

Comparative Evaluation of LSTM and GRU for Long-Term Temperature Forecasting: A Case Study

Dipta Divakara Pius Purwadaria, Ting Lie*, Fong-Ching Yuan

Department of Information Management, Yuan Ze University, Taoyuan

Email: *tinglie@saturn.yzu.edu.tw

How to cite this paper: Purwadaria, D.D.P., Lie, T. and Yuan, F.-C. (2026) Comparative Evaluation of LSTM and GRU for Long-Term Temperature Forecasting: A Case Study. *Applied Mathematics*, 17, 249-264.

<https://doi.org/10.4236/am.2026.175015>

Received: March 17, 2026

Accepted: May 23, 2026

Published: May 26, 2026

Copyright © 2026 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Accurate long-term temperature forecasting is essential for climate-sensitive planning in subtropical regions characterized by pronounced seasonal variability. This study compares two recurrent neural network architectures Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) for long-term monthly temperature prediction in Taiwan region. Meteorological data from 2010 - 2025 were used, with correlation-based feature selection and Min-Max normalization applied prior to training. The dataset was chronologically partitioned into training, validation, and testing subsets to ensure temporal integrity. Model performance was evaluated using Mean Absolute Percentage Error (MAPE), and convergence behavior was examined to assess computational efficiency. Both models successfully captured nonlinear seasonal dynamics and achieved high forecasting accuracy. However, GRU converged in approximately half the number of training epochs required by LSTM, demonstrating superior computational efficiency. These results suggest that greater architectural complexity does not necessarily yield materially improved long-term forecasting performance. The findings provide empirical evidence to inform recurrent neural network selection for extended-horizon monthly climate prediction tasks.

Keywords

Forecasting, Temperature, LSTM, GRU

1. Introduction

Accurate temperature forecasting plays a critical role in modern energy system management. In electricity-intensive regions, particularly during extended heat periods, even minor forecasting errors can substantially alter cooling demand,

peak load intensity, and overall grid stability [1]. As climate variability intensifies and extreme temperature events become more frequent, reliable long-term temperature prediction is no longer solely a meteorological challenge. It has become a strategic component of sustainable energy planning.

Taiwan region provides a compelling context for examining this issue. As a highly industrialized and densely populated economy, Taiwan region's electricity consumption is strongly influenced by temperature fluctuations, especially during summer months when air-conditioning demand surges [2]. The case of Kaohsiung, Taiwan region's largest industrial and port city, further illustrates the stakes involved. According to the Taiwan Power Company, Kaohsiung consistently records the highest electricity consumption in Taiwan region due to its heavy industrial concentration, dense urbanization, and prolonged high-temperature conditions [3]. In such environments, inaccurate temperature forecasts may lead to supply-demand mismatches, increasing the risk of grid stress and power shortages [4].

Historical disruptions continue to demonstrate the severe economic consequences of energy instability, with modern outages scaling in cost alongside increased technological dependency. The March 2022 nationwide power outage in Taiwan region, triggered by a localized failure at the Hsinta Power Plant, impacted over 5.49 million households and caused industrial losses reaching tens of billions of New Taiwan dollars, a modern echo of the systemic vulnerabilities seen in 1999 [5]. Similarly, the United States has faced unprecedented instability; the February 2021 Texas power crisis (Winter Storm Uri) resulted in economic damages estimated between \$10 billion and \$20 billion in direct losses, with total societal costs nearing \$200 billion [6]. These events highlight a crucial insight: energy resilience depends not only on generation capacity but also on accurate demand anticipation, of which temperature forecasting increasingly volatile due to climate change is a foundational element [5] [7].

In recent years, deep learning approaches have emerged as powerful tools for time-series forecasting. Recurrent neural network architectures such as Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) are particularly effective in modeling non-linear dynamics, long-term dependencies, and seasonal variations inherent in climate data [8]. Compared with traditional statistical models, these architectures offer greater flexibility in capturing complex temporal relationships without strict stationarity assumptions [9].

Despite their increasing application, comprehensive comparisons of LSTM and GRU for long-term temperature forecasting in Taiwan region remain limited. Many existing studies focus on short-term prediction horizons or purely meteorological performance, without explicitly linking forecasting accuracy to energy system implications. Furthermore, extended multi-year evaluations that assess model robustness across varying climatic conditions are relatively scarce. To address this gap, this study develops and compares LSTM and GRU models for long-term monthly temperature forecasting in Taiwan region, using official meteorological data obtained from the Climate Observation Data Inquire Service (CODiS).

The dataset spans from 2010 - 2025, enabling a comprehensive evaluation of model performance across more than 15 years of continuous climate variability. Importantly, this research situates forecasting accuracy within the broader framework of electricity demand management and energy system stability. Accordingly, this study seeks to answer the following research questions:

1. What is the predictive performance of Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) models for long-term monthly temperature forecasting in Taiwan region during the period 2010 - 2025?
2. How do LSTM and GRU compare in terms of forecasting accuracy, convergence behavior and computational efficiency during model training?

By addressing these questions, this study bridges methodological advances in deep learning with practical energy system planning considerations, offering both theoretical and applied contributions to climate-energy research.

2. Literature Review

2.1. Temperature Forecasting and Energy Consumption

Temperature forecasting plays a critical role in the operation and management of modern energy systems. As one of the most influential meteorological parameters, air temperature directly affects electricity demand patterns, particularly during summer peak periods when cooling requirements dominate total energy use. Variations in temperature can lead to significant fluctuations in electricity consumption, making accurate forecasting essential for optimizing generation schedules, managing reserve margins, and preventing power shortages [10].

