Scientific Journal of Informatics
Vol. 13, No. 2, May 2026

p-ISSN 2407-7658 https://journal.unnes.ac.id/journals/sji/index e-ISSN 2460-0040

Hybrid CNN–LSTM–Transformer for Electrical Energy Consumption Forecasting: A Multi-Scenario Evaluation

Alief Flostyo Zulfiqor Roshif [1]*, Oky Dwi Nurhayati[2], Dinar Mutiara Kusumo Nugraheni[3]
[1]Master of Information Systems, Diponegoro University, Semarang, Indonesia
[2]Computer Engineering Department, Diponegoro University, Semarang, Indonesia
[3]Informatics Department, Diponegoro University, Semarang, Indonesia

Abstract.
Purpose: The purpose of this study is to develop hybrid deep learning model combining Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM), and Transformer mechanisms for short-term electricity consumption forecasting. The motivation arises from the limitations of existing studies, which are mostly based on single or dual hybrid models and have not fully captured complex temporal dependencies in electricity demand. To address this gap, the proposed CNN–LSTM–Transformer architecture is introduced to jointly model local patterns, sequential dependencies, and global temporal relationships, which have not been simultaneously explored in prior electricity forecasting studies. In addition, the study investigates the impact of multivariate inputs and feature engineering on forecasting performance.
Methods: The proposed model is evaluated using the Tétouan City Electricity Consumption dataset, which includes load data from three zones and meteorological variables. Data preprocessing involves cleaning, Min–Max normalization, and sequence windowing for supervised learning. Model performance is assessed using MAE, RMSE, and MAPE. To ensure comprehensive evaluation, six experimental scenarios are designed, including univariate and multivariate settings, per-zone and combined-zone configurations, as well as feature selection-based scenarios, to analyze accuracy, robustness, and generalization capability.
Result: The experimental results demonstrate that the proposed model achieves consistent forecasting performance across all scenarios. The best performance is obtained in the multivariate combined scenario (S4), with MAE values of 232–257 kWh, RMSE values of 385–433 kWh, and MAPE values ranging from 0.82% to 1.43% across all zones. The feature selection scenarios (S5 and S6) also show competitive performance, with MAE ranging from 234 to 324 kWh, RMSE from 393 to 590 kWh, and MAPE between 0.82% and 1.83%, indicating that engineered features can maintain prediction accuracy while reducing input complexity. In contrast, the univariate per-zone (S1) and multivariate per-zone (S3) scenarios produce higher errors, while the combined univariate scenario (S2) yields moderate improvements in certain zones. Overall, these findings confirm that integrating multivariate features with cross-zone data leads to the most accurate forecasting performance.
Novelty: The novelty of this study lies in the proposed integration of CNN, LSTM, and Transformer architectures into a unified hybrid framework for electricity consumption forecasting, which has not been widely explored in the energy demand domain. The model effectively combines local feature extraction, sequential dependency learning, and global attention mechanisms, and demonstrates strong capability in handling both univariate and multivariate electricity consumption data, including additional meteorological variables.
Practical Implications: This study shows that hybrid CNN-LSTM-Transformer can applied in electrical domain with multiple meteorological variables.

Keywords: CNN, LSTM, Transformer, Electrical energy, Electrical forecasting, Hybrid model
Received March 2026 / **Revised** April 2026 / **Accepted** May 2026

INTRODUCTION

Electricity consumption forecasting represents one of the most prominent examples of complex time-series modeling within the broader field of information systems. The increasing demand for electricity, driven by population growth, technological expansion, and economic development, has intensified the need for reliable forecasting frameworks. Indonesia, for instance, continues to experience sustained growth in electricity demand in parallel with demographic expansion and urban development. National statistics indicate that the country's population reached approximately 285.7 million in 2024 with an annual growth rate of around 1.11%. Correspondingly, the "Rencana Umum Penyediaan Tenaga Listrik (RUPTL)" issued

[1]*Corresponding author.
Email addresses: aliefflostyozr@gmail.com (Roshif)
DOI: 10.15294/sji.v13i2.44973

by the national electricity provider projects an average electricity demand growth of approximately 4.9% per year for the 2021–2030 period [1]. These trends highlight the importance of accurate electricity demand forecasting in supporting infrastructure planning, grid stability, and long-term energy policy formulation.

From an information systems perspective, electricity consumption data exhibit several characteristics that complicate predictive modeling. Load demand is typically influenced by a wide range of factors, including temporal patterns, environmental conditions, socioeconomic activities, and technological developments. As a result, electricity consumption data often display nonlinear dynamics, non-stationary structures, and high levels of volatility across different time horizons. Furthermore, electricity demand may vary significantly across geographical regions or consumption zones, reflecting differences in urban density, industrial activity, and climate conditions. These characteristics make electricity consumption datasets particularly suitable for evaluating the effectiveness of advanced predictive analytics techniques designed for complex time-series environments [2].

Despite the strategic importance of electricity forecasting, the implementation of advanced predictive models within operational information systems remains relatively limited in many developing contexts [3]. One major challenge lies in the dynamic and multifactorial nature of electricity consumption behavior. Environmental variables such as temperature, wind speed, solar radiation, and precipitation may influence electricity demand through their effects on heating, cooling, and renewable energy production [4]. At the same time, emerging technological trends—including the growing adoption of electric vehicles and smart home devices—further contribute to the variability of electricity consumption patterns [5]. These interacting influences produce data patterns that are often highly nonlinear and difficult to model using conventional statistical approaches.

Traditional time-series forecasting techniques such as Autoregressive Integrated Moving Average (ARIMA) and Vector Autoregression (VAR) have been widely used in electricity demand modeling, particularly due to their ability to capture linear temporal patterns and statistical relationships in energy systems. For example, ARIMA has been extensively applied in electricity consumption forecasting because of its capability to model trend and seasonality components in time-series data. However, ARIMA is fundamentally limited by its assumption of linearity and stationarity, making it less effective when dealing with highly nonlinear and complex electricity consumption patterns influenced by dynamic factors such as weather variations and user behavior [6]. Similarly, VAR models have been used to analyze multivariate relationships among energy-related variables, but they also rely on linear interdependencies and often struggle to capture nonlinear interactions between variables, especially in high-dimensional and highly dynamic datasets. For instance, VAR-based approaches in load forecasting may fail to accurately model sudden demand fluctuations caused by external conditions such as temperature spikes or irregular consumption behavior [7].

These models rely primarily on linear relationships within historical observations and have demonstrated effectiveness in relatively stable forecasting environments. However, their predictive capabilities often deteriorate when applied to large-scale datasets characterized by nonlinear structures and long-term dependencies [8]. In particular, conventional statistical models struggle to capture complex interactions among multiple variables and often fail to adapt to rapidly evolving consumption dynamics in modern energy systems [9]. These limitations have motivated researchers to explore more flexible modeling approaches capable of learning complex patterns directly from data.

In recent years, deep learning has emerged as a promising paradigm for predictive modeling in time-series analysis. Unlike traditional statistical methods, deep learning architectures can automatically learn hierarchical representations from raw data, enabling them to capture both local and global temporal relationships within sequential datasets [10]. This capability makes deep learning models particularly effective for handling complex, nonlinear patterns in electricity consumption data. For example, recurrent neural network (RNN)-based architectures such as LSTM and GRU have been widely applied in energy forecasting tasks due to their ability to model long-term dependencies [11]. A recent study demonstrated that an autoregressive GRU (AR-GRU) model achieved significantly higher forecasting accuracy compared to LSTM and GRU, reducing Mean Absolute Scaled Error (MASE) by up to 67% in short-term electricity consumption prediction [12]. Moreover, the model was able to maintain stable performance across both peak and off-peak demand periods, highlighting the robustness of deep learning approaches in capturing dynamic consumption patterns [13]. These findings provide strong evidence that deep learning architectures offer substantial advantages over traditional methods in modeling complex temporal dependencies in energy systems.

Among the various architectures that have been applied to energy forecasting tasks, Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM) networks, and Transformer models have demonstrated particularly strong performance. CNN architectures are effective in identifying local temporal patterns through convolutional feature extraction [14], while LSTM networks are specifically designed to model long-term dependencies within sequential data [15]. Transformer architectures, on the other hand, employ self-attention mechanisms that allow the model to capture global relationships across entire sequences without relying on strict sequential processing [16].

