AI Engineering
https://www.sciltp.com/journals/aieng

Article

Building Electricity Load Forecasting Considering Climate Change Impacts: A Multi-Factor Deep Learning Approach

Yuqing Wang, Qiaoqiao Li and Yan Xu *

School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798, Singapore
* Correspondence: xuyan@ntu.edu.sg

How To Cite: Wang, Y.; Li, Q.; Xu, Y. Building Electricity Load Forecasting Considering Climate Change Impacts: A Multi-Factor Deep Learning Approach. *AI Engineering* **2025**, *1*(1), 3.

Received: 3 May 2025
Revised: 10 June 2025
Accepted: 26 June 2025
Published: 30 June 2025

Abstract: Accurate electricity load forecasting is crucial for efficient power system operation and planning, especially in urban environments where building energy dynamics play a significant role. This study focuses on improving building load forecasting by accounting for the differences between indoor and outdoor climate conditions, which significantly influence energy consumption patterns in buildings. In this paper, a comprehensive comparative study was conducted to assess the performance of several deep neural network models, with multi-factors incorporated to enhance forecast performance. The input features considered in this paper include the input sequence structures, lagged correlation-based temperature conditions, calendar information and categorical Building-Adjusted Internal Temperature (BAIT) index. The proposed approach innovatively integrates the disparities between indoor and outdoor climatic conditions to better capture the complexities of energy demand influenced by building characteristics and external weather conditions. The proposed method was tested on Singapore EMA load data from 2019 to 2022. The experimental outcomes demonstrate that the BAIT-enhanced models exhibit strong performance across various metrics, which yields superior predictions.

Keywords: electricity load forecasting; building-adjusted internal temperature; deep learning; long short-term memory; transformer; climatic variables

1. Introduction

Short-term power load forecasting is crucial for effective power system operation and control, ensuring reliable power supply while enhancing the efficiency of power generation, transmission, and distribution. However, precise electricity consumption prediction is challenging due to its susceptibility to a range of factors, including weather conditions, economic activities, and societal routines [1].

The building and construction sector stands as one of the largest consumers of energy in the world, contributing to 25–40% of the total load [2]. The forecasting of building energy consumption is inherently complex, requiring the consideration of numerous factors, including climatic conditions, occupancy patterns, and energy usage behaviors. In Singapore, the consistently hot climate throughout the year results in most energy-intensive activities occurring indoors, significantly impacting the overall electricity load profile. In reality, the disparities between indoor and outdoor climatic conditions can significantly influence the energy demands of buildings [3]. Therefore, both indoor-adjusted and outdoor climate variables are essential for developing accurate load forecasting models that capture the nuanced interactions affecting building energy consumption.

In recent years, advancements in neural network architectures have driven substantial improvements in electric load forecasting, classical recurrent approaches were first applied to handle electricity load data. Recurrent Neural Networks (RNNs), especially Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU), are widely used for electric load forecasting due to their ability to capture temporal dependencies. Studies have shown that LSTM models outperform traditional statistical methods by effectively learning complex, non-linear relationships, enhancing accuracy [4,5]. GRU, with its simpler architecture and shorter training time, achieves comparable performance to LSTM in load forecasting [6,7]. More recently, advances in transformer-based models, have significantly improved forecasting accuracy by effectively capturing long-range temporal dependencies and complex variable interactions. Among these, Inverted Transformer (iTransformer) stands out due to its innovative approach of utilizing channels as tokens, rather than the traditional positional encoding, enabling the efficient handling of multivariate time series data through improved computational efficiency and representation learning capability [8]. The success of iTransformer has inspired various derivative models, including iFlashformer, iFlowformer, iInformer, and iReformer. Specifically, iFlashformer incorporates Flash Attention mechanisms to drastically reduce computational complexity and enhance scalability for extended sequence lengths [9]. iFlowformer integrates flow-based attention strategies to dynamically capture and represent the non-stationary interactions typical in climatic and electricity load data [10]. The iInformer model utilizes the ProbSparse self-attention mechanism to prioritize more informative data segments, significantly reducing redundancy and computational overhead [11]. Additionally, iReformer applies locality-sensitive hashing (LSH) attention to efficiently approximate full attention, effectively capturing local and global patterns in extensive sequence data with reduced computational demand [12]. In most current studies, outdoor meteorological data—including temperature, humidity, wind speed, and precipitation rate—are directly incorporated into load forecasting models [13].

Overall, these models demonstrate substantial improvements over classical approaches by capturing both global and local patterns in time series data, and their integration into multivariate short-term forecasting frameworks has led to state-of-the-art results across various application domains. However, such methods that incorporate raw weather data often fail to adequately account for the effect of environmental conditions on indoor temperatures, which is crucial for understanding heating and cooling demands. The Building-Adjusted Internal Temperature (BAIT) index was proposed in [14] to encapsulate these factors, offering a more accurate representation of indoor conditions.

In this paper, a multi-factor data-driven method is proposed considering the indoor climate correlations to improve short-term load forecasting, where several deep learning architectures are evaluated on multivariate single and multi-step forecasting tasks against a detailed representation of indoor conditions and other factors to effectively delineate the impact of building-specific climate dynamics on energy usage. The analysis was conducted utilizing electricity consumption data from Singapore spanning 2019 to 2022, with a four-fold cross-validation approach to substantiate the robustness of the findings. The results of this study verify the advantages of integrating indoor and outdoor climatic factors into deep learning frameworks for enhancing power load prediction. This approach significantly enhances the models' capacity to simulate energy demand variability by capturing the distinctive climatic dynamics affecting buildings. Traditional short-term load forecasting studies usually feed raw outdoor weather variables or simple thermal indices such as Cooling-Degree-Days (CDD) and Effective Temperature (ET) into neural networks. These proxies implicitly assume an instantaneous equilibrium between outdoor and indoor conditions, thereby ignoring the thermal inertia of building envelopes. The Building-Adjusted Internal Temperature (BAIT) proposed by Staffell et al. (2023) successfully captures such inertia, yet its application has hitherto been confined to macro-scale heating and cooling demand assessment. This work is the first to operationalize BAIT for hourly building-level electricity load forecasting and to embed it in data-driven deep-learning frameworks.

The rest of the paper is organized as follows. Firstly, Figure 1 illustrates the overall workflow, where pre-processing (top) converts raw data into an $11 \times h$ feature matrix before training the deep-learning models (bottom). The input matrix contains traditional factors discussed in Section 2.1 and BAIT index introduced in Section 2.2. Then, the input matrix is fed into deep neural networks for training. After that, multivariate single-step and multi-step forecasting tasks are described in Sections 2.3 and 2.4. Finally, all the results and discussion can be found in the case study.