The relationship between air temperature and electricity demand is a documented pillar of energy management, with recent data confirming a significant correlation between extreme heat and record-breaking consumption. In Taiwan region, contemporary patterns from 2023 to 2025 show that daily electricity consumption from June to September remains consistently higher than during the rest of the year due to elevated ambient temperatures, with residential and commercial cooling placing substantial pressure on the grid [11]. Consequently, the Percent Operating Reserves (POR) of the grid often decreases as temperatures rise, frequently falling into single digits and triggering supply alerts [5]. These findings emphasize that accurate temperature forecasting is a cornerstone of national energy security, especially as climate extremes increase the frequency of large-scale outages [6].

The case of Kaohsiung, Taiwan region's primary industrial hub, further illustrates these stakes. Kaohsiung consistently records high electricity demand driven by heavy industry, dense urbanization, and prolonged high-temperature conditions [12]. Power shortages in such high-demand areas can trigger severe economic losses; for example, the March 2022 nationwide blackout, caused by a failure at the Hsinta Power Plant in Kaohsiung, impacted 5.49 million households and resulted in substantial economic disruption, a modern parallel to the devastating 1999 outage [5]. Comparable events have been observed globally; in the United States, the February 2021 Texas grid collapse (Winter Storm Uri) resulted in esti-

mated economic damages upwards of \$130 billion [6].

Global warming brings the importance of temperature effect on energy demand more severely. In 2025, the U.S. experienced 23 separate billion-dollar weather disasters, many involving severe storms and “heat-driven droughts” that threatened regional grid stability across the central and western states. The increasing frequency of these events can be attributed in part to global warming and climate change, which have intensified temperature extremes and made energy systems more sensitive to weather fluctuations. Notably, around 25% of this increase is expected to result directly from climate change, while the remainder stems from socioeconomic development and rising living standards. This growing demand for cooling energy underscores the importance of reliable temperature forecasting in sustainable power system management [13].

Given this context, temperature forecasting during peak electricity demand periods becomes particularly crucial. Reliable temperature forecasts can help power utilities anticipate load fluctuations, plan generation capacity, and prevent blackouts. They also support long-term energy planning and climate adaptation strategies, ensuring that energy infrastructure remains resilient under changing climatic conditions. However, accurately forecasting temperature, especially over long-term horizons, remains challenging due to the nonlinear and dynamic nature of atmospheric processes. These challenges have motivated researchers to explore advanced computational methods, including machine learning and deep learning models, to improve forecast precision and robustness [14].

2.2. Deep Learning Approaches for Temperature

Forecasting tasks involving sequential or time-dependent data, such as meteorological and energy demand prediction, are highly complex due to the intricate temporal dependencies inherent in the data. Traditional models, including basic Artificial Neural Networks (ANN) and early Recurrent Neural Networks (RNN), frequently encounter challenges when processing long sequences, primarily due to the “vanishing” and “exploding” gradient problems during the training phase. To overcome these limitations, advanced gated architectures such as LSTM and GRU have been developed to effectively capture long-range dependencies.

2.2.1. Long Short-Term Memory (LSTM)

The LSTM network, introduced in [15], is a specialized RNN architecture designed to maintain an internal memory state over extended periods. Unlike standard feedforward networks where inputs are independent, LSTM utilizes a cell state (memory) and a series of “gates” to regulate the flow of information.

- **Forget Gate:** This initial stage determines which information from the previous cell state should be discarded. It utilizes a sigmoid function to evaluate the previous hidden state and the current input, outputting a value between 0 (completely discard) and 1 (completely retain).
- **Input Gate:** This gate identifies which new information will be stored in the cell state. A sigmoid layer decides which values to update, while a tanh layer creates

a vector of new candidate values; these are multiplied to update the memory.

- **Output Gate:** Finally, the output is determined based on the filtered version of the cell state. A sigmoid function decides which parts of the state to output, which is then multiplied by the cell state passed through a tanh function to ensure the final output remains within a standardized range.

While LSTMs are highly effective for complex time-series problems, their sophisticated architecture—consisting of three distinct gates—can lead to increased computational demands and longer training times.

2.2.2. Gated Recurrent Unit (GRU)

As a response to the computational complexity of LSTMs, [16] introduced the Gated Recurrent Unit (GRU), a streamlined variant that offers comparable performance with a simplified structure. The GRU optimizes efficiency by merging the cell state and hidden state and reducing the number of gates to two:

- **Update Gate:** This component combines the functions of the LSTM's forget and input gates, determining how much of the previous memory should be retained and how much new information should be added.
- **Reset Gate:** This gate controls how much of the past information is used to calculate the new candidate state, effectively deciding how to combine the new input with the previous memory.

Because the GRU features fewer parameters, it often achieves faster convergence during training compared to the LSTM, making it an increasingly popular choice for resource-efficient climate modeling. Recent studies have indicated that for multi-year monthly forecasting, the GRU can provide predictive accuracy like that of the more complex LSTM while significantly reducing the number of training epochs required.

2.3. Correlation Analysis

In temperature and energy forecasting studies, input variable selection plays a crucial role in determining model accuracy and computational efficiency. Including irrelevant or redundant variables can degrade model performance, while carefully chosen predictors enhance generalization and reduce overfitting. One of the most widely used statistical techniques for identifying relevant input features is correlation analysis, which measures the strength and direction of a linear relationship between variables.