Although each of these architectures offers valuable modeling capabilities, relying on a single deep learning model may not be sufficient to fully capture the multifaceted characteristics of electricity consumption data. Electricity demand patterns often contain a combination of short-term fluctuations, long-term dependencies, and global contextual relationships that may not be adequately represented by any single modeling paradigm [17]. As a result, recent research has increasingly explored hybrid deep learning architectures that integrate multiple models within a unified predictive framework. Such hybrid approaches aim to leverage the complementary strengths of different architectures while mitigating their individual limitations [18].

Several recent studies have demonstrated the potential of hybrid deep learning models in energy forecasting applications. For example, Al-Ali et al. proposed a hybrid CNN–LSTM–Transformer architecture for solar energy production forecasting and reported significant improvements in prediction accuracy compared to conventional approaches [19]. Similarly, Ranjbar and Rahimzadeh developed a hybrid Transformer–LSTM–CNN model for forecasting gasoline consumption, highlighting the advantages of integrating multiple neural architectures for modeling complex consumption patterns [20]. These studies confirm that hybrid architectures can effectively combine local feature extraction, sequential memory mechanisms, and global attention-based modeling.

At a more detailed level, Binbusayyis and Sha introduced a deep CNN–BiLSTM model enhanced with an attention mechanism for predicting household electricity consumption, demonstrating improved performance relative to baseline machine learning approaches [21]. Other studies have also emphasized the effectiveness of Transformer-based architectures in capturing long-term temporal dependencies within smart grid environments [22]. However, despite these advances, most existing studies remain limited to specific application domains, single-region datasets, or relatively homogeneous data distributions [23]. As a result, their findings may not generalize well to more complex real-world scenarios involving heterogeneous electricity consumption patterns across multiple zones.

More importantly, previous studies often do not critically address the integration of all three key modeling aspects simultaneously: local temporal feature extraction, long-term sequential dependency learning, and global contextual relationship modelling [24]. In many cases, hybrid models emphasize only partial combinations of these components. For instance, Lou et al. proposed a CNN–Transformer architecture for predicting electric vehicle driving range [25], which effectively captures local and global patterns but lacks recurrent mechanisms such as LSTM to explicitly model sequential dependencies over time. Conversely, models that incorporate recurrent layers often do not fully leverage global attention mechanisms or fail to optimize cross-feature interactions in multivariate settings [16]. This indicates that existing approaches are still insufficient in capturing the full complexity of temporal dynamics in electricity consumption data.

In addition, the majority of prior studies do not explicitly address the challenge of multi-zone forecasting, where consumption patterns vary significantly across different spatial regions [26]. The absence of integrated cross-zone modeling limits the ability of these models to learn shared patterns and reduces their robustness when applied to large-scale energy systems [27]. Consequently, the problem of developing a unified predictive framework that can effectively handle multivariate inputs, heterogeneous spatial characteristics, and multi-scenario training strategies remains unresolved.

Based on these limitations, this study aims to address the identified research gap by proposing a hybrid CNN–LSTM–Transformer model designed to simultaneously capture local patterns, sequential dependencies, and global contextual relationships within multi-zone electricity consumption data. In addition, this study introduces a multi-scenario experimental framework to evaluate the robustness of the model under different data configurations, thereby providing a more comprehensive understanding of its performance in real-world forecasting applications.

The objective of this research is therefore to design, implement, and evaluate the proposed hybrid CNN–LSTM–Transformer architecture for short-term electricity consumption forecasting. Model performance is assessed through quantitative evaluation metrics including Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE) [24]. Through this evaluation, the study aims to provide empirical insights into the effectiveness of hybrid deep learning architectures for modeling complex electricity consumption.

Ultimately, this research contributes to the advancement of predictive analytics within the field of information systems by demonstrating the applicability of hybrid deep learning models for large-scale time-series forecasting. The findings of this study are expected to support the development of intelligent energy management systems capable of delivering accurate electricity demand predictions, thereby facilitating more efficient operational planning and decision-making in modern smart grid infrastructures.

METHODS

Research Framework

This study adopts a structured experimental framework consisting of several sequential stages, beginning with data acquisition and preparation, followed by model development, training, and performance evaluation [28]. The overall research workflow is illustrated in Figure 1. The process begins with collecting and preprocessing the electricity consumption dataset to ensure that the data are suitable for machine learning analysis. After the preprocessing stage, the proposed hybrid deep learning architecture is constructed and trained using the prepared dataset.

Model performance is then evaluated using several regression-based evaluation metrics. If the model performance does not meet the expected criteria, and model refinement are conducted iteratively until satisfactory predictive accuracy is obtained. This iterative process ensures that the model configuration is optimized prior to the final evaluation stage [29]. The evaluation phase assesses the predictive capability of the model using quantitative indicators including Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE). These metrics provide complementary perspectives for assessing prediction accuracy and stability [30].

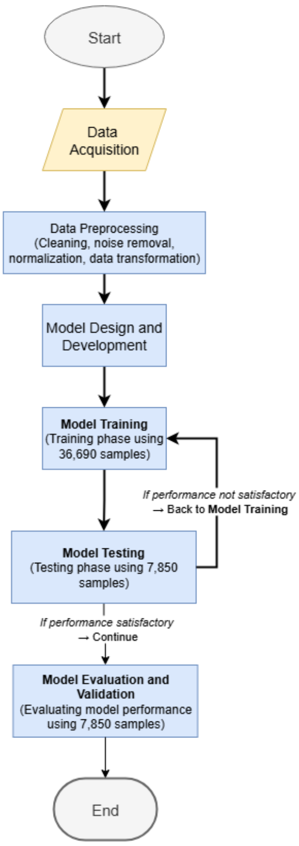

Figure 1. Research flow for electricity consumption forecasting using hybrid CNN–LSTM–transformer

Dataset and Preprocessing

The dataset utilized in this research is the Tétouan City Electricity Consumption Dataset, which contains electricity demand measurements collected from three distribution zones in Tétouan, Morocco, namely Quads, Smir, and Boussafou. The dataset consists of a total of approximately 52,416 observations, recorded at a time resolution of ten minutes over a one-year period. Each observation includes electricity consumption values for all three zones, resulting in a balanced data distribution across zones, where each zone has an equal number of time-series records [31]. This structure enables both per-zone and combined multi-zone modeling scenarios to be effectively implemented and evaluated.

In addition to electricity load data, the dataset includes several meteorological variables that potentially influence energy consumption, such as temperature, humidity, diffuse flow, and general diffuse flow. These variables provide important contextual information for modeling the relationship between environmental conditions and electricity demand.

The Tétouan dataset is selected in this study due to several key advantages. First, it offers high temporal resolution data, which is essential for capturing fine-grained consumption patterns and short-term fluctuations. Second, the availability of multi-zone electricity consumption data allows for the analysis of spatial variability and supports the development of models capable of handling heterogeneous consumption behavior. Third, the inclusion of multiple meteorological features makes the dataset suitable for multivariate time-series forecasting [32]. These characteristics make the dataset an appropriate benchmark for evaluating the performance of advanced hybrid deep learning models in complex electricity consumption forecasting scenarios.

Prior to model training, the raw dataset underwent several preprocessing procedures to improve data quality and ensure compatibility with the deep learning framework. The preprocessing stage began with data cleaning, focusing on handling missing values and reducing noise in the dataset. Missing values were addressed using time-based linear interpolation, which estimates missing observations based on adjacent temporal points [33]. This approach was selected because electricity consumption data exhibit strong temporal continuity, where values at nearby time steps are highly correlated. Therefore, linear interpolation provides a reasonable approximation without introducing abrupt changes in the data distribution. To further ensure data consistency, forward-fill and backward-fill methods were applied to handle any remaining boundary missing values.

In addition, outliers in electricity consumption variables were detected using the z-score method and replaced using a rolling median approach. This strategy was chosen to mitigate the influence of extreme values while preserving the underlying temporal patterns of the data. For meteorological variables, a smoothing process using a rolling mean was applied to reduce short-term noise and enhance signal stability. Following data cleaning, feature scaling was applied to facilitate efficient neural network training. All numerical variables were normalized using Min–Max scaling within the range of 0 to 1. This normalization technique was selected because deep learning models are sensitive to the scale of input features [34]. By transforming all variables into a uniform range, Min–Max scaling helps prevent features with larger magnitudes from dominating the learning process, improves numerical stability, and accelerates convergence during optimization.