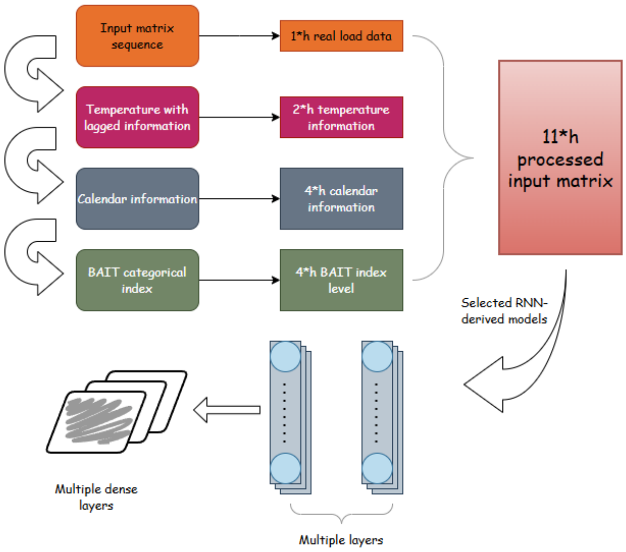

Figure 1. Framework of the proposed method (Blue blocks denote conventional features, orange blocks denote BAIT-enhanced features).

2. Proposed Method

In this study, different deep neural network models are used to predict the load incorporated with multi-factors. The inputs of the models include historical load, calendar information and the BAIT index. A comprehensive comparative study is then conducted with BAIT index replacing weather data. Section 2 outlines how raw data are transformed into model-ready features, starting with traditional factors (Section 2.1) and then the proposed BAIT index (Section 2.2).

2.1. Traditional Factors

As displayed in Figure 1, the input matrix is created through four stages: input sequence length selection, consideration of preceding temperature conditions, classification of calendar information and the BAIT index, which is introduced to replace outdoor meteorological data used in traditional methods [13]. This section contains the first three stages.

In this paper, the 96-h electricity load was chosen as the first-dimension input feature. This duration allows the model to learn inter-day relationships, which is essential for accurate electricity demand prediction.

Then calendar effect is discussed in detail. Significant fluctuations in electricity usage on Mondays and Saturdays, are observed which is due to changes in human behavior. Mondays typically exhibit high energy loads due to the resumption of work and industrial activities after the weekend. In contrast, energy loads on Saturdays are lower, reflecting workplace closures and a shift in energy consumption patterns as more people stay at home. To capture these variations, Mondays and Saturdays are delineated in the input data. There is also a clear distinction between weekday and weekend energy consumption. Weekdays are dominated by commercial and industrial use, while weekends are characterized by residential consumption. During summer, higher temperatures lead to increased air conditioning use, especially during daylight hours, which boosts electricity demand. To incorporate these elements, weekend and summer indicators are also introduced, finally creating a $4 \times h$ calendar matrix that was integrated into the initial input matrix.

Finally, given the pivotal role of temperature in predicting power load demand, the analysis incorporates two rows within the input matrix to represent temperature conditions. The first row contains a sequence of temperatures recorded for each of the past h hours, formatted as a $1 \times h$ string. The second row represents the most correlated lagged temperature. The substantial impact of temperature fluctuations on electricity usage necessitates examining historical weather-electricity consumption patterns to identify the most pertinent temperature characteristics. Consequently, the second-row data is adjusted such that the final value corresponds to the temperature with the highest correlation, while the preceding values reflect time points prior to this.

2.2. BAIT Index Factor

Buildings-Adjusted Internal Temperature (BAIT) estimates perceived indoor temperature in buildings without heating or cooling systems. It incorporates natural conditions of temperature, solar irradiation, wind speed, and humidity to determine indoor climate conditions. Simplified, BAIT simulates the thermostat reading in a home without air conditioning, thereby elucidating the correlation between external climate conditions and a building's energy usage [14]. The calculation of BAIT involves integrating multiple factors that significantly influence the perception of indoor climate condition.

(1) Basic formula

The BAIT at each day (d) and grid location (l) is calculated as follows:

$$\mathrm{BAIT}_{d,l} = \mathrm{T} + x(\mathrm{S} - \mathrm{S}^*) - y(\mathrm{W} - \mathrm{W}^*) + z(\mathrm{H} - \mathrm{H}^*)(\mathrm{T} - \mathrm{T}^*) \tag{1}$$

where d represents days, l represents location, and BAIT represents the temperature in Celsius. BAIT is affected by four factors: T (outdoor air temperature), S (solar intensity), W (wind speed) and H (relative humidity). The coefficients x, y and z correspond to the effects of solar irradiation, wind speed and relative humidity on temperature changes. In addition, S*, W*, H*, and T* are the ideal air temperature, solar irradiation, wind speed, and humidity values to be used in the estimation.

(2) Smoothing for Thermal Inertia

Due to thermal inertia, indoor temperatures change slowly and do not immediately reflect outdoor weather variations. This is due to insulation and the building's structure. This phenomenon is modeled by smoothing the BAIT values over 48 h, reducing short-term fluctuations and their impact on perceived room temperature.

$$\mathrm{BAIT}_{d,l} = \frac{\mathrm{BAIT}_{d,l} + \sigma\mathrm{BAIT}_{d-1,l} + \sigma^2\mathrm{BAIT}_{d-2,l}}{1 + \sigma + \sigma^2} \tag{2}$$

In this context, σ is a smoothing parameter, usually between 0 and 1, which describes the influence of previous temperature data, assigning weights to previous days' temperatures.

(3) Blending with Raw Temperature

During hot weather, unadjusted temperature data may be analyzed to assess the effectiveness of cooling measures, such as ventilation and air conditioning. This approach was adopted to investigate the impact of these measures, including heat dissipation from opening windows.

$$\mathrm{B} = \frac{\mathrm{B}_{\max}}{1 + e^{-\mathrm{B}'}} \tag{3}$$

$$\mathrm{B}' = (\mathrm{T} - 0.5(\mathrm{B_U} - \mathrm{B_L}))\frac{10}{\mathrm{B_U} - \mathrm{B_L}} \tag{4}$$

In this model, B represents the mixing value and $\mathrm{B}_{\max}$ refers to the maximum mixing degree. The model also involves an input to a sigmoid curve, denoted as B′. The model sets boundaries: the highest is $\mathrm{B_U}$ = 23 °C, and the lowest is $\mathrm{B_L}$ = 15 °C. When $\mathrm{B_U}$ is reached or exceeded, the building's insulation effectiveness decreases, and human behavior changes, reducing the weather's impact on the building by half.

A sigmoid function is used to adjust the BAIT,

$$\mathrm{BAIT}_{d,l} = (\mathrm{BAIT}_{d,l}(1 - \mathrm{B}_{d,l})) + (\mathrm{T}_{d,l}\mathrm{B}_{d,l}) \tag{5}$$

$\mathrm{B}_{d,l}$ represents blending factors based on occupant behavior and specific environmental conditions, such as opening a window during warm weather.

The BAIT index is evaluated numerically but is categorized into four distinct levels, as illustrated in Table 1. Similar to the date factor, One-hot encoding is employed to enrich the date information. This method transforms the categorical BAIT data into a $4\times h$ matrix, which is then appended to the input matrix [15].