Among various correlation measures, the Pearson Product-Moment Correlation (PPMC) coefficient is commonly employed to quantify the relationship between two continuous variables. The Pearson correlation coefficient ρ is mathematically defined as:

$$\rho = \frac{\sum_{i=1}^n (x_i - \mu_x)(y_i - \mu_y)}{\sqrt{\sum_{i=1}^n (x_i - \mu_x)^2 \sum_{i=1}^n (y_i - \mu_y)^2}} \quad -1 \leq \rho \leq 1 \quad (1)$$

where n is the number of observations; x_i is the value of x (for i -th observation);

y_i is the value of y (for i -th observation); μ_x is the mean of x variable; μ_y is the mean of y variable. The r values range between -1.0 and -0.7 shows a strong negative correlation, while a correlation between 0.7 and 1.0 indicates a strong positive correlation. A correlation of 0.0 indicates no relationship at all.

According to statistical conventions follows [17]:

- $|r| \geq 0.7$: Strong correlation
- $0.5 \leq |r| < 0.7$: Moderate correlation
- $|r| < 0.5$: Weak correlation

By applying correlation analysis before model training, redundant and weakly correlated predictors are eliminated. This preprocessing not only reduces computational complexity but also improves model convergence and accuracy. The approach aligns with recommendations from [18] and [19], who emphasized that feature selection based on correlation can significantly improve the performance of forecasting models in energy and environmental applications.

2.4. Evaluating the Performance Index

Several measurement indicators have been proposed and employed to assess the prediction accuracy of models. The most frequently used measure is the mean absolute percentage error (MAPE) [18]. A significant advantage of this measure is that it does not depend on the magnitudes of the forecast variables. [19] suggested that MAPE is the most appropriate error measure for evaluating forecasting performance. In this study, the mean absolute percentage error (MAPE) is used to evaluate the forecasting accuracy through cross-sectional analysis.

$$MAPE = \frac{\sum_{t=1}^n |(A_t - F_t)/A_t|}{n} \quad (2)$$

where A_t is the actual values for period t , F_t the expected value for period t , and n is the number of training samples. The smaller the $MAPE$, the closer to the actual historical data, the predicted values are, and the more accurate is the forecasting model.

3. Methodology

3.1. Flow Chart of LSTM and GRU Models

Figure 1 shows the flowchart of LSTM and GRU models which have been used in this investigation. The data set is partitioned to three parts—training data set, validation data set, and testing data set. The training data set was used to determine the optimal forecasting model, the validation data set was used to prevent overfitting of the various models, and the testing data set was used to investigate the performance of the various models [20]. The optimization scheme of two models is summarized as follows: Step 1 is to use correlation analysis to obtain the best features. Step 2 is to initialize the epoch of the models. Step 3 is to evaluate the fitness of each epoch using MAPE given in Equation (2) for each training and validation data set. Step 4 is to repeat the process, as shown in **Figure 1** until the epoch's number is equal to 1000 with MAPE convergence. When the termination

condition is satisfied, the well-trained models and testing data set will be used with the best epoch to investigate the performance of the models.