The final preprocessing step involved transforming the time-series data into a supervised learning format using a sliding window approach. A fixed window size of 24 time steps was used as input, representing the most recent historical observations prior to prediction. This choice was made to capture short-term temporal dependencies within a four-hour period, given the dataset's 10-minute resolution [35]. The model is trained to predict the subsequent time step based on this input sequence, following a single-step forecasting strategy. The use of a 24-step window balances the need to capture sufficient temporal context while maintaining computational efficiency and reducing the risk of overfitting.

After preprocessing, the dataset was divided into training and testing sets using a temporal split, where 80% of the data was allocated for training and the remaining 20% for testing. This split preserves the chronological order of the data, ensuring that future information is not used during model training and maintaining the integrity of the forecasting evaluation [36].

Hybrid CNN–LSTM–Transformer Architecture

The predictive model proposed in this research integrates three deep learning architectures into a unified hybrid framework: Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM) networks, and Transformer encoders. Each component contributes a specific capability in capturing different temporal characteristics of electricity consumption data. The overall model architecture is illustrated in Figure 2.

CNN layers are employed to extract localized temporal features from sequential input data. By applying convolutional filters across time-series segments, CNNs are able to identify short-term fluctuations and repeating patterns in electricity consumption [36].

The extracted features are then passed to LSTM layers, which are designed to capture long-term temporal dependencies within sequential data. LSTM networks incorporate memory cells and gating mechanisms that enable the model to retain relevant information across extended time intervals while selectively discarding less important signals [37].

Finally, the Transformer encoder component introduces an attention-based mechanism that allows the model to analyze relationships among different time steps simultaneously. Through multi-head self-attention, the Transformer can identify global dependencies across the entire sequence without relying solely on sequential processing [38]. The combination of CNN, LSTM, and Transformer components enables the model to represent local patterns, sequential dynamics, and global contextual relationships within electricity consumption data.

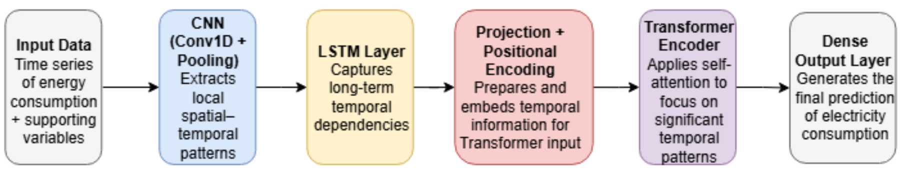

Figure 2. Hybrid CNN–LSTM–transformer architecture for electricity consumption forecasting

Model Configuration

The hybrid CNN–LSTM–Transformer model was implemented using the TensorFlow and Keras deep learning framework. The architecture was designed to integrate convolutional feature extraction, sequential modeling, and attention-based learning within a unified pipeline.

The input layer receives a sequence of time-series data with a fixed window size of 24 time steps. The first component of the model is a one-dimensional Convolutional Neural Network (CNN) layer with 64 filters and a kernel size of 3, followed by a Rectified Linear Unit (ReLU) activation function. This layer is responsible for extracting local temporal patterns and short-term fluctuations in electricity consumption data.

The extracted features are then passed to a Long Short-Term Memory (LSTM) layer with 64 hidden units and the *return_sequences* parameter enabled. This configuration allows the model to retain sequential information across all time steps and pass the full sequence representation to the subsequent attention mechanism.

Next, a Transformer-inspired attention block is applied using a Multi-Head Attention layer with 4 attention heads and a key dimension of 32. The attention output is combined with the LSTM output through a residual connection, followed by layer normalization. This component enables the model to capture global dependencies and contextual relationships across the entire sequence.

The final stage of the model consists of an additional LSTM layer with 32 units to aggregate the learned temporal features, followed by a fully connected dense layer with 32 neurons and ReLU activation. The output layer uses a single neuron with a linear activation function to produce continuous electricity consumption predictions.

Model Training and Optimization

The model training process was conducted using the Adam optimizer with a learning rate of 0.001, which was selected due to its effectiveness in handling non-stationary and noisy time-series data. The loss function used during training was Mean Squared Error (MSE), as it penalizes large prediction errors and promotes stable convergence [39].

Training was performed for a maximum of 50 epochs with a batch size of 32. A validation split of 10% was applied to the training data to monitor model generalization during training. To prevent overfitting, an Early Stopping strategy was implemented, where training was automatically terminated if the validation loss did not improve for 5 consecutive epochs. The best model weights were restored based on the lowest validation loss [40].

The model training process follows a structured pipeline. First, the preprocessed and normalized dataset is divided into training and testing sets using a temporal split. Second, input-output sequences are generated using a sliding window approach with a window size of 24. Third, the model is trained using the training sequences while validation performance is monitored. Finally, the trained model is evaluated on the unseen test data using MAE, RMSE, and MAPE metrics.

To ensure reproducibility, a fixed random seed (SEED = 42) was applied across all libraries, including NumPy, TensorFlow, and Python's random module. This ensures consistent weight initialization and data processing across experiments. All experiments were conducted in a GPU-enabled environment, which significantly reduced training time and improved computational efficiency [41].

It should be noted that hyperparameter tuning was not extensively performed in this study. Instead, model parameters such as the number of filters, hidden units, and attention heads were selected based on commonly used configurations in prior studies and preliminary experiments. This approach ensures a fair evaluation of the proposed architecture across multiple experimental scenarios while maintaining computational feasibility.

Model Components

Input layer

The input layer receives sequential electricity consumption observations together with auxiliary meteorological variables such as temperature, humidity, wind speed, and solar radiation. These variables provide contextual information that may influence electricity demand.

CNN Layer (Conv1D + Pooling)
The convolutional layer extracts local temporal patterns from sequential data by applying convolutional filters across the input sequence. The convolution operation can be expressed as equation (1):

$$y_i = f(\sum_{k=1}^{K} w_k \cdot x_{i+k-1} + b) \tag{1}$$

where y_i denotes the output feature at position i, x_i represents the input sequence, w_k denotes the convolution filter weights of size K, and b is the bias term. The function $f(\cdot)$ refers to the ReLU activation function applied element-wise to introduce non-linearity into the model [42].
Pooling operations are then applied to the feature maps to reduce spatial dimensionality and redundancy while retaining the most salient features.

LSTM layer
The LSTM captures long-term temporal dependencies from CNN features, as shown in equation (2).

$$\begin{aligned} f_t &= \sigma(W_f[h_{t-1}, x_t] + b_f) \\ i_t &= \sigma(W_i[h_{t-1}, x_t] + b_i) \\ \tilde{C}_t &= \tanh\,(W_C[h_{t-1}, x_t] + b_C) \\ C_t &= f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \\ h_t &= \tanh(C_t) \odot \sigma(W_o[h_{t-1}, x_t] + b_o) \end{aligned} \tag{2}$$

The LSTM layer is designed to capture long-term temporal dependencies from features extracted by the CNN. At each timestep, the LSTM receives input from the CNN (x_t) as well as the previous hidden state and cell state (h_{t-1} and C_{t-1}). Three main gates—forget gate (f_t), input gate (i_t), and output gate (o_t)—control the flow of information into and out of the cell state. The candidate cell state ($\tilde{C}_t$) generates new information to be stored. The current hidden state (h_t) is computed by combining the output gate and the updated cell state. Weights (W_f, W_i, W_C, W_o) and biases (b_f, b_i, b_C, b_o) govern the linear transformations within each gate. Element-wise multiplication ($\odot$) selectively merges information between components, and $\sigma(\cdot)$ represents the sigmoid activation function used to regulate information flow within the gates [43].

Projection + Positional Encoding
The Projection and Positional Encoding layer introduces temporal order information into the LSTM output before it enters the Transformer, using sinusoidal encoding as shown in equation (3).

$$PE_{(pos,2i)} = \sin\,(\frac{pos}{10000^{2i/d}}), PE_{(pos,2i+1)} = \cos\,(\frac{pos}{10000^{2i/d}}) \tag{3}$$

After the LSTM layer, the output is projected and augmented with temporal order information using positional encoding. Positional encoding employs sinusoidal functions to mark the position of each timestep within the sequence, allowing the Transformer to capture temporal relationships even though it lacks the sequential structure of an RNN. The variable pos represents the timestep position, i is the embedding dimension index, and d is the total embedding dimension. This ensures that temporal information is incorporated into the Transformer input without losing sequential relationships [44].