Finally, the input matrix, with a size of $11\times h$, is composed, where 11 represents the category number of calendar effects, and h denotes the length of historical input. Subsequently, this structured grid is fed into multi-layer RNN-derived forecast structures to predict power load. In summary, Sections 2.1 and 2.2 construct an $11\times h$ feature matrix that feeds the networks described next.

Table 1. bait levels (categorical index).

BAIT Level	BAIT Range	One-Hot Encoding
Very Low	(0, 26.738)	[1, 0, 0, 0]
Low	[26.738, 27.348)	[0, 1, 0, 0]
Moderate	[27.348, 27.850)	[0, 0, 1, 0]
High	[27.850, +∞)	[0, 0, 0, 1]

2.3. Multivariate Single-Step and RNN-Derived Forecasting Methods

The multivariate single-step forecasting task involves predicting electricity load for the immediate next time step based on multiple influencing variables recorded up to the current time step. This immediate next step can vary depending on the application context and data resolution, typically being half-hourly, hourly, or daily predictions. Formally, given historical multivariate time series data $\mathbf{X}_{t-h+1:t} \in \mathbb{R}^{h\times m}(h = 96, m = 11)$, where h is the length of the historical observation window and m represents the number of influencing factors (variables), the single-step forecasting objective is defined as:

$$\hat{x}_{t+1} = f(\mathbf{X}_{t-h+1:t}; \theta)$$

where $\hat{x}_{t+1} \in \mathbb{R}^m$ indicates the predicted electricity load and influencing factors at the subsequent single time step $t + 1$, and $f(\cdot; \theta)$ is the forecasting function parameterized by model parameters θ.

To examine the effect of BAIT index on multivariate single-step forecasting task, conventional recurrent neural network models are applied. Recurrent Neural Networks (RNNs) form the foundation for these forecasting tasks by processing sequential data and updating their hidden state at each time step according to the current input. Among these, Long Short-Term Memory (LSTM) networks mitigate the challenges of traditional RNNs by employing three gates—input, forget, and output—to regulate information flow over long sequences. The Gated Recurrent Unit (GRU) simplifies this process with a more streamlined architecture that uses fewer gates, thus improving computational efficiency while addressing the vanishing gradient problem. Additionally, bidirectional models like Bidirectional LSTM (Bi-LSTM) and Bidirectional GRU (Bi-GRU) further enhance performance by processing data in both forward and reverse directions, enabling them to capture contextual information from both past and future time points effectively [16].

2.4. Multivariate Multi-Step Forecasting and Transformer-Based Models

To further explore the effectiveness of the BAIT index in a multi-step forecasting context and investigate the interactions among multiple variables, we extend our analysis beyond conventional recurrent neural network models to advanced transformer-based models on multivariate multi-step forecasting task. Transformer-based architectures are selected due to their proven capability in capturing long-range dependencies and efficiently modeling interactions among multiple input variables, making them highly suitable for multi-step electricity load forecasting.

Multivariate multi-step forecasting extends this scenario to predict electricity load and relevant influencing factors across multiple future time steps. This prediction spans a forecast horizon of multiple consecutive time points, such as several hours or days ahead, offering greater utility for planning and management purposes. Given historical observations $\mathbf{X}_{t-h+1:t} \in \mathbb{R}^{h\times m}$, the multi-step forecasting problem is formally represented as:

$$\widehat{\mathbf{X}}_{t+1:t+M} = F(\mathbf{X}_{t-h+1:t}; \phi)$$

where $\widehat{\mathbf{X}}_{t+1:t+M} \in \mathbb{R}^{M\times m}$ is the predicted multivariate time series for the upcoming M time steps, and $F(\cdot; \phi)$ is a multi-step forecasting model parameterized by parameters ϕ. Here h (look-back window) is fixed to 96 for single-step and 336 for multi-step experiments, whereas m = 11 includes load, calendar flags and BAIT categories, as detailed in Sections 2.1 and 2.2.

In this study, we evaluate five transformer-based models derived from the Informer Transformer (iTransformer) framework. The baseline iTransformer utilizes sparse attention and distillation mechanisms for efficient long-sequence forecasting. Building on this foundation, the iFlashformer integrates Flash Attention to significantly reduce memory and computational complexity, enabling faster training and inference. The iFlowformer incorporates flow-based attention mechanisms to capture dynamic temporal correlations among climatic variables, historical load, and calendar features, thereby enhancing the modeling of non-stationary interactions. Furthermore, the iInformer employs a ProbSparse self-attention mechanism to prioritize informative queries, optimizing computational load for extended forecasting horizons. Finally, the iReformer leverages locality-sensitive hashing (LSH) attention to efficiently approximate full-attention, facilitating the capture of both local and global temporal dependencies.

Along with the single-step forecasting, these two frameworks provide a robust platform to examine the effects of varying input lengths and forecasting horizons on the accuracy and stability of each model. The comprehensive comparative analysis not only highlights the strengths and weaknesses of individual models in capturing multivariate temporal patterns but also offers critical insights into the role of the BAIT parameter in improving electricity load forecasting performance.

3. Case Study

The impact of the BAIT index on electricity demand predictions was investigated using a comparative experiment. Initially, a comprehensive input matrix was compiled, incorporating critical variables such as historical power grid usage, weather temperature, solar irradiation, wind speed and temporal patterns. Then for multivariate single-step forecasting task, four deep learning models—LSTM, Bi-LSTM, GRU, and Bi-GRU—were trained and evaluated. Next, the weather-related features (temperature, solar radiation, wind speed, humidity) were replaced with the categorical BAIT index, and the models were re-evaluated. After that, the same process was performed for multivariate multistep prediction where several SOTA models were chosen: iTransformer**,** iFlashformer**,** iFlowformer**,** iInformer and iReformer**.** To rigorously evaluate these models in a multi-step forecasting scenario, the input historical sequence length is set to 336 h (two weeks), with multiple forecast horizons considered, namely 92 h (4 days), 192 h (8 days), 336 h (14 days), and 720 h (30 days). This comprehensive range of prediction lengths facilitates an extensive assessment of the models' capability to handle short-term to very long-term forecasting scenarios.

Several metrics are used to evaluate the predictive model's effectiveness: mean absolute error (MAE), root mean square error (RMSE), mean absolute percentage error (MAPE), coefficient of determination (R^2), and mean square error (MSE). These metrics collectively provide a comprehensive assessment of the model's efficiency. This ensures a nuanced understanding of prediction accuracy.

This study employed data spanning four years (2019–2022), with a specific emphasis on electricity consumption patterns in Singapore. To evaluate the model's performance, a four-fold cross-validation approach is utilized.