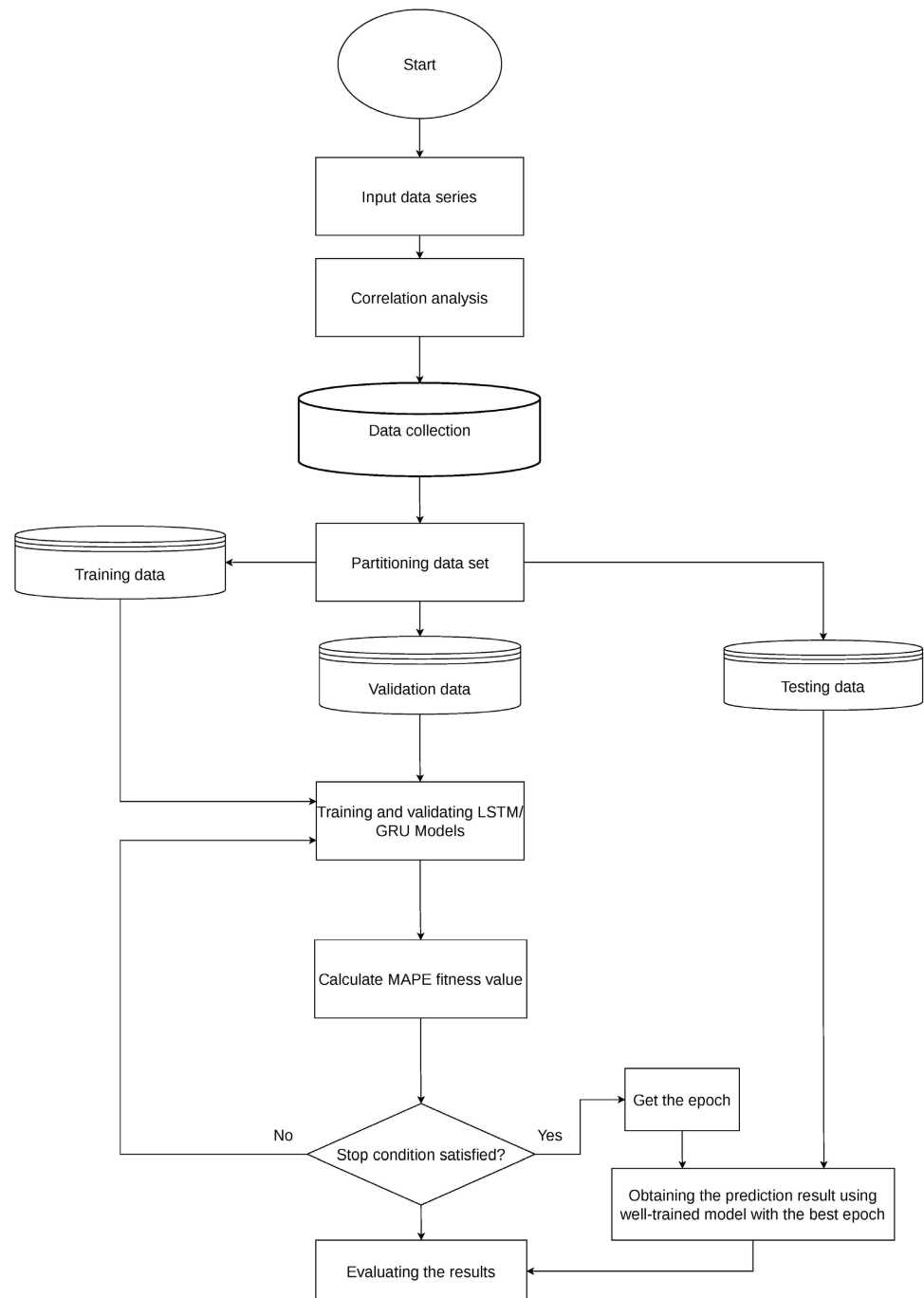


Figure 1. Flowchart of LSTM and GRU models.

3.2. Study Area and Data Description

This study focuses on Kaohsiung, Taiwan region's largest industrial and port city, which consistently records the highest electricity consumption nationwide according to the Taiwan Power Company. Due to its dense industrial concentration

and prolonged summer heat exposure, Kaohsiung represents a high temperature-elastic electricity demand environment, making it an ideal case for evaluating forecasting performance under peak-load-sensitive conditions.

The meteorological data used in this study were obtained from the Central Weather Administration (CWA) of Taiwan region, specifically focusing on the Kaohsiung station records from 2010 to 2025. To ensure the reliability of the long-term series, a rigorous completeness check was performed on the hourly records. Although the station identifier transitioned from 467440 to 467441 during the study period because of the updating the equipment from the old to the new, the geographic coordinates and sensor calibration remained consistent, allowing for a seamless integration of the two records. Missing hourly observations accounts for less than 0.5% of the total dataset. Because the study focuses on monthly temperature aggregated from the 10:00 - 17:00 window, which are the working hours for the factories in Taiwan region using the electricity, any days with more than two missing hours within this specific timeframe were excluded from the monthly mean calculation to maintain high data fidelity. This ensures that the final dataset accurately reflects the peak thermal characteristics of the region without being skewed by instrumental downtime or station relocation.

3.3. Data Preprocessing

To establish a rigorous data foundation for the forecasting models, the meteorological parameters initially identified in **Table 1** underwent a systematic screening process using Pearson correlation analysis. This stage is critical for enhancing model efficiency and predictive accuracy by isolating the most influential drivers of thermal dynamics in Kaohsiung.

To ensure the model adheres to real-world forecasting constraints and to avoid target leakage, all input features were defined as lagged variables. Specifically, for a target temperature at month t , the model strictly utilizes meteorological predictors from the preceding months ($t-1$ through $t-12$). As detailed in **Table 2**, variables demonstrating a strong linear relationship with monthly temperature defined by a Pearson correlation coefficient $|r| \geq 0.7$ were prioritized for model integration. These selected predictors, which include station pressure, sea-level pressure, dew-point temperature, and body temperature, serve as the foundational lagged inputs for the LSTM and GRU architectures.

To reduce redundancy and improve model efficiency, Pearson correlation analysis was conducted between monthly temperature and candidate meteorological variables. Variables with absolute correlation coefficients $|r| \geq 0.7$ were selected as input predictors.

As detailed in **Table 2**, the relationships between the potential predictors and the target monthly temperature were quantified, with the strongest correlations highlighted in bold type. These bolded entries indicate variables that meet the selection threshold of a Pearson correlation coefficient $|r| \geq 0.7$, signaling a high degree of linear association.

Table 1. Frequently used variables in temperature predictions models.

Variables	Frequency	Unit	Explanation
Station test pressure	Monthly (M)	hPa	Atmospheric pressure is directly observed by the observatory.
Sea-level pressure	Monthly (M)	hPa	Height affects the air pressure, converting the observed air pressure to that at sea level.
Temperature	Monthly (M)	°C	Represents the current cold and hot values.
Dew-point temperature	Monthly (M)	°C	The temperature required to which the gaseous water contained in the air at the fixed pressure is saturated and condensed into liquid water.
Relative humidity	Monthly (M)	%	The percentage of the partial pressure and the pressure of saturated vapor at the same temperature and volume. That is the ratio between absolute humidity to maximum humidity, whose value shows how saturated water vapor is.
Wind speed	Monthly (M)	m/s	Speed of the wind.
Wind direction	Monthly (M)	%	The direction the wind blows.
Peak gust	Monthly (M)	m/s	A gust refers to an instantly changing wind, with the maximum gust being the maximum gust during that period.
Max. gust wind direction	Monthly (M)	360°	The wind direction comes from the maximum wind gust.
Precipitation	Monthly (M)	mm	The amount of accumulated rainwater.
Hours of precipitation	Monthly (M)	hr	The duration when the rain lasts.
Hours of sunshine	Monthly (M)	hr	The duration of the sun.
Full sky radiation	Monthly (M)	MJ/nf	Used to measure solar radiation energy at a given time and region.
Visibility	Monthly (M)	km	The observer can still clearly see the object at how far away from the object.
Ultra-violet Index	Monthly (M)	UVI	Solar UV radiation intensity.
Total amount of cloud	Monthly (M)	0 - 10	The proportion of the sky shaded by the clouds, 0 cloudless, 10 clouds for all day.
Body temperature	Monthly (M)	°C	The human body's feeling of temperature.