Transformer Encoder
The encoder employs *multi-head self-attention* to identify global dependencies, as shown in equation (4).

$$\text{Attention}(Q, K, V) = \text{softmax}(\frac{QK^T}{\sqrt{d_k}})V \tag{4}$$

The Transformer encoder leverages multi-head self-attention to capture global dependencies within sequential data. Queries (Q), keys (K), and values (V) are derived from the projected LSTM outputs. Attention is computed by comparing queries and keys and normalizing via the softmax function. The attention output is then multiplied by the values to generate sequence representations that account for relationships between all timesteps. The key dimension (d_k) is used to stabilize the attention score calculations [45].

Dense output layer

The final dense layer produces predictions for future electricity consumption. Its input comes from the output of the Transformer encoder. Weights (W_{dense}) and biases (b_{dense}) perform linear transformation, and the linear activation function ensures that the output values can be continuous, corresponding to the scale of electricity consumption.

Model Evaluation

To assess the predictive performance of the proposed model, three commonly used regression evaluation metrics were employed: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE).

MAE measures the average magnitude of prediction errors without considering their direction. RMSE penalizes larger errors more strongly by computing the square root of the mean squared prediction errors. MAPE expresses prediction error in percentage form, enabling easier comparison across datasets with different scales [46].

The formulas for these metrics are shown in Equations (5), (6), and (7), where $y_i$represents the actual value, $\hat{y}_i$represents the predicted value, and $n$denotes the number of observations.

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n} | y_i - \hat{y}_i | \tag{5}$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{6}$$

$$\text{MAPE} = \frac{100}{n}\sum_{i=1}^{n} | \frac{y_i - \hat{y}_i}{y_i} | \tag{7}$$

Experimental Scenarios

To comprehensively evaluate the robustness and generalization capability of the proposed model, six experimental scenarios were designed. Each scenario introduces different data configurations, feature engineering strategies, and training paradigms, while maintaining a consistent Hybrid CNN–LSTM–Transformer architecture and training setup across all experiments.

Scenario 1 – Univariate Per-Zone Evaluation (S1).

The model is trained and evaluated independently for each electricity consumption zone. Only historical electricity consumption is used as input. This scenario serves as a baseline to assess zone-specific predictive performance under homogeneous data conditions.

Scenario 2 – Univariate Combined Evaluation (S2).

All zones are combined into a single dataset, and a zone identifier is included as an additional feature. Only electricity consumption is used as the primary predictor. This scenario evaluates the benefit of leveraging a larger training dataset while preserving zone distinction.

Scenario 3 – Multivariate Per-Zone Evaluation (S3).

The model is trained separately for each zone using both electricity consumption and meteorological variables (temperature, humidity, wind speed, atmospheric pressure, and solar radiation). This scenario evaluates the impact of external environmental factors on zone-specific forecasting performance.

Scenario 4 – Multivariate Combined Evaluation (S4).

All zones are merged into a single dataset, and both electricity consumption and meteorological variables are used as input features, along with a zone identifier. This scenario evaluates whether combining datasets across zones improves learning of complex temporal–spatial relationships.

Scenario 5 – Feature Selection Per-Zone (S5).

The model is trained separately for each zone using a reduced set of selected features, including lag-based and rolling statistics (e.g., lag-1, lag-24, rolling mean). This scenario evaluates the effectiveness of feature engineering in improving predictive performance under zone-specific conditions.

Scenario 6 – Feature Selection Combined (S6).
All zones are combined into a single dataset with a zone identifier, and selected engineered features (lag and rolling-based features) are used as inputs. This scenario evaluates the combined effect of feature engineering and cross-zone learning on forecasting accuracy and generalization capability.

Summary of Experimental Framework
Through these six experimental scenarios, the proposed evaluation framework systematically examines model accuracy, feature sensitivity, the effect of data aggregation across zones, and the impact of engineered temporal features. This design enables a comprehensive assessment of the Hybrid CNN–LSTM–Transformer model in both isolated and generalized settings.

To facilitate clear comparison, the experimental configurations are summarized in Table 1, including data structure, feature usage, training strategy, and evaluation objectives. All scenarios are evaluated using MAE, RMSE, and MAPE to ensure consistent performance comparison across different experimental settings.

Table 1. Summary of experimental scenarios

Scenario	Description	Data Configuration	Feature Strategy	Evaluation Objective
S1	Univariate Per-Zone Evaluation	Each zone (Zone 1, Zone 2, Zone 3) trained and tested independently	Raw electricity consumption only	Establish baseline performance for each zone under isolated conditions
S2	Univariate Combined Evaluation	All zones combined with zone identifier	Raw electricity consumption + zone_id	Evaluate benefit of data aggregation across zones while preserving zone identity
S3	Multivariate Per-Zone Evaluation	Each zone trained separately	Electricity consumption + meteorological variables	Assess impact of external environmental factors on zone-specific forecasting
S4	Multivariate Combined Evaluation	All zones combined with zone identifier	Electricity consumption + meteorological variables + zone_id	Evaluate joint effect of multivariate features and cross-zone learning
S5	Feature Selection Per-Zone	Each zone trained separately	Selected engineered features	Evaluate effectiveness of feature engineering in zone-specific modeling
S6	Feature Selection Combined Evaluation	All zones combined with zone identifier	Selected engineered features + zone_id	Assess combined impact of feature engineering and cross-zone generalization

RESULTS AND DISCUSSIONS
The experimental results demonstrate that the proposed Hybrid CNN–LSTM–Transformer model consistently achieves high prediction accuracy across six experimental scenarios involving different data configurations, feature strategies, and training paradigms. The overall performance comparison is summarized in Table 2, which reports MAE, RMSE, and MAPE values for each zone. As shown in the table, the best performance is achieved in the combined and feature-engineered scenarios, particularly S4 (Multivariate Combined) and S6 (Feature Selection Combined), where the model attains MAE values as low as 257 kWh (Zone 1) and 232–234 kWh (Zone 3), with MAPE values generally below 1.5%. In addition, Zone 2 consistently produces the lowest prediction errors across all scenarios (MAE 229–250 kWh), indicating that its consumption pattern is relatively more stable and easier to learn compared to the other zones.

Table 2. Performance comparison across experimental scenarios

Scenario	Zone	MAE (kWh)	RMSE (kWh)	MAPE (%)
S1	Zone 1	364	549	1.14
	Zone 2	232	440	1.14
	Zone 3	338	640	2.24
S2	Zone 1	353	536	1.10

Scenario	Zone	MAE (kWh)	RMSE (kWh)	MAPE (%)
	Zone 2	229	457	1.14
	Zone 3	247	588	1.53
S3	Zone 1	489	663	1.54
	Zone 2	235	426	1.15
	Zone 3	228	468	1.40
S4	Zone 1	257	423	0.82
	Zone 2	245	385	1.23
	Zone 3	232	433	1.43
S5	Zone 1	324	520	1.02
	Zone 2	248	454	1.19
	Zone 3	282	590	1.83
S6	Zone 1	257	433	0.82
	Zone 2	250	393	1.27
	Zone 3	234	466	1.47

The results in Table 2 highlight several important insights. First, the transition from univariate to multivariate configurations (S1 → S3 and S2 → S4) shows that incorporating meteorological variables improves prediction accuracy, particularly in combined training settings. However, the improvement is not substantial, indicating that historical electricity consumption remains the dominant predictive signal. Second, combining data across zones (S2, S4, S6) consistently yields better or more stable performance compared to per-zone training (S1, S3, S5), demonstrating the benefit of larger and more diverse training data. Third, feature engineering through lag-based and rolling statistics (S5 and S6) further enhances performance, especially when combined with cross-zone learning, as seen in S6. Among all configurations, S4 and S6 provide the most balanced and robust results, confirming that the integration of multivariate features, feature engineering, and combined training contributes significantly to model accuracy. Nevertheless, slightly higher RMSE values in certain cases (e.g., up to 663 kWh in S3) indicate that the model remains sensitive to sudden fluctuations and irregular consumption patterns.