A. Calendar information analysis by SHAP

SHapley **A**dditive ex**P**lanations (SHAP) is an interpretable method rooted in cooperative game theory, designed to quantify and visualize the impact of individual features on the predictions made by machine learning models. By calculating Shapley values, SHAP determines the contribution of each feature, effectively capturing their relative importance and interaction effects within the predictive model.

Figure 2 provides a detailed SHAP summary plot illustrating the influence of calendar-based features and environmental variables on model predictions for electricity demand. Each data point in the plot represents an individual observation, positioned horizontally according to its SHAP value—positive values indicate a positive impact on electricity demand predictions, while negative values reflect a reduction in predicted demand. The color gradient from blue to red conveys the magnitude of feature values, with red signifying higher values and blue indicating lower ones.

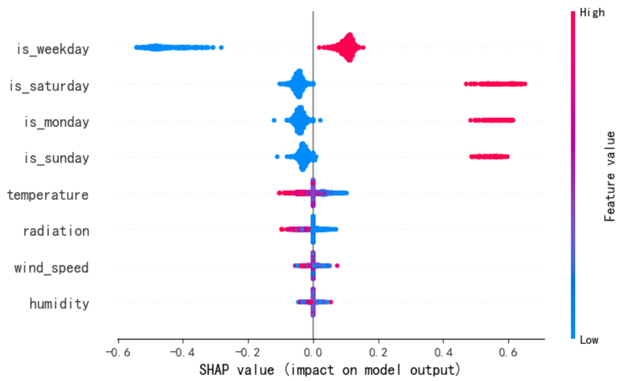

Figure 2. SHAP graph of the model with calendar info.

From the analysis depicted in Figure 2, it is evident that calendar information significantly impacts electricity demand predictions. Among these features, is_saturday and is_monday demonstrate the most substantial contributions. Specifically, the feature is_monday, associated with the commencement of the working week, displays a clear positive impact on electricity demand predictions. This observation aligns with the expected rise in industrial and commercial energy consumption as operations resume after the weekend. Conversely, is_saturday exhibits predominantly negative SHAP values, signifying a reduction in electricity demand consistent with reduced industrial and commercial activities during weekends.

In comparison, features such as is_weekday and is_sunday show relatively less significant influence, indicating that although these days contribute to variations in electricity consumption patterns, their effects are comparatively modest. Environmental factors such as temperature, radiation, wind_speed, and humidity display balanced distributions around zero, suggesting that their influence, while present, is context-dependent and does not dominate the prediction outcomes as clearly as calendar-related features.

Overall, this SHAP analysis effectively underscores the critical role of temporal and environmental features in influencing electricity demand forecasts, highlighting the importance of incorporating domain-specific calendar indicators and contextual environmental conditions into predictive modeling frameworks.

B. Temperature correlation analysis

Figure 3 illustrates the influence of temperature data from the previous 24 h on subsequent hourly power consumption. The X-axis represents the time lag (in hours), elucidating the correlation between temperature fluctuations and electricity demand several hours prior. The Y-axis denotes the correlation strength, and three curves correspond to distinct measurements. The blue curve depicts the direct relationship between temperature and electricity demand, peaking at an 11-h lag. The orange curve represents the R^2 coefficient from linear regression, following the trend of the blue curve but with a lower value. The green curve represents an R^2 fit derived from a quadratic equation, typically offering a more pertinent correlation than the orange curve.

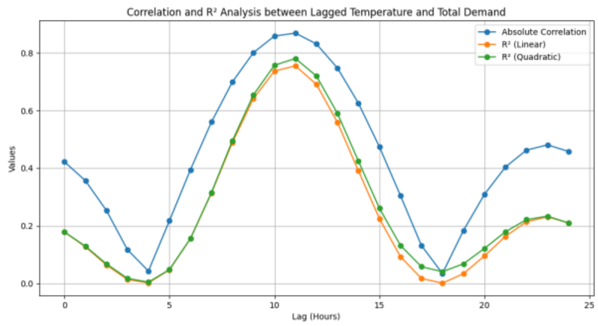

Figure 3. Correlation and R^2 analysis result between lagged temperature and real load.

The analysis reveals that temperature conditions from around 11 h earlier have the strongest correlation with electricity consumption. This may reflect delayed consumer responses to temperature changes for indoor climate adjustments. Identifying the most correlated scenarios significantly enhances the precision of electricity demand forecasts. Additionally, the graph highlights temperature readings from 11 h earlier.

C. Improve load forecasting with categorical BAIT index

Figure 4 presents the correlation between outdoor temperature and indoor BAIT for Singapore using a scatter plot. The dot color, ranging from light yellow to dark red, corresponds to the intensity of solar radiation. Same as the wind speed, ranging from light purple to dark purple, corresponds to the intensity of it. The plot clearly demonstrates the significant influence of outdoor temperature on indoor comfort. The black dotted line represents a 1:1 linear correlation, confirming outdoor temperature as a key factor affecting the indoor environment. Solar irradiation plays a crucial role. Strong solar irradiation (indicated by red) correlates with higher outdoor temperatures, leading to a concomitant rise in BAIT due to increased heat absorption and retention by building exteriors. Wind speed also impacts BAIT, though less directly. Mild winds (shown in dark purple) are often associated with higher temperatures, suggesting natural cooling is less effective during hot conditions.

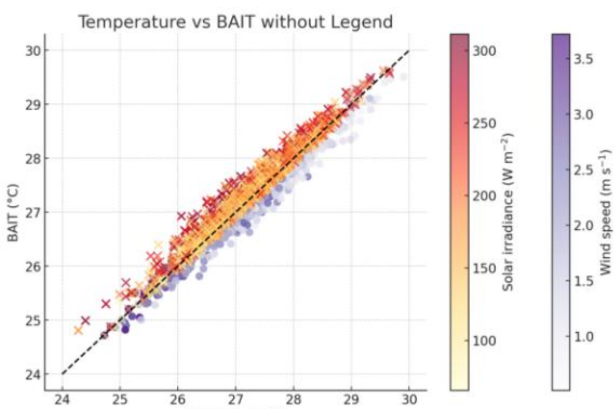

Figure 4. Influence of BAIT by solar irradiance and wind speed respect to real temperature.

Figure 5 denotes a bar chart depicting the distribution of BAIT values across all samples. As depicted in the figure, the majority of BAIT values are grouped within a specific range, indicating consistent internal temperatures across buildings of different ages. This consistency is likely due to Singapore's stable climate with minimal temperature variation.

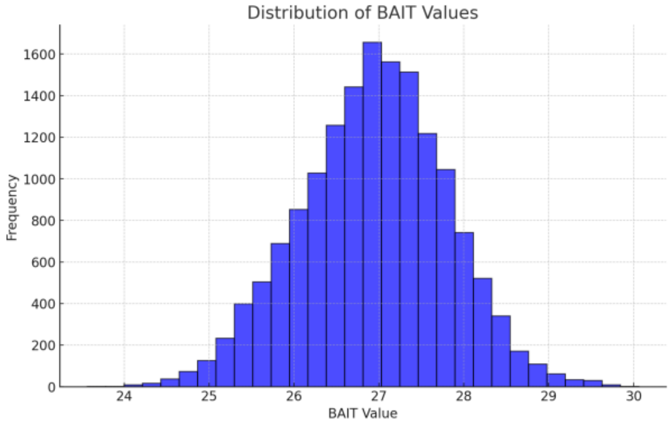

Figure 5. the distribution of the BAIT value on frequency.