Implementing this threshold ensures the retention of only the most statistically significant predictors, effectively reducing dimensionality and streamlining the model architecture. By eliminating redundant or weakly correlated variables, this approach improves convergence stability during the estimation process and mitigates the risk of overfitting, thereby enhancing the model's predictive generalizability.

Table 2. Correlation analysis.

Variables	Pearson correlation coefficient	Variables	Pearson correlation coefficient
Station test pressure	0.8764	Precipitation	0.4953
Sea-level pressure	0.8953	Hours of precipitation	0.4504
Temperature	0.9214	Hours of sunshine	0.4075
Dew-point temperature	0.9727	Full sky radiation	0.6596
Relative humidity	0.4423	Visibility	0.6853
Wind speed	0.1542	Ultra-violet index	0.6946
Wind direction	0.7184	Total amount of cloud	0.2674
Peak gust	0.3901	Body temperature	0.9171
Max. gust wind direction	0.1151		

By filtering out weakly correlated variables and strictly utilizing historical lags, the study reduces dimensionality and ensures that the forecasting results are not artificially inflated by contemporaneous data leakage. Finally, all selected variables were normalized using Min-Max scaling prior to model training to ensure numerical stability and accelerate gradient-based optimization:

$$X_{\text{scaled}} = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \quad (3)$$

3.4. Data Partitioning

To preserve temporal integrity and prevent data leakage, the dataset was chronologically divided into three non-overlapping subsets. The training set, spanning the years 2010 through 2022, was utilized to estimate the models' internal parameters and weights across thirteen years of continuous meteorological observations. Subsequent data from 2023 formed the validation set, which was used exclusively to tune hyperparameters and monitor convergence behavior to prevent overfitting during the training process. Finally, the period of 2024 - 2025 was designated as the testing set, reserved for out-of-sample performance evaluation. In this phase, the models utilized the 2024 data to generate the 2025 temperature forecasts, ensuring that the reported accuracy metrics reflect the models' ability to predict a truly independent future horizon.

3.5. Model Training Strategy and Parameter

To transition from raw time-series observations to a supervised learning format, a sliding window technique was implemented. This study utilized a look-back window of 12 months, meaning the meteorological features from the previous year were used as input to predict the temperature for the following month, establishing a forecast horizon of $t + 1$. Given the total dataset span from January 2010 to December 2025, this 12-month sliding window generated a total of 180 overlapping sequences. This sequence count provides a sufficient dataset for deep learning, allowing the LSTM and GRU architectures to effectively learn the recurring seasonal dependencies and long-term thermal dynamics inherent in Taiwan region's climate. Both models were trained using Backpropagation Through Time (BPTT). The objective of optimization was to minimize Mean Absolute Percentage Error (MAPE). Training continued until validation MAPE stabilized, preventing overfitting. The parameter settings of LSTM and GRU are shown in **Table 3**.

Table 3. Parameter setting of the optimization searching algorithms.

Algorithm	Parameter	Algorithm value
LSTM	Number of training and validation epochs	1000
	Number of testing epochs	Based on loss
	Batch size	32
	Optimizer	Adam

Continued

LSTM	First LSTM layer unit	50
	Second LSTM layer unit	50
	Third LSTM layer unit	50
Number of training and validation epochs		1000
Number of testing epochs		Based on loss
GRU	Batch size	32
	Optimizer	Adam
	First GRU layer unit	50
	Second GRU layer unit	50

To ensure a rigorous and controlled comparison, both the LSTM and GRU architectures were evaluated using an identical hyperparameter search space. This included a shared configuration for the Adam optimizer, a consistent batch size of 32, and a standardized unit count of 50 per layer. While the LSTM model was configured with three recurrent layers, the GRU architecture utilized two. This structural variation was determined through preliminary hyperparameter tuning on the validation set; it was observed that a third GRU layer introduced unnecessary architectural complexity that led to immediate overfitting and increased computational latency without improving predictive accuracy. Therefore, the chosen configurations represent the optimal depth for each specific model type under the same experimental constraints, ensuring that the performance comparison reflects the best-tuned version of each architecture.

Based on the empirical observations during the training phase, the two recurrent neural network architectures exhibited distinct convergence behaviors. The Long Short-Term Memory (LSTM) network reached convergence at approximately 800 epochs, whereas the Gated Recurrent Unit (GRU) network converged more rapidly, requiring only about 400 epochs.

To ensure optimal forecasting performance and mitigate the risk of overfitting, the final weights for both models were not necessarily taken from the final epoch. Instead, the models were strictly selected based on the epoch that yielded the minimum Mean Absolute Percentage Error (MAPE) on the validation set.