To further validate these findings, Figure 3 and Figure 4 illustrate the comparison between actual and predicted electricity consumption for the best-performing scenarios, namely S4 (Multivariate Combined) and S6 (Feature Selection Combined). These scenarios are selected because they consistently achieve the lowest prediction errors and represent the most effective configurations in this study.

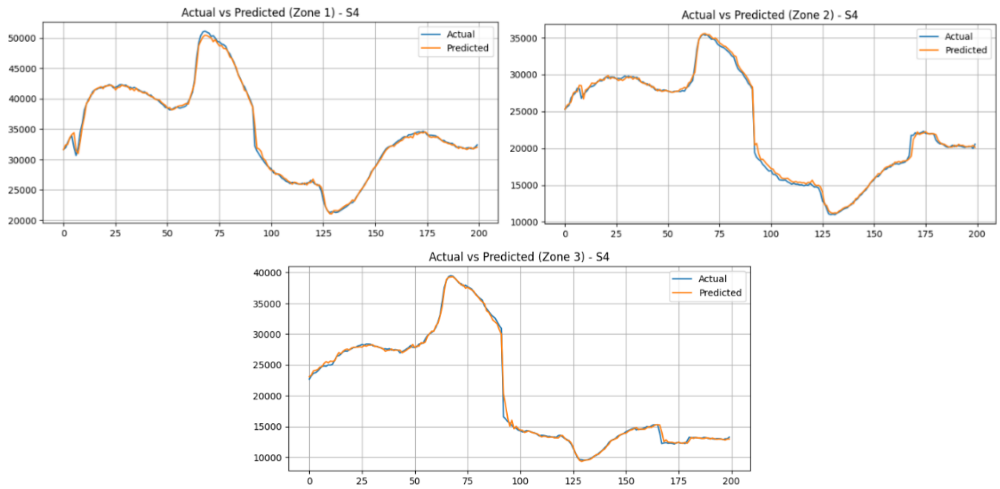

Figure 3. Comparison between actual and predicted electricity consumption for Scenario S4 across three zones

The results in Figure 3 show that the predicted values closely follow the actual consumption patterns across all zones, effectively capturing both short-term fluctuations and overall temporal trends. The model demonstrates strong alignment with the actual data, particularly in Zone 1 and Zone 3, where minimal deviation is observed even during peak demand periods. This indicates that the combination of multivariate features and cross-zone learning enables the model to better understand complex temporal–spatial relationships in electricity consumption.

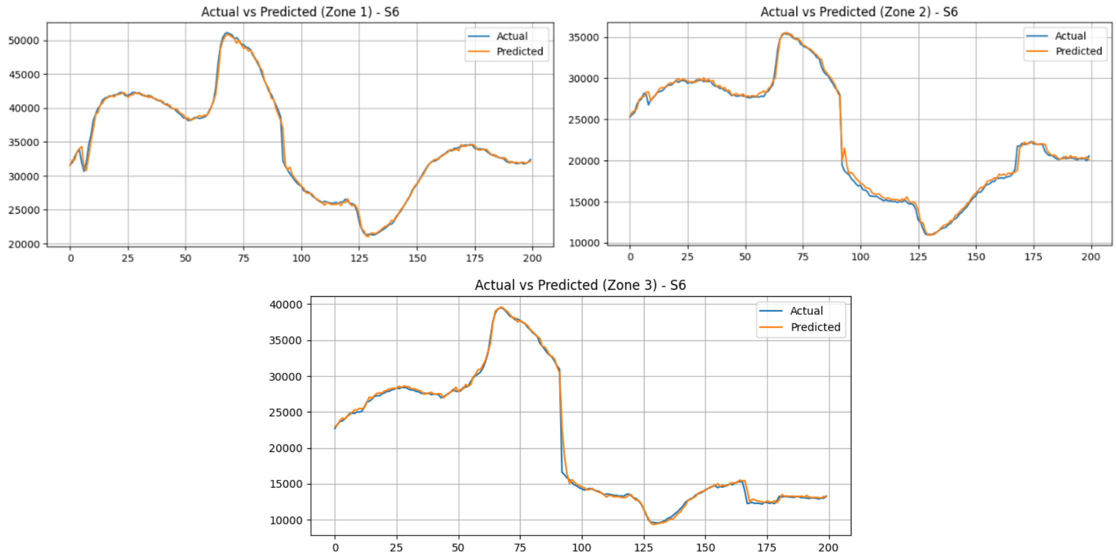

Figure 4. Comparison between actual and predicted electricity consumption for Scenario S6 across three zones

Similarly, Figure 4 demonstrates that the model maintains stable and accurate prediction performance under the feature-engineered combined configuration. The predicted curves consistently track the actual consumption patterns, confirming that the integration enhances the model's ability to capture temporal dependencies. Although minor discrepancies are observed during abrupt changes in demand, these deviations remain relatively small and do not significantly affect overall forecasting accuracy.

Overall, both quantitative and visual results confirm that the proposed Hybrid CNN–LSTM–Transformer model is capable of effectively modeling complex temporal dynamics in electricity consumption data. The integration of convolutional feature extraction, sequential memory, and attention-based global modeling enables the model to capture local patterns, long-term dependencies, and global temporal relationships simultaneously. Despite its strong performance, the model still exhibits limitations in handling highly irregular consumption patterns, suggesting that future improvements may focus on enhancing robustness under extreme variability conditions.

Comparison with Baseline Models

Table 3 presents a comparative evaluation between the proposed Hybrid CNN–LSTM–Transformer model and several baseline models, including LSTM, CNN, Transformer, and their hybrid variants (CNN–LSTM, CNN–Transformer, and LSTM–Transformer) across three electricity consumption zones in the Tétouan dataset. The evaluation is conducted using standard regression metrics, namely MAE, RMSE, and MAPE, to measure prediction accuracy and robustness.

Table 3. Performance comparison with baseline models [20], [21]

Zone	Model	MAE (kWh)	RMSE (kWh)	MAPE (%)
	LSTM	2500	3400	11,85
	CNN	2500	4300	17,45
	Transformer	3400	3600	11,31
Zone 1	CNN–LSTM	2851,70	3815,98	11,47
	CNN–Transformer	1351,99	989,17	3,05
	LSTM–Transformer	1194,03	837,42	2,61
	CNN–LSTM–Transformer	257	423	0,82
	LSTM	1500	2200	14,86
	CNN	2700	3300	17,93
Zone 2	Transformer	1000	2200	13,54
	CNN–LSTM	2851,70	3815,98	11,47
	CNN–Transformer	980,06	707,08	3,43

	LSTM–Transformer	879,01	616,74	3,01
	CNN–LSTM–Transformer	245	385	1,23
Zone 3	LSTM	1000	1500	24,62
	CNN	3500	4500	20,24
	Transformer	1000	1700	26,16
	CNN–LSTM	2851,70	3815,98	11,47
	CNN–Transformer	1286,70	885,65	5,28
	LSTM–Transformer	1106,20	765,07	4,74
	CNN–LSTM–Transformer	232	433	1,43

The experimental results clearly indicate that the proposed Hybrid CNN–LSTM–Transformer model consistently outperforms all baseline models across all zones in terms of MAE, RMSE, and MAPE. This confirms the effectiveness of combining convolutional feature extraction, sequential modeling, and attention-based global dependency learning within a unified architecture.

In Zone 1, the proposed model achieves a substantial reduction in prediction error, with an MAE of 257 kWh, RMSE of 423 kWh, and MAPE of 0.82%. These values are significantly lower compared to standalone models such as LSTM (MAE 2500 kWh, MAPE 11.85%), CNN (MAE 2500 kWh, MAPE 17.45%), and Transformer (MAE 3400 kWh, MAPE 11.31%). Even compared to hybrid baselines, such as CNN–Transformer (MAE 1351.99 kWh) and LSTM–Transformer (MAE 1194.03 kWh), the proposed model demonstrates a markedly superior performance, indicating improved capability in capturing both local and global temporal patterns.

A similar trend is observed in Zone 2, where the proposed model achieves an MAE of 245 kWh, RMSE of 385 kWh, and MAPE of 1.23%. These results significantly outperform all baseline architectures, including LSTM (MAE 1500 kWh), CNN (MAE 2700 kWh), and Transformer (MAE 1000 kWh), as well as hybrid variants such as CNN–Transformer (MAE 980.06 kWh) and LSTM–Transformer (MAE 879.01 kWh). The reduction in error indicates that the proposed model is particularly effective in learning more stable consumption patterns when trained on combined zone structures.