D. Multivariate Single-step Forecasting Results

As is illustrated in Figure 6, the study revealed that the Bi-GRU model outperformed other models trained on conventional climate factors, including temperature, solar irradiation, wind speed, and humidity, in terms of all evaluation metrics.

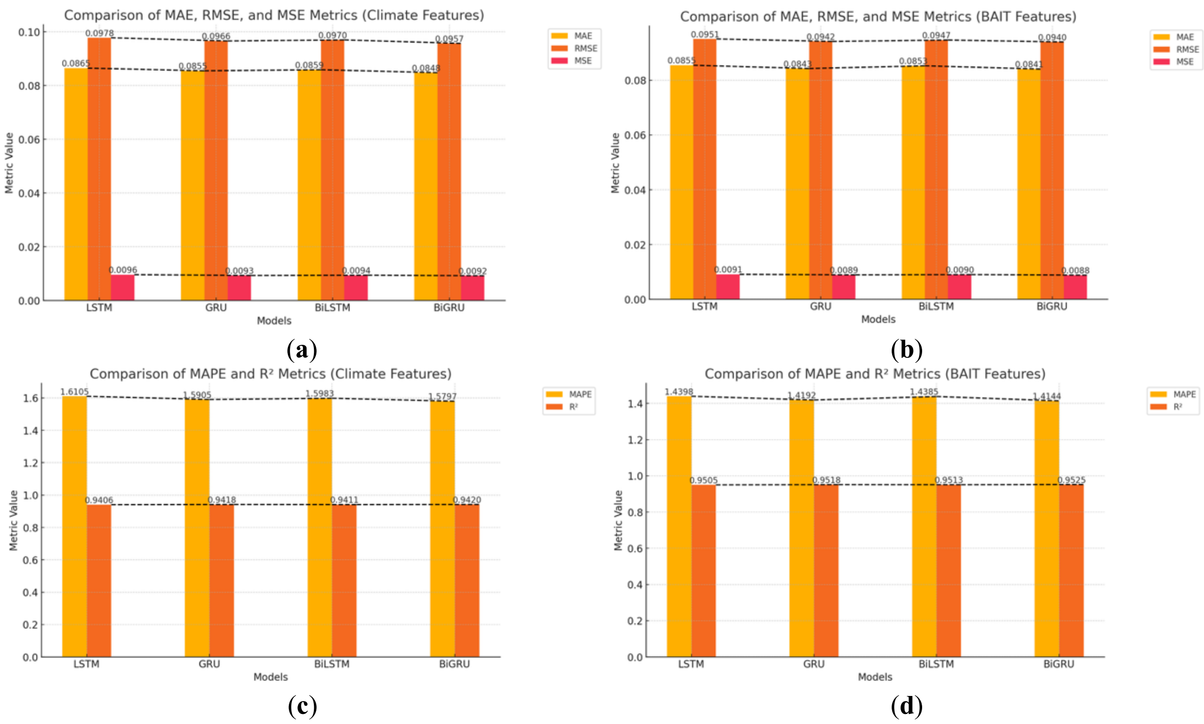

Figure 6. Performance from different models and different features applied. (**a**) MAE, RMSE and MSE with climate features (**b**) MAE, RMSE and MSE with BAIT features (**c**) MAPE and R^2 with climate features (**d**) MAPE and R^2 with BAIT features.

LSTM and Bi-LSTM models performed well but were slightly less accurate with a marginally higher error rate compared to GRU models. This trend is also evident when evaluating data based on BAIT index features. Bi-directional GRU (Bi-GRU) models consistently rank highest across multiple criteria, with GRU, Bi-dimensional GRU, and LSTM models following respectively. The GRU's simple structure efficiently handles BAIT features without the added complexity that can hinder LSTM performance.

The comparative outcome for each model is shown in Table 2. The application of our proposed categorical BAIT index led to improvements across all evaluation metrics. For LSTM models, MAE, RMSE, MAPE and MSE were decreased by 1.16%, 2.76%, 10.60% and 5.21%, and R^2 was increased by 1.04%. For GRU model, MAE, RMSE, MAPE and MSE were decreased by 1.40%, 2.48%, 10.77% and 4.30%, and R^2 was increased by 1.06%. For Bi-LSTM model, MAE, RMSE, MAPE and MSE were decreased by 0.70%, 2.37%, 10.00% and 4.44%, and R^2 was increased by 1.07%. For Bi-GRU model, MAE, RMSE, MAPE and MSE were decreased by 0.83%, 1.78%, 10.46% and 4.35%, and R^2 was increased by 1.10%.

Table 2. Comparative study of proposed method and baseline methods.

	LSTM		**GRU**		**Bi-LSTM**		**Bi-GRU**	
	Original	***BAIT***	***Original***	***BAIT***	***Original***	***BAIT***	***Original***	***BAIT***
MAE (GW)	0.0865	0.0855	0.0855	0.0843	0.0859	0.0853	0.0848	**0.0841**
RMSE (GW)	0.0978	0.0951	0.0966	0.0942	0.097	0.0947	0.0957	**0.0940**
MAPE (%)	1.6105	1.4398	1.5905	1.4192	1.5983	1.4385	1.5797	**1.4144**
R^2	0.9406	0.9505	0.9418	0.9518	0.9411	0.9513	0.9420	**0.9525**
MSE (GW^2)	0.0096	0.0091	0.0093	0.0089	0.0094	0.0090	0.0092	**0.0088**

E. Visualization Analysis

To provide deeper insights into the predictive capabilities of the models evaluated, visualizations for the predicted electricity load using the LSTM model are presented, both without and with the integration of the BAIT index, for comparative purposes.

Figure 7A,B illustrate the performance of the LSTM model without employing the BAIT index. Specifically, Figure 7A presents the predictions across the entirety of the first-year data. The predictions generally align with the observed load trends, but some noticeable deviations and fluctuations suggest the model's limitations in capturing intricate short-term variations. A more detailed view provided in Figure 7B, showing predictions for the final 1000 time samples of the first year, further highlights these discrepancies, particularly in periods exhibiting sharp fluctuations or unusual consumption patterns.

In contrast, Figures 7C,D demonstrate the improved predictive accuracy of the LSTM model when the BAIT index is incorporated. Figure 7C, covering the same full year as Figure 7A, exhibits significantly enhanced alignment between predictions and actual load data, demonstrating the efficacy of the BAIT index in capturing the nuanced climatic influences on electricity demand. Figure 7D reinforces these findings through a closer examination of the last 1000 time steps of the same year, where the BAIT-enhanced predictions substantially reduce deviations and more accurately capture fluctuations, especially during critical demand periods.