4. Result and Discussion

This section presents the empirical findings derived from the implementation of the LSTM and GRU models for long-term monthly temperature forecasting in Kaohsiung, Taiwan region. The results are evaluated based on graphical comparison, numerical prediction accuracy, and error-based performance metrics. Emphasis is placed on assessing the models' ability to capture seasonal temperature dynamics, maintain stability across transitional climatic periods, and deliver reliable forecasts during peak summer months when electricity demand sensitivity is the highest. In addition to predictive accuracy, convergence behavior and compu-

tational efficiency are considered to provide a comprehensive evaluation of model suitability for temperature-driven energy demand management applications.

The following analysis systematically interprets the forecasting outcomes and discusses their implications within the broader context of climate variability and energy system resilience.

Figure 2 presents the graphical comparison between observed monthly temperatures in Kaohsiung and the corresponding predictions generated by the LSTM and GRU models. The figure demonstrates that both architectures effectively reproduce the pronounced seasonal cycle typical of subtropical climates, including lower temperatures in winter (January-February), a steady increase during spring, a sustained peak during summer (June-September), and a gradual decline toward late autumn and early winter. The predicted curves from both models closely track the observed trajectory without structural divergence, indicating strong generalization capability beyond the training period. Minor deviations are observable during transitional months, where both models slightly overestimate temperatures in early-year months and show marginal fluctuations during the peak summer period.

However, these discrepancies remain limited in magnitude and do not disrupt the overall seasonal structure. The graphical alignment suggests that both recurrent architectures successfully capture long-term nonlinear dependencies embedded in the monthly temperature series.

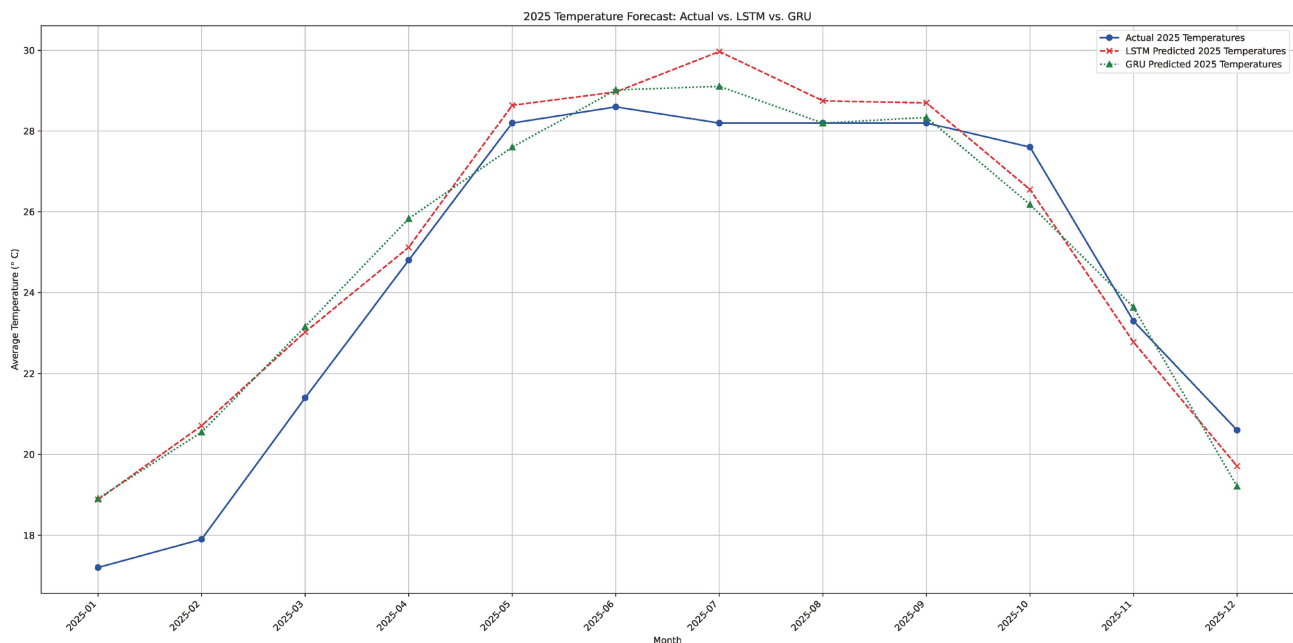


Figure 2. Comparison of actual value and prediction results using LSTM and GRU.

Table 4 provides a detailed numerical comparison of actual and predicted values across all twelve months. The results indicate that both models maintain forecasting errors within a narrow range, generally below $\pm 1^\circ\text{C}$. During winter months, both LSTM and GRU tend to slightly overestimate temperature, which may reflect

reduced variance and greater atmospheric instability during colder periods. In spring and early summer, predictions closely approximate observed values, demonstrating stable learning of transitional seasonal dynamics. Notably, during peak summer months when electricity demand sensitivity is the highest in both models exhibit strong predictive consistency, with only minor overestimation in July and marginal deviations in late summer. This stability during high-temperature periods is particularly significant from an energy management perspective, as forecasting reliability during peak demand seasons directly influences load allocation and reserve planning decisions. Toward the end of the year, both models show slight underestimation, yet the magnitude remains operationally acceptable. Overall, the monthly breakdown confirms that forecasting performance is consistent across seasons and does not exhibit systematic bias.