In Zone 3, which exhibits higher variability in consumption patterns, the proposed model still achieves the best performance with an MAE of 232 kWh, RMSE of 433 kWh, and MAPE of 1.43%. This is a significant improvement over LSTM (MAE 1000 kWh), CNN (MAE 3500 kWh), and Transformer (MAE 1000 kWh). Although hybrid baselines such as LSTM–Transformer (MAE 1106.20 kWh) and CNN–Transformer (MAE 1286.70 kWh) provide partial improvements, they remain considerably less accurate than the full proposed architecture.

Overall, the results demonstrate that each component of the hybrid architecture contributes to performance improvement in different ways. CNN effectively extracts local temporal patterns, LSTM captures sequential dependencies, while the Transformer enhances global temporal relationship modeling through self-attention mechanisms. However, the full integration of all three components provides complementary strengths, resulting in a more robust representation of electricity consumption dynamics. In comparison to standalone deep learning models, the proposed hybrid model achieves an overall error reduction of approximately 80–90% in MAE and MAPE, highlighting its strong predictive capability and stability across different zones and consumption patterns.

Comparison with Classical and Single Deep Learning Models

The dataset used in this study includes electricity consumption statistics across multiple zones, with the average power consumption across three zones used as the target variable shown in equation (8).

$$y_{avg} = \frac{\text{PowerConsumption Zone1} + \text{PowerConsumption Zone2} + \text{PowerConsumption Zone3}}{3} \quad (8)$$

Several forecasting models were evaluated in this study, including traditional statistical methods such as ARIMA and Exponential Smoothing, as well as deep learning-based models such as LSTM, CNN, Transformer, and hybrid architectures. Prior to model training, data normalization was applied to ensure numerical stability and to improve the learning process across different model architectures.

Table 4 presents the comparative performance of all evaluated models in terms of MAE, RMSE, and MAPE. The results clearly show that the proposed Hybrid CNN–LSTM–Transformer model significantly outperforms all baseline methods across all evaluation metrics.

Table 4. Performance comparison with other study [20], [21]

Model	MAE (kWh)	RMSE (kWh)	MAPE (%)
LSTM	3046,18	4041,98	12,76
CNN	4161,78	5113,12	18,40
Transformer	3177,31	4067,59	13,87
ARIMA	5381,19	6882,49	20,67
Exponential Smoothing	4731,19	5960,81	19,63
CNN–LSTM–Transformer	244,67	413,67	1,16

The proposed model achieves the lowest error values, with an MAE of 244.67 kWh, RMSE of 413.67 kWh, and MAPE of 1.16%. These results demonstrate a substantial improvement over both traditional statistical models and standalone deep learning approaches.

Among the classical forecasting methods, ARIMA and Exponential Smoothing show relatively higher error values, with MAE reaching 5381.19 kWh and 4731.19 kWh, respectively. This indicates that these linear models are less capable of capturing the nonlinear and highly dynamic characteristics of electricity consumption data.

Similarly, standalone deep learning models such as LSTM, CNN, and Transformer achieve better performance than classical methods but still exhibit significantly higher prediction errors compared to the proposed hybrid model. For instance, LSTM achieves an MAE of 3046.18 kWh, while CNN and Transformer record MAE values of 4161.78 kWh and 3177.31 kWh, respectively. These results suggest that each individual architecture captures only partial aspects of the underlying temporal patterns.

The superior performance of the proposed Hybrid CNN–LSTM–Transformer model can be attributed to its integrated architecture. The CNN component extracts local temporal patterns, the LSTM captures sequential dependencies over time, and the Transformer enhances global dependency modeling through self-attention mechanisms. The combination of these three components enables a more comprehensive representation of electricity consumption dynamics.

Overall, the proposed model achieves an error reduction of more than 90% in MAE and MAPE compared to traditional and baseline deep learning models, demonstrating its effectiveness, robustness, and suitability for complex time-series forecasting tasks.

Discussion of Model Behavior, Feature Influence, and Comparison with Previous Studies

To provide a deeper interpretation of the experimental findings, this study analyzes performance variation across zones and evaluates the impact of different data configurations, including univariate, multivariate, and feature-engineered scenarios (S1–S6). The results consistently show that Zone 2 achieves the lowest prediction errors across all scenarios (MAE 229–250 kWh), while Zone 3 exhibits relatively higher RMSE and MAPE values. This pattern reflects the intrinsic characteristics of electricity consumption in each zone. Zone 2 demonstrates more stable and regular temporal behavior, allowing the model to learn consistent patterns effectively under both per-zone (S1, S3, S5) and combined configurations (S2, S4, S6). In contrast, Zone 3 presents higher variability and irregular fluctuations, which increase prediction uncertainty and lead to larger error values, particularly in scenarios with less contextual information (e.g., S1 and S3). These findings indicate that forecasting performance is not only influenced by model architecture but also strongly dependent on the volatility and complexity of the underlying data distribution.

From a configuration perspective, the comparison across scenarios highlights the impact of data aggregation and feature design. Moving from per-zone modeling (S1) to combined learning with a zone identifier (S2) leads to consistent performance improvement, suggesting that cross-zone data aggregation helps the model learn more generalized temporal representations while still preserving zone-specific characteristics. This trend becomes more evident in multivariate settings, where the combined scenario (S4) outperforms the per-zone multivariate scenario (S3), indicating that integrating both spatial (cross-zone) and contextual (meteorological) information enhances the model's ability to capture complex temporal–spatial dependencies. Similarly, feature engineering scenarios (S5 and S6) demonstrate improves prediction stability, with S6 (feature selection combined) achieving performance comparable to S4. This suggests that engineered temporal features can effectively substitute part of the raw multivariate information by explicitly encoding historical patterns.

The analysis also reveals a clear relationship between meteorological variables and prediction accuracy. Scenarios that incorporate weather-related features (S3 and S4) generally achieve better performance than purely univariate approaches (S1 and S2), confirming that external environmental factors contribute to explaining variations in electricity consumption. Variables such as temperature and humidity influence cooling and heating demand, while solar radiation and wind conditions affect energy usage patterns. However, the improvement remains relatively moderate, particularly when compared to feature-engineered scenarios (S5 and S6). This indicates that while meteorological variables provide useful contextual signals, historical electricity consumption remains the dominant predictive factor. In other words, meteorological features improve model sensitivity to external variations, whereas engineered temporal features strengthen the model's ability to capture intrinsic consumption dynamics [47].

Overall, the results demonstrate that the best performance is achieved in scenarios that combine multiple strengths, particularly S4 and S6, where the model benefits from both cross-zone learning and enriched feature representations. These findings highlight the importance of jointly considering data structure (per-zone vs combined), feature type (raw vs engineered vs multivariate), and temporal characteristics when designing electricity forecasting models. At the same time, the presence of higher RMSE values in certain zones indicates that the model remains sensitive to sudden fluctuations, suggesting that highly volatile consumption patterns continue to pose challenges even under advanced hybrid architectures [48].

When compared with previous studies, the superiority of the proposed Hybrid CNN–LSTM–Transformer model becomes more evident. Prior research shows that standalone deep learning models such as LSTM, CNN, and Transformer produce significantly higher errors, with MAE values reaching 2500–3400 kWh in Zone 1 and MAPE values above 11%. Even hybrid two-model approaches, such as CNN–Transformer and LSTM–Transformer, still yield MAE values above 800–1300 kWh. In contrast, the proposed model achieves substantially lower errors, with MAE values of 257 kWh (Zone 1), 245 kWh (Zone 2), and 232 kWh (Zone 3), and MAPE values as low as 0.82%. Similarly, when compared with classical statistical models such as ARIMA and Exponential Smoothing, which exhibit MAE values above 4700–5300 kWh and MAPE exceeding 19%, the proposed model demonstrates an error reduction of more than 90%.

These results clearly indicate that previous approaches are limited in capturing the full complexity of electricity consumption data. Traditional models such as ARIMA are effective for linear patterns but fail to model nonlinear dependencies [49], while standalone deep learning models often capture only partial temporal characteristics [50]. Even hybrid two-model architectures improve performance but still lack the ability to simultaneously capture local patterns, sequential dependencies, and global temporal relationships [51]. The proposed CNN–LSTM–Transformer architecture addresses this gap by integrating convolutional feature extraction, recurrent temporal modeling, and attention-based global context learning within a unified framework. This combination enables the model to learn multi-scale temporal representations more effectively, resulting in significantly improved forecasting accuracy.