Overall, these visualizations underscore the practical advantages and predictive improvements achieved through the integration of the BAIT index into forecasting models. They vividly illustrate the index's capability to encapsulate complex indoor-outdoor climatic interactions, thereby enhancing the reliability of short- to medium-term electricity load predictions.

F. Multivariate Multi-step Forecasting Results

This section presents and analyzes the results obtained from the proposed transformer-based multi-step forecasting models. Table 3 and Figure 8 summarize the performance comparison between models trained with traditional input features (Original) and those integrated with the BAIT index across different forecast horizons (96, 192, 336, and 720 h).

From Table 3, it is evident that incorporating the BAIT index consistently improves forecasting accuracy across all transformer-based models and forecast horizons. Specifically, the iReformer model with BAIT features achieved the best MSE results (bolded), outperforming other models across most horizons, particularly at the shorter prediction length (96 h, MSE = 0.0813). Similarly, the MAE metrics showed substantial improvements with BAIT integration, with iReformer and iInformer models demonstrating superior performance, achieving the lowest MAE values of 0.2172 and 0.2250 respectively at the 96-h horizon.

Figure 8 further visualizes these performance differences. Clearly, all models exhibit improved accuracy with BAIT across both MSE and MAE metrics, particularly noticeable in shorter-term forecasts (96 and 192 h). The iTransformer and iFlashformer models exhibit larger relative improvements upon BAIT integration, especially noticeable at the 336-h forecast length, highlighting the advantage of BAIT in capturing nuanced climatic interactions for intermediate-term predictions.

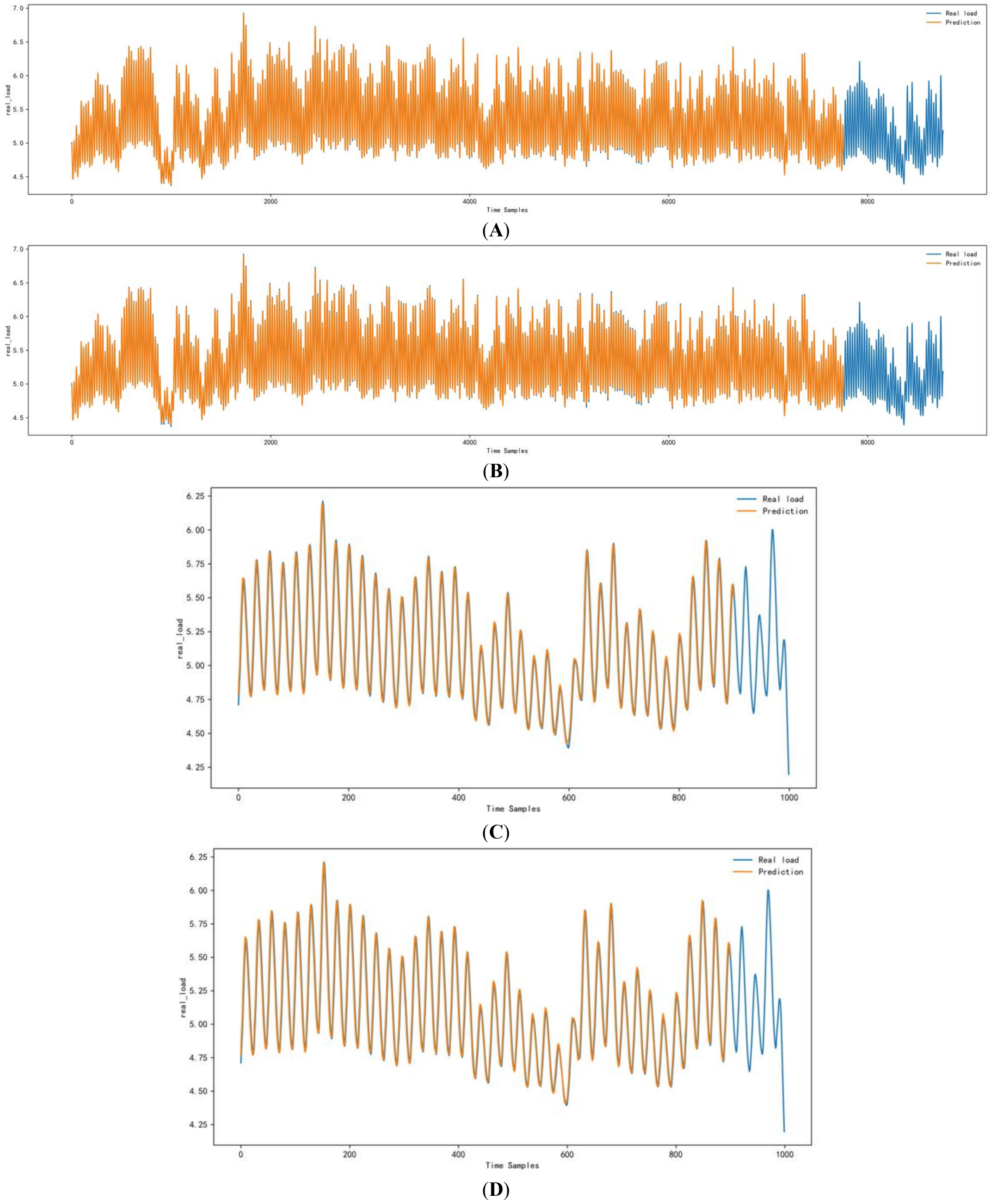

Figure 7. Real load and prediction by LSTM with and without BAIT index for demonstration. (**A**) First year prediction by LSTM without BAIT; (**B**) First year prediction by LSTM with BAIT; (**C**) 1000 time samples prediction by LSTM without BAIT; (**D**) 1000 time samples prediction by LSTM with BAIT.

However, as the prediction horizon extends to 720 h, the improvement margins offered by BAIT tend to diminish, suggesting that long-term load predictions become increasingly challenging, even with improved climatic representations. Nonetheless, the BAIT-enhanced transformer models consistently maintain lower error rates compared to their counterparts using original features.

Notably, several models exhibited superior or comparable predictive accuracy for the 720-h forecast horizon (e.g., iTransformer, iFlashformer for MSE) when contrasted with the 336-h horizon. This apparently counterintuitive phenomenon can be attributed to distinct characteristics of the time series and model capabilities. The 720-h interval, equivalent to 30 days, may align more effectively with dominant and stable long-term cyclical patterns, such as monthly variations, inherent in power load data. The Transformer-based architectures employed are particularly adept at capturing such long-range dependencies. Consequently, these models might identify more robust and predictable signals at this extended horizon, potentially less obscured by the intermediate-term volatility or less defined periodicities that could exert a greater influence on predictions at the 336-h mark.