Table 4. Forecasting model results

Model	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Actual value	17.2	17.9	21.4	24.8	28.2	28.6	28.2	28.2	28.2	27.6	23.3	20.6
LSTM	18.9	20.7	23.0	25.1	28.6	29.0	30.0	28.8	28.7	26.6	22.8	19.7
GRU	18.9	20.6	23.2	25.8	27.6	29.0	29.1	28.2	28.3	26.2	23.6	19.2

To further evaluate and compare the predictive performance of the two recurrent architectures, Mean Absolute Percentage Error (MAPE) was employed as the primary accuracy metric. MAPE provides a scale-independent assessment of forecasting precision and allows for direct comparison between models. The results indicate that both LSTM and GRU achieve MAPE values below the commonly accepted 5% threshold for high-accuracy forecasting, demonstrating strong reliability in long-term monthly temperature prediction. The LSTM model records a slightly lower MAPE than GRU; however, the difference between the two values is minimal and does not indicate a practically significant performance gap.

Table 5 summarizes the comparative predictive accuracy of the recurrent architecture using Mean Absolute Percentage Error (MAPE), Mean Absolute Error (MAE), and Root Mean Square Error (RMSE). The results indicate that both LSTM and GRU achieve MAPE values below the commonly accepted 0.05 threshold for high-accuracy forecasting, demonstrating strong reliability in long-term monthly temperature prediction. Specifically, the LSTM model records a MAPE of 0.0481, while the GRU model achieves a nearly identical value of 0.0483. In terms of absolute error, both models maintain an MAE of approximately 1.03°C and an RMSE of 1.28°C, confirming that the average prediction deviation remains within a narrow and operationally acceptable range for climate-sensitive applications.

Table 5. Performance comparison of the different forecasting models.

Model	MAPE	MAE (°C)	RMSE (°C)
LSTM	0.0481	1.0460	1.2826
GRU	0.0483	1.0334	1.2775

Although the LSTM yields a marginally lower MAPE, the absolute difference between the two models is negligible and does not indicate a practically significant performance advantage in the context of temperature-driven energy planning. However, when considering training dynamics where the GRU reached convergence in approximately 400 epochs compared to the 800 epochs required by the LSTM the results suggest that the GRU achieves comparable predictive precision with substantially greater computational efficiency. These findings highlight that increased architectural complexity does not necessarily translate into materially superior long-term forecasting performance for monthly climate series. Consequently, the GRU offers a more balanced and resource-efficient solution for practical implementation in temperature-elastic electricity markets.

5. Discussion

This study developed and systematically compared Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) models for long-term monthly temperature forecasting in Kaohsiung, Taiwan region, using more than 15 years of continuous meteorological observations. By integrating correlation-based feature selection, rigorous data partitioning, and validation-based model selection, the research aimed to evaluate not only predictive accuracy but also computational efficiency and practical relevance within a temperature-driven electricity demand context.

The empirical results demonstrate that both LSTM and GRU effectively capture the nonlinear seasonal dynamics inherent in subtropical climate time series. Both models achieved Mean Absolute Percentage Error (MAPE) values below 0.05, indicating high forecasting reliability for long-term monthly prediction tasks. For comparison purpose, preliminary testing with a Seasonal Naïve model was run as being a benchmark model for this study, yielded a MAPE of 0.0551, which exceeded the commonly accepted 0.05 threshold for high-accuracy forecasting. In conclusion, although LSTM produced a marginally lower MAPE than GRU, the difference was negligible in practical terms. Notably, GRU converged in approximately half the number of epochs required by LSTM, highlighting its superior computational efficiency without sacrificing predictive precision.

These findings suggest that increased architectural complexity does not necessarily guarantee materially improved performance in long-horizon monthly climate forecasting. For applications where computational cost, scalability, and real-time adaptability are critical considerations, GRU may offer a balanced and resource-efficient alternative to LSTM. From a methodological standpoint, this study contributes to the ongoing debate regarding recurrent neural network selection for environmental time series analysis, particularly under extended multi-year forecasting horizons.

Beyond methodological implications, the study situates forecasting accuracy within the broader framework of energy system resilience. In temperature-elastic electricity markets such as Kaohsiung, maintaining forecasting errors within a

narrow margin is essential for effective load scheduling, reserve allocation, and grid stability planning. Accurate monthly temperature prediction enhances the capacity of power utilities to anticipate demand fluctuations and mitigate supply-demand mismatches.

Nevertheless, several limitations should be acknowledged. The study focuses on a single metropolitan case, and future research could extend the comparative framework to multiple climatic zones to evaluate spatial generalizability. Furthermore, while this study prioritized high-accuracy gated architectures, future investigations should explicitly compare these results against a wider array of traditional statistical benchmarks and hybrid models to further quantify the performance gains of deep learning in varied environments. Additionally, incorporating attention mechanisms or exogenous socioeconomic variables may further enhance predictive robustness.

In conclusion, this research demonstrates that both LSTM and GRU are highly effective for long-term monthly temperature forecasting, with GRU offering comparable accuracy and superior computational efficiency. By bridging deep learning methodology with practical energy demand considerations, this study provides empirical evidence supporting the strategic role of advanced forecasting models in strengthening climate-adaptive energy system management.

Funding

This work was supported by the Ministry of Science and Technology (MOST) of the Taiwan region, under Project MOST 110-2410-H-155-003.