Despite its strong performance, the model still shows limitations when handling highly volatile consumption patterns, as indicated by higher RMSE values in certain zones and scenarios. This suggests that extreme fluctuations and heterogeneous consumption behaviors remain challenging, even for advanced hybrid architectures. Nevertheless, the findings of this study provide clear empirical evidence that integrating CNN, LSTM, and Transformer offers a substantial improvement over both traditional statistical models and existing deep learning approaches. This not only reinforces the effectiveness of hybrid architectures in electricity forecasting but also highlights the scientific contribution of this work in advancing multi-component deep learning models for complex time-series prediction tasks.

CONCLUSION

This study aimed to evaluate the effectiveness of a hybrid CNN–LSTM–Transformer architecture for short-term electricity consumption forecasting under multiple data configurations. The results demonstrate that the proposed model achieves consistently high predictive performance across all experimental scenarios, with MAE values ranging from 228 to 489 kWh, RMSE values between 385 and 663 kWh, and MAPE generally below 2%, confirming its capability in modeling complex electricity consumption patterns. Among the evaluated scenarios, the multivariate combined configuration (S4) produced the most accurate and stable results, with MAE values of 232–257 kWh, RMSE values of 385–433 kWh, and MAPE ranging from 0.82% to 1.43%, highlighting the importance of integrating both multivariate features and cross-zone information. In contrast, scenarios relying on single-zone or limited feature inputs showed relatively higher errors, indicating that model performance is sensitive to data richness and feature representation. The primary contribution of this study lies in the development and systematic evaluation of a unified hybrid architecture that simultaneously captures local temporal patterns (via CNN), sequential dependencies (via LSTM), and global contextual relationships (via attention mechanisms). Unlike prior studies that focus on single configurations, this research introduces a multi-scenario experimental framework (S1–S6) that provides a more comprehensive assessment of model robustness under varying data conditions. Despite these contributions, this study has several limitations, particularly reduced generalization performance in certain cross-zone configurations and the absence of extensive hyperparameter optimization, which may limit the model's optimal performance.

Future research should focus on improving model generalization and adaptability through more advanced techniques. Potential directions include the application of transfer learning to leverage knowledge across zones, domain adaptation methods to address distribution differences between regions, and automated hyperparameter optimization strategies to enhance model performance. Furthermore, incorporating additional contextual data, such as socioeconomic indicators or real-time grid information, may further improve forecasting accuracy in large-scale energy systems.

REFERENCES

[1] P. T. Pln, "Rencana usaha penyediaan tenaga listrik (ruptl) 2025 - 2034," pp. 2025–2034, 2025.

[2] A. Agaal, "Data-driven Insights for Informed Decision- Making : Applying LSTM Networks for Robust Electricity Forecasting in Libya," vol. 17, no. 3, pp. 65–89, 2025, doi: 10.5815/ijisa.2025.03.05.

[3] K. Chen, J. Jiang, F. Zheng, and K. Chen, "A novel data-driven approach for residential electricity consumption prediction based on ensemble learning," *Energy*, vol. 150, pp. 49–60, 2018, doi: 10.1016/j.energy.2018.02.028.

[4] H. Shen, F. Li, L. Shi, S. Sun, Z. Yang, and T. Yang, "Short-Term Power Load Forecasting Based on Reduction of Meteorological Data Dimensionality and Hybrid Deep Learning," *Dianli Jianshe/Electric Power Constr.*, vol. 45, no. 1, pp. 13–21, 2024, doi: 10.12204/j.issn.1000-7229.2024.01.002.

[5] M. Ala'raj, M. Radi, E. Alsisi, M. Majdalawieh, and M. Darwish, "Machine Learning-Based Electric Vehicle Charging Demand Forecasting: A Systematized Literature Review," *Energies*, vol. 18, no. 17, 2025, doi: 10.3390/en18174779.

[6] M. Ahmed, M. Elsayed, and K. Emara, "Enhancing Short-Term Electricity Consumption Forecasting with Deep Learning Models," *2025 Twelfth Int. Conf. Intell. Comput. Inf. Syst.*, pp. 454–461, 2025, doi: 10.1109/ICICIS66182.2025.11313214.

[7] A. Moradzadeh, H. Moayyed, K. Zare, and B. Mohammadi-ivatloo, "Short-term electricity demand forecasting via variational autoencoders and batch training-based bidirectional long short-term memory," *Sustain. Energy Technol. Assessments*, vol. 52, 2022, doi: 10.1016/j.seta.2022.102209.

[8] A. Wan, Q. Chang, K. AL-Bukhaiti, and J. He, "Short-term power load forecasting for combined heat and power using CNN-LSTM enhanced by attention mechanism," *Energy*, vol. 282, no. March, p. 128274, 2023, doi: 10.1016/j.energy.2023.128274.

[9] S. Fang, S. Shu, J. Pan, H. Huang, G. Wang, and K. Wang, "A long-term and short-term prediction method for power transformer top oil temperature based on comprehensive thermal factor and CNN-LSTM-attention," *Int. J. Electr. Power Energy Syst.*, vol. 172, no. August, p. 111170, 2025, doi: 10.1016/j.ijepes.2025.111170.

[10] N. Oliveira, N. Sousa, and I. Praça, "Deep Learning for Short-Term Instant Energy Consumption Forecasting in the Manufacturing Sector," in *Lecture Notes in Networks and Systems*, 2023, pp. 165–175. doi: 10.1007/978-3-031-20859-1_17.

[11] R. Raju, T. Thachankattil, and N. P. Ramlal, "Dual Deep Learning model for Electricity Price Forecasting: Bi-LSTM and GRU fusion," 2023, pp. 13–17. doi: 10.1145/3607947.3607951.

[12] D. M. Asad, M. Shahriar, and A. H. Chowdhury, "Load Forecasting for Dhaka City Using RNN, LSTM, and GRU Architectures with Meteorological and Temporal Data," 2025. doi: 10.1109/ECCE64574.2025.11013779.

[13] K. Gaur and S. Sharma, "CNN-Bi-LSTM based household energy consumption prediction," 2021, pp. 233–237. doi: 10.1109/ICSPC51351.2021.9451797.

[14] Y. K. Ahranjani, M. Beiraghi, and R. Ghanizadeh, "Short time load forecasting for Urmia city using the novel CNN-LTSM deep learning structure," *Electr. Eng.*, vol. 107, no. 1, pp. 1253–1264, 2025, doi: 10.1007/s00202-024-02361-4.

[15] F. Chen, C. Wu, J. Zhang, and G. Liu, “LSTM Based Short-Term Data Center Electrical Consumption Forecasting,” 2023, pp. 730–735. doi: 10.1145/3594739.3612915.

[16] Z. Zhao *et al.*, “Short-Term Load Forecasting Based on the Transformer Model,” *Inf.*, vol. 12, no. 12, 2021, doi: 10.3390/INFO12120516.

[17] S. Bouktif, A. Fiaz, A. Ouni, and M. A. Serhani, “Single and multi-sequence deep learning models for short and medium term electric load forecasting,” *Energies*, vol. 12, no. 1, 2019, doi: 10.3390/en12010149.

[18] G. M. Ramadan, A. S. Abosinnee, M. K. Al-Dulaimi, Z. Jaffer, A. alkhayyat, and Z. L. Naser, “Hybrid Deep Learning Enabled Load Prediction for Energy Storage Systems,” 2023, pp. 466–472. doi: 10.1109/IICETA57613.2023.10351206.

[19] E. M. Al-ali, Y. Hajji, Y. Said, M. Hleili, A. M. Alanzi, and A. H. Laatar, “Solar Energy Production Forecasting Based on a Hybrid CNN-LSTM-Transformer Model,” pp. 1–19, 2023.

[20] M. Ranjbar and M. Rahimzadeh, “Advancing Gasoline Consumption Forecasting : A Novel Hybrid Model Integrating Transformers , LSTM , and”.

[21] A. Binbusayyis and M. Sha, “Heliyon Energy consumption prediction using modified deep CNN-Bi LSTM with attention mechanism,” *Heliyon*, vol. 11, no. 1, p. e41507, 2025, doi: 10.1016/j.heliyon.2024.e41507.