Overall, the empirical results underscore the effectiveness of integrating the BAIT index into transformer-based forecasting frameworks, confirming its utility in capturing relevant climatic conditions and enhancing predictive performance across multiple forecasting horizons.

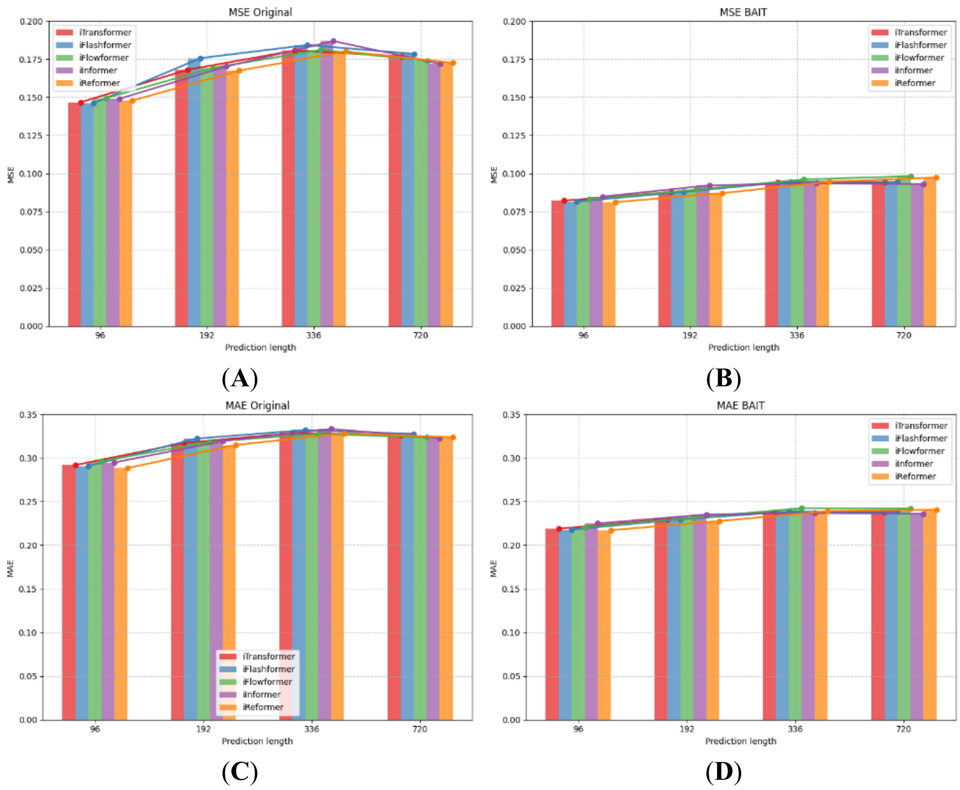

Figure 8. Performance from different models with different features and prediction lengths. (**A**) MAE with climate features; (**B**) MSE with BAIT features; (**C**) MAE with climate features; (**D**) MAE with BAIT features.

Table 3. Comparative study of proposed method and baseline methods on multivariate task.

		iTransformer		iFlashformer		iFlowformer		iInformer		iReformer	
Pred		***Original***	***BAIT***	***Original***	***BAIT***	***Original***	***BAIT***	***Original***	***BAIT***	***Original***	***BAIT***
96	**MSE**	0.1467	0.0822	0.1461	0.0816	0.1495	0.0831	0.1490	0.0849	0.1478	**0.0813**
	MAE	0.2919	0.2190	0.2909	0.2180	0.2964	0.2207	0.2949	0.2250	0.2885	**0.2172**
192	**MSE**	0.1681	0.0879	0.1757	0.0880	0.1694	0.0894	0.1702	0.0923	0.1675	**0.0871**
	MAE	0.3171	0.2297	0.3222	0.2297	0.3186	0.2315	0.3194	0.2352	0.3150	**0.2277**
336	**MSE**	0.1807	0.0943	0.1843	0.0947	0.1811	0.0962	0.1869	**0.0937**	0.1801	0.0947
	MAE	0.3285	0.2379	0.3320	0.2384	0.3276	0.2425	0.3333	**0.2370**	0.3283	0.2392
720	**MSE**	0.1769	0.0943	0.1785	0.0946	0.1739	0.0983	0.1720	**0.0930**	0.1726	0.0974
	MAE	0.3261	0.2382	0.3276	0.2385	0.3236	0.2423	0.3227	**0.2359**	0.3237	0.2408

4. Discussion

The empirical findings presented in this study underscore several critical insights regarding the effectiveness of integrating multi-factor inputs, particularly the Building-Adjusted Internal Temperature (BAIT) index, in electricity load forecasting models.

Firstly, the comparative results demonstrate that employing the BAIT index substantially enhances forecasting accuracy across various models and prediction horizons. The improvements observed in evaluation metrics such as MAE, RMSE, MAPE, and R^2 suggest that the BAIT index effectively encapsulates the complex interactions between climatic variables and building-specific characteristics, significantly outperforming traditional weather features alone. For instance, the Bi-GRU model, which leverages bidirectional contextual information processing, consistently delivered superior performance in single-step forecasting scenarios when integrated with BAIT. The observed reduction in errors, such as a 10.46% decrease in MAPE for the Bi-GRU model, clearly illustrates the benefit of capturing nuanced climatic dynamics with BAIT.

The transformer-based multi-step forecasting results further reinforce the robustness and adaptability of advanced neural network architectures in handling extended forecasting horizons. Among the evaluated transformer models, the iReformer notably exhibited exceptional predictive accuracy across shorter forecast horizons, particularly at 96-h intervals, achieving the lowest MAE and MSE values. This highlights the effectiveness of locality-sensitive hashing attention mechanisms in managing both local and global dependencies within multivariate time series data. However, as forecasting horizons extended beyond two weeks, the marginal benefits offered by integrating BAIT decreased, indicating that long-term load forecasting remains inherently challenging, even when leveraging advanced climate representation techniques.

The visual analyses of BAIT demonstrated the significant impact of outdoor climatic conditions, particularly solar irradiance and wind speed, on indoor temperature dynamics. The clear correlation observed between outdoor temperatures and BAIT values validates the underlying premise of the BAIT index as an effective proxy for perceived indoor comfort levels and subsequent energy demands.

Finally, the performance analysis across various forecast lengths confirmed the practical advantages of transformer-based models combined with BAIT, particularly in mid-range forecasting (up to two weeks). Although predictive accuracy inevitably declines as the forecast horizon lengthens, the consistent superiority of BAIT-enhanced models suggests their suitability for reliable short- to medium-term load management decisions.

In summary, this study provides strong theoretical justification and empirical validation for adopting multi-factor approaches, particularly those incorporating nuanced indoor-outdoor climatic interactions, to improve electricity load forecasting accuracy. The insights derived from this research not only advance forecasting methodologies but also offer practical implications for enhancing the efficiency and sustainability of power system operations.

