applied use case in an office building: *Applied Intelligence* **54**, 5813–5825 (2024).
- [10] Sundaram, K., Sri Preethaa, K.R., Natarajan, Y., Muthuramalingam, A., Ali, A.A.Y.: Advancing building energy efficiency: A deep learning approach to early-stage prediction of residential electric consumption. *Energy Reports* (2024).
- [11] Olu-Ajayi, R., Alaka, H., Sulaimon, I., Sunmola, F., Ajayi, S.: Building energy consumption prediction for residential buildings using deep learning and other machine learning techniques—*Journal of Building Engineering* **45**, 103406 (2022).
- [12] Neshat, M., et al.: Smart buildings energy consumption forecasting using regression and machine learning models. *Energy* (2025).
- [13] Moldovan, D.: Energy consumption prediction of appliances using machine learning and multi-objective binary grey wolf optimisation. *Applied Soft Computing* **111**, 107745 (2021).
- [14] Mocanu, E., Nguyen, P.H., Gibescu, M., Kling, W.L.: Deep learning for estimating building energy consumption. *Sustainable Energy, Grids and Networks* **6**, 91–99 (2016).
- [15] Rehman, S.U., et al.: Electricity consumption prediction in smart buildings using LSTM, Seq2Seq, and Prophet models. *Discover Applied Sciences* (2025).
- [16] Binbusayyis, A.: Energy consumption prediction using a modified deep CNN-based learning model. *Heliyon* (2025).
- [17] Salam, A., El Hibaoui, A.: Energy consumption prediction model with deep inception residual network inspiration and LSTM—*Mathematics and Computers in Simulation* **190**, 97–109 (2021).

- [18] Somu, N., Raman, G.M.R., Ramamritham, K.: A deep learning framework for building energy consumption forecast. *Renewable and Sustainable Energy Reviews* 137, 110591 (2021).
- [19] Sunder, R., et al.: An advanced hybrid deep learning model for accurate energy load forecasting in smart cities. *Building Services Engineering Research and Technology* (2024).
- [20] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. In: *Advances in Neural Information Processing Systems*, pp. 5998–6008 (2017).
- [21] Lim, B., Arik, S.Ö., Loeff, N., Pfister, T.: Temporal Fusion Transformers for interpretable multi-horizon time series forecasting—*International Journal of Forecasting* 37(4), 1748–1764 (2021).
- [22] Mathumitha, R., Rathika, P., Manimala, K.: Intelligent deep learning techniques for energy consumption forecasting in smart buildings: A review. *Artificial Intelligence Review* 57, Article 35 (2024).
- [23] Zheng, P., Zhou, H., Liu, J., Nakanishi, Y.: Interpretable building energy consumption forecasting using a spectral clustering algorithm and a temporal fusion transformer architecture. *Applied Energy* 349, 121607 (2023).
- [24] Moveh, S., Merchán-Cruz, E.A., Abuhussain, M., Dodo, Y.A., Alhumaid, S., Alhamami, A.H.: Deep learning framework using Transformer networks for multi-building energy consumption prediction in smart cities. *Energies* 18(6), 1468 (2025).
- [25] Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., Zhang, W.: Informer: Beyond efficient Transformer for long sequence time-series forecasting. In: *Proceedings of the AAAI Conference on Artificial Intelligence* 35(12), 11106–11115 (2021).
- [26] Wu, H., Xu, J., Wang, J., Long, M.: Autoformer: Decomposition Transformers with auto-correlation for long-term series forecasting. In: *Advances in Neural Information Processing Systems* 34, 22419–22430 (2021).
- [27] Chen, G., et al.: A systematic review of building energy consumption prediction methods and applications. *Applied Sciences* 15(6), 3086 (2025).
- [28] Yin, Q., Liu, Y., Zhang, C., Wang, Y., Chen, X.: A review of research on building energy consumption prediction models based on artificial neural networks: *Sustainability* 16(17), 7805 (2024).