[22] K. Boumais and F. Messaoudi, “Short-Term Load Forecasting Method for Renewable Energy Integration and Grid Stability Using CNN, LSTM, and Transformer Models,” *Stat. Optim. Inf. Comput.*, vol. 13, no. 4, pp. 1578–1594, 2025, doi: 10.19139/soic-2310-5070-2256.

[23] Y. Gong, Y. Wang, Y. Xie, X. Peng, Y. Peng, and W. Zhang, “Dynamic fusion LSTM-Transformer for prediction in energy harvesting from human motions,” *Energy*, vol. 327, no. April, p. 136192, 2025, doi: 10.1016/j.energy.2025.136192.

[24] J. Chen and X. Xue, “A novel deep learning-based convolutional neural network - long short-term memory model for predicting weekly significant wave heights,” *Ocean Eng.*, vol. 334, no. May, p. 121647, 2025, doi: 10.1016/j.oceaneng.2025.121647.

[25] B. Lou, J. Tang, L. Hu, and J. Ye, “Multi-source data-driven short-term remaining driving range prediction for electric vehicles : A hybrid CNN-transformer framework,” *Energy*, vol. 334, no. June, p. 137564, 2025, doi: 10.1016/j.energy.2025.137564.

[26] N. Sergeev and P. Matrenin, “Improving Accuracy of Machine Learning Based Short-Term Load Forecasting Models with Correlation Analysis and Feature Engineering,” in *International Conference of Young Specialists on Micro/Nanotechnologies and Electron Devices, EDM*, 2023, pp. 1000–1004. doi: 10.1109/EDM58354.2023.10225058.

[27] M. Wang, Y. Zheng, B. Wang, and Z. Deng, “Household Electricity Load Forecasting Based on Multitask Convolutional Neural Network with Profile Encoding,” *Math. Probl. Eng.*, vol. 2021, 2021, doi: 10.1155/2021/6661798.

[28] C. Yang, Z. Liu, and Q. Li, “Fusion of Attention Mechanisms in CNN-LSTM Models for Paid Digital Labor Prediction,” *ACM Int. Conf. Proceeding Ser.*, no. January, pp. 211–217, 2025, doi: 10.1145/3735014.3735901.

[29] D. Li, C. Xiao, X. Zeng, and Q. Shi, “Short-mid term electricity consumption prediction using non-intrusive attention-augmented deep learning model,” *Energy Reports*, vol. 8, pp. 10570–10581, 2022, doi: 10.1016/j.egyr.2022.08.195.

[30] T. Deepa, P. Vishnu, and P. Monika, “Assessing The Efficacy of Linear Regression and Long-Short Term Memory in Refining the Accuracy of Energy Consumption Forecasts for The Purpose of Optimized Energy Distribution,” *2nd Int. Conf. Emerg. Res. Comput. Sci. ICERCS 2024*, pp. 1–4, 2024, doi: 10.1109/ICERCS63125.2024.10894894.

[31] Z. Talib, L. Ali, A. Rahaim, and A. S. Aljuboury, “Modified CNN Learning Model for Tetouan City Energy Consumption Prediction,” vol. 3, pp. 465–479, 2025, doi: 10.58346/JISIS.2025.I3.032.

[32] M. Bensalah, “Empowering Smart Cities : SARIMA Forecasting of Power Consumption in Tetouan ’ s Urban Grid,” 2024.

[33] W. Yang, J. Shi, S. Li, Z. Song, Z. Zhang, and Z. Chen, “A combined deep learning load forecasting model of single household resident user considering multi-time scale electricity consumption behavior,” *Appl. Energy*, vol. 307, 2022, doi: 10.1016/j.apenergy.2021.118197.

[34] A. S. Attention, “Short-term load forecasting of CNN-LSTM integrating Markov state transition and self-attention mechanism,” *2025 5th Int. Conf. Comput. Sci. Blockchain*, pp. 93–96, 2025, doi: 10.1109/CCSB66722.2025.11154311.

[35] P. Chakraborty, S. Kalaivani, A. Ambika, K. Ramya, C. Engineering, and T. Nadu, “Forecasting Electricity Consumption Using ARIMA-LSTM Model,” vol. 17, no. 4, pp. 4–7, 2025, doi:

10.21272/jnep.17(4).04011.
[36] P. Chen and L. Chen, “Prediction method of intelligent building electricity consumption based on deep learning,” *Evol. Intell.*, vol. 16, no. 5, pp. 1637–1644, 2023, doi: 10.1007/s12065-023-00815-5.
[37] M. Qureshi, M. A. Ahmad, and S. U. Ur Rehman, “Deep learning-based forecasting of electricity consumption,” *Sci. Rep.*, vol. 14, no. 1, 2024, doi: 10.1038/s41598-024-56602-4.
[38] J. W. Chan and C. K. Yeo, “Electrical Power Consumption Forecasting with Transformers,” 2022, pp. 255–260. doi: 10.1109/EPEC56903.2022.10000228.
[39] T. M. Al-Jaafreh, A. Al-Odienat, and Y. A. Altaharwah, “The Solar Energy Forecasting Using LSTM Deep Learning Technique,” 2022. doi: 10.1109/ETCEA57049.2022.10009717.
[40] M. T. Ucar and A. Kaygusuz, “Short-Term Energy Consumption Forecasting Analysis Using Different Optimization and Activation Functions with Deep Learning Models,” *Appl. Sci.*, vol. 15, no. 12, 2025, doi: 10.3390/app15126839.
[41] K. Zhang, R. Zhao, S. Liu, Q. Song, and J. Zhang, “Deep Learning-Based Forecasting of Power Load and Management System,” 2023, pp. 753–755. doi: 10.1109/ISPCEM60569.2023.00142.
[42] F. Ouyang, J. Wang, and H. Zhou, “Electricity Consumption Forecasting Based on PC-CNN-BiLSTM Combined with Layered Transfer Learning Strategy,” 2022, pp. 628–632. doi: 10.1109/ICCEAI55464.2022.00133.
[43] H. Zang *et al.*, “Residential load forecasting based on LSTM fusing self-attention mechanism with pooling,” *Energy*, vol. 229, 2021, doi: 10.1016/j.energy.2021.120682.
[44] R. A. Nugroho, S. Mu’min, A. Saputro, A. Andhyka, A. S. Editya, and A. Mufliq, “Forecasting of Electricity Usage of Indonesian Consumer Using Transformer Time Series,” 2023, pp. 707–710. doi: 10.1109/ICAMIMIA60881.2023.10427799.
[45] Z. Lin, “A hybrid CNN-LSTM-Attention model for electric energy consumption forecasting,” 2023, pp. 167–173. doi: 10.1145/3588155.3588182.
[46] B. Zhang, H. Sun, and Z. Yu, “Evaluation and Research of Power Forecasting Based on LSTM Time Series,” 2022, pp. 176–181. doi: 10.1109/AIAM57466.2022.00040.
[47] K. Lin, Y. Zhao, and J. hong Kuo, “Data-driven models applying in household hazardous waste: Amount prediction and classification in Shanghai,” *Ecotoxicol. Environ. Saf.*, vol. 263, no. July, p. 115249, 2023, doi: 10.1016/j.ecoenv.2023.115249.
[48] Y.-J. Lee and H.-J. Choi, “Forecasting building electricity power consumption using deep learning approach,” 2020, pp. 542–544. doi: 10.1109/BigComp48618.2020.000-8.
[49] F. V Atabay, R. M. Pagkalinawan, S. D. Pajarillo, A. R. Villanueva, and J. V Taylar, “Multivariate Time Series Forecasting using ARIMAX, SARIMAX, and RNN-based Deep Learning Models on Electricity Consumption,” 2022. doi: 10.1109/IISEC56263.2022.9998301.
[50] H. Sun, Y. Ren, S. Wang, B. Zhao, and R. Yin, “Smart grid power load type forecasting: research on optimization methods of deep learning models,” *Front. Energy Res.*, vol. 11, 2023, doi: 10.3389/fenrg.2023.1321459.
[51] A. Saranj and M. Zolfaghari, “The electricity consumption forecast: Adopting a hybrid approach by deep learning and ARIMAX-GARCH models,” *Energy Reports*, vol. 8, pp. 7657–7679, 2022, doi: 10.1016/j.egyr.2022.06.007.