References

- [1] IPCC (2023) Climate Change 2023: Synthesis Report. Intergovernmental Panel on Climate Change.
- [2] Huang, Y. (2020) Examining Impact Factors of Residential Electricity Consumption in Taiwan Region Using Index Decomposition Analysis Based on End-Use Level Data. *Energy*, **213**, Article ID: 119067. <https://doi.org/10.1016/j.energy.2020.119067>
- [3] Liu, D. (2022) Residential and Industrial Electricity Consumption in Taiwan Region: Weather or Macroeconomic Condition (or Both). *Energy Strategy Reviews*, **39**, Article ID: 100795. <https://doi.org/10.1016/j.esr.2021.100795>
- [4] Duan, Z., de Wilde, P., Attia, S. and Zuo, J. (2025) Challenges in Predicting the Impact of Climate Change on Thermal Building Performance through Simulation: A Systematic Review. *Applied Energy*, **382**, Article ID: 125331. <https://doi.org/10.1016/j.apenergy.2025.125331>
- [5] Zahraoui, Y., Korótko, T., Rosin, A., Mekhilef, S., Seyedmahmoudian, M., Stojcevski, A., *et al.* (2024) AI Applications to Enhance Resilience in Power Systems and Microgrids—A Review. *Sustainability*, **16**, Article 4959. <https://doi.org/10.3390/su16124959>
- [6] Pan, W. and Li, Y. (2022) Improving Power Grid Resilience under Extreme Weather Conditions with Proper Regulation and Management of Ders—Experiences Learned from the 2021 Texas Power Crisis. *Frontiers in Energy Research*, **10**, Article 921335. <https://doi.org/10.3389/fenrg.2022.921335>

- [7] Melaku, N.D., Fares, A. and Awal, R. (2023) Exploring the Impact of Winter Storm Uri on Power Outage, Air Quality, and Water Systems in Texas, Usa. *Sustainability*, **15**, Article 4173. <https://doi.org/10.3390/su15054173>
- [8] Hewamalage, H., Bergmeir, C. and Bandara, K. (2021) Recurrent Neural Networks for Time Series Forecasting: Current Status and Future Directions. *International Journal of Forecasting*, **37**, 388-427. <https://doi.org/10.1016/j.ijforecast.2020.06.008>
- [9] Yunita, A., Pratama, M.I., Almuzakki, M.Z., Ramadhan, H., Akhir, E.A.P., Firdausiah Mansur, A.B., *et al.* (2025) Performance Analysis of Neural Network Architectures for Time Series Forecasting: A Comparative Study of RNN, LSTM, GRU, and Hybrid Models. *MethodsX*, **15**, Article ID: 103462. <https://doi.org/10.1016/j.mex.2025.103462>
- [10] Madin, V., Salykova, O., Ivanova, I., Bizhanova, O. and Aldasheva, D. (2025) Enhancing Electricity Consumption Forecasting in the Republic of Kazakhstan Using Machine Learning. *Journal of Applied Engineering and Technological Science (JAETS)*, **6**, 1166-1196. <https://doi.org/10.37385/jaets.v6i2.7425>
- [11] Huang, K. and Flowers, B. (2025) Parametric Optimisation of Building Envelopes for Energy Efficiency in Taiwan Region's Cooling-Dominated Climates. *International Journal of Sustainable Energy*, **44**, Article ID: 2560863. <https://doi.org/10.1080/14786451.2025.2560863>
- [12] Tsai, W. (2021) Trend Analysis of Taiwan Region's Greenhouse Gas Emissions from the Energy Sector and Its Mitigation Strategies and Promotion Actions. *Atmosphere*, **12**, Article 859. <https://doi.org/10.3390/atmos12070859>
- [13] Khosravi, F., Lowes, R. and Ugalde-Loo, C.E. (2023) Cooling Is Hotting up in the UK. *Energy Policy*, **174**, Article ID: 113456. <https://doi.org/10.1016/j.enpol.2023.113456>
- [14] Krzywanski, J., Sosnowski, M., Grabowska, K., Zylka, A., Lasek, L. and Kijo-Kleczkowska, A. (2024) Advanced Computational Methods for Modeling, Prediction and Optimization—A Review. *Materials*, **17**, Article 3521. <https://doi.org/10.3390/ma17143521>
- [15] Hochreiter, S. and Schmidhuber, J. (1997) Long Short-Term Memory. *Neural Computation*, **9**, 1735-1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- [16] Cho, K., van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., *et al.* (2014) Learning Phrase Representations Using RNN Encoder-Decoder for Statistical Machine Translation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Doha, 25-29 October 2014, 1724-1734. <https://doi.org/10.3115/v1/d14-1179>
- [17] Cohen, J. (1988) *Statistical Power Analysis for the Behavioral Sciences*. 2nd Edition, Routledge.
- [18] Song, H. and Li, G. (2008) Tourism Demand Modelling and Forecasting—A Review of Recent Research. *Tourism Management*, **29**, 203-220. <https://doi.org/10.1016/j.tourman.2007.07.016>
- [19] Witt, S.F. and Witt, C.A. (1991) *Modeling and Forecasting Demand in Tourism*. Academic Press.
- [20] Bergmeir, C., Hyndman, R.J. and Benítez, J.M. (2018) A Note on the Validity of Cross-Validation for Evaluating Autoregressive Time Series Prediction. *Computational Statistics & Data Analysis*, **120**, 70-83. <https://doi.org/10.1016/j.csda.2017.11.003>