5. Conclusions and Future Work

This study emphasizes the critical role of integrating multi-factor inputs, particularly the Building-Adjusted Internal Temperature (BAIT) index, in enhancing the accuracy and reliability of electricity load forecasting models. Through extensive comparative analyses, the incorporation of BAIT into various deep learning architectures—including RNN-derived models (LSTM, GRU, Bi-LSTM, Bi-GRU) and advanced transformer-based models (iTransformer, iFlashformer, iFlowformer, iInformer, iReformer)—consistently demonstrated superior predictive performance over traditional climatic factors alone. Empirical results underscored notable improvements across multiple evaluation metrics, affirming BAIT's effectiveness in capturing intricate indoor-outdoor climatic dynamics and addressing the delayed responses of building energy consumption to external weather fluctuations.

Although the proposed BAIT-enhanced framework was validated on a tropical commercial-building dataset, its transfer to other regions is straightforward: the statistical coefficients (x, y, z, σ) can be recalibrated with just a few weeks of local data. In heating-dominated climates, Heating Degree Days (HDD) should be appended, and the sigmoid switching threshold should be shifted upwards to reflect heating activation. For arid regions with large diurnal swings, incorporating absolute humidity (or enthalpy) and shortening the smoothing window improves responsiveness, whereas heavy-mass buildings may require extending the window to 72–96 h. The current study still assumes a single thermal zone without advanced HVAC control; validating these adaptations on other cities' datasets constitutes our immediate future work.

Future research directions include extending the proposed approach to other geographic regions with distinct climatic and building characteristics, exploring additional influencing factors such as socio-economic indicators and occupant behavior, and incorporating real-time adaptive mechanisms for model updating and dynamic forecasting. Further investigation into large language models might also provide enhanced predictive capabilities for long-term forecasting horizons, ultimately contributing to more sustainable, efficient, and resilient urban power systems. Future work will also focus on further enhancing model interpretability. In addition to SHAP-based feature importance analysis, we plan to employ advanced explanation techniques such as attention visualization, attention rollout, and integrated gradients to better elucidate the decision-making processes of both RNN and transformer-based models. We will also investigate counterfactual and example-based explanations to uncover how specific input changes influence predictions. These efforts aim to bridge the gap between model accuracy and transparency, making our forecasting system more trustworthy for practical applications.

Author Contributions

Y.W.: conceptualization, methodology, software, data curation, formal analysis, visualization, investigation, writing—original draft preparation; Q.L.: writing—reviewing and editing; Y.X.: supervision, project administration. All authors have read and agreed to the published version of the manuscript.

Funding

This research is supported by the National Research Foundation, Singapore, and National Environment Agency (NEA) under its Climate Impact Science Research (CISR) Programme (project ID NRF-000413-01). Y.Wang, Q. Li and Y. Xu are with Nanyang Technological University, Singapore.

Data Availability Statement

The electricity load data used in this study were obtained from the Energy Market Authority (EMA) of Singapore and cover the period from 2019 to 2022. These data can be accessed through the official EMA website (https://www.ema.gov.sg), though detailed time-series records may require formal request or institutional affiliation. The climate-related variables used to compute the Building-Adjusted Internal Temperature (BAIT) index, including temperature, solar irradiation, wind speed, and humidity, were obtained from the Renewables.ninja platform (https://www.renewables.ninja/). This platform provides hourly meteorological and solar data for global locations. Access to the data requires user registration and is subject to terms of academic or non-commercial use.

Conflicts of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Hong, T.; Fan, S. Probabilistic electric load forecasting: A tutorial review. *Int. J. Forecast.* **2016**, *32*, 914–938.
2. Pomponi, F.; Moncaster, A. Circular economy for the built environment: A research framework. *J. Clean. Prod.* **2017**, *143*, 710–718. https://doi.org/10.1016/j.jclepro.2016.12.055.
3. Zhao, H.; Magoulès, F. A review on the prediction of building energy consumption. *Renew. Sustain. Energy Rev.* **2012**, *16*, 3586–3592. https://doi.org/10.1016/j.rser.2012.02.049.
4. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780.
5. Kong, W.; Dong, Z.Y.; Jia, Y.; et al. Short-term residential load forecasting based on LSTM recurrent neural network. *IEEE Trans. Smart* Grid **2019**, *10*, 841–851.
6. Cho, K.; van Merriënboer, B.; Bahdanau, D.; et al. On the properties of neural machine translation: Encoder-decoder approaches. *arXiv* **2014**, arXiv:1409.1259.
7. Zheng, Z.; Zhu, J.; Lu, Y. Load forecasting using a GRU-based RNN model. In Proceedings of the International Conference on Energy, Electrical and Power Engineering, Munich, Germany, 17–18 July 2017.
8. Li, S.; Jin, X.; Xuan, Y.; et al. Enhancing Transformer Efficiency with Channel-wise Tokenization for Time Series Forecasting. In *Advances in Neural Information Processing Systems*; NeurIPS: La Jolla, CA, USA, 2023; Volume 36.
9. Dao, T.; Fu, D.; Ermon, S.; et al. FlashAttention: Fast and Memory-Efficient Exact Attention with IO-Awareness. In *Advances in Neural Information Processing Systems* NeurIPS: La Jolla, CA, USA, 2022; Volume 35.
10. Liu, Y.; Gu, J.; Zeng, X.; et al. Flowformer: A Continuous-Time Transformer for Time Series Forecasting. In Proceedings of the 39th International Conference on Machine Learning, PMLR, Baltimore, Maryland, USA, 17–23 July 2022; Volume 162, pp. 14329–14343.
11. Zhou, H.; Zhang, S.; Peng, J.; et al. Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting. In Proceedings of the AAAI Conference on Artificial Intelligence, Virtually, 2–9 February 2021; Volume 35, pp. 11106–11115.
12. Kitaev, N.; Kaiser, Ł.; Levskaya, A. Reformer: The Efficient Transformer. In Proceedings of the International Conference on Learning Representations, Addis Ababa, Ethiopia, 26–30 April 2020.
13. Chen, Y.; Zhang, D. Theory-guided deep-learning for electrical load forecasting (TgDLF) via ensemble long short-term memory, *Adv. Appl. Energy* **2021**, *1*, 100004. https://doi.org/10.1016/j.adapen.2020.100004.
14. Staffell, I.; Pfenninger, S.; Johnson, N. A global model of hourly space heating and cooling demand at multiple spatial scales. *Nat. Energy* **2023**, *8*, 1328–1344. https://doi.org/10.1038/s41560-023-01341-5.
15. Yang, W.; Sparrow, S.N.; Wallom, D.C. A comparative climate-resilient energy design: Wildfire Resilient Load Forecasting Model using multi-factor deep learning methods. *Appl. Energy* **2024**, *368*, 123365. https://doi.org/10.1016/j.apenergy.2024.123365.
16. Schuster, M.; Paliwal, K.K. Bidirectional recurrent neural networks. *IEEE Trans. Signal Process.* **1997**, *45*, 2673–2681.