Int J Comput Intell Syst https://doi.org/10.1007/s44196-026-01336-6

Article in Press

A Hybrid Temporal-spectral Load Forecasting Model with Static Context Fusion for Smart Cities

Received: 14 December 2025

Accepted: 17 April 2026

Published online: 22 May 2026

Cite this article as: Rehman H.M.R.U., Younas R., Changhyun P. *et al.* A Hybrid Temporal-spectral Load Forecasting Model with Static Context Fusion for Smart Cities. *Int J Comput Intell Syst* (2026). https://doi.org/10.1007/s44196-026-01336-6

Hafiz Muhammad Raza Ur Rehman, Rabbiya Younas, Park Changhyun, Urfa Gul, Roberto Marcelo Alvarez, Yini Miro & Imran Ashraf

A Hybrid Temporal-Spectral Load Forecasting Model with Static Context Fusion for Smart Cities

Hafiz Muhammad Raza Ur Rehman[1], Rabbiya Younas[1], Park Changhyun[1], Urfa Gul[1*], Roberto Marcelo Alvarez[2,3,4], Yini Miro[2,5,6,7], Imran Ashraf[1*]

[1]School of Computer Science and Engineering, Yeungnam University, Gyeongsan, 38541, Republic of Korea.
[2]Universidad Europea del Atlántico., Isabel Torres 21, Santander, 39011, Spain.
[3]Universidad Internacional Iberoamericana Arecibo, Puerto Rico, 00613, USA.
[4]Universidade Internacional do Cuanza, Cuito, Bie,, Angola.
[5]Universidad Internacional Iberoamericana, Campeche, 24560, Mexico.
[6]Fundacion Universitaria Internacional de Colombia, Bogota, Colombia.
[7]Universidad de La Romana, La Romana, Republica Dominicana.

*Corresponding author(s). E-mail(s): urfagul@yu.ac.kr; imranashraf@yu.ac.kr;
Contributing authors: mrazaurrehman@ynu.ac.kr; younis.rabbiya@gmail.com; park@yu.ac.kr; roberto.alvarez@uneatlantico.es ; yini.miro@uneatlantico.es;

Abstract

Accurate short-term electricity load forecasting is essential for reliable and efficient smart city energy management, particularly in environments characterized by high-dimensional, heterogeneous, and noisy multivariate signals. However, existing forecasting models often struggle to simultaneously capture nonlinear temporal dependencies, multi-scale periodicity, and static contextual influences within a unified framework. To address this challenge, this study proposes a hybrid deep learning architecture that integrates Bidirectional Long Short-Term Memory (BiLSTM) for temporal modeling, an additive attention mechanism for adaptive time-step weighting, Fast Fourier Transform (FFT)-based frequency residual learning for periodicity extraction, and embedding-based static feature fusion for contextual representation. The model is evaluated on the ISO-NE Smart City Energy Dataset for next-hour electricity load forecasting using a two-week input window (336 hours). Experimental results demonstrate that the proposed hybrid framework significantly improves predictive accuracy, achieving an RMSE of 25.51 kW and an $\mathbf{R^2}$ of 0.9905, outperforming recurrent, convolutional, and transformer-based baselines under identical evaluation settings. Ablation analysis confirms that temporal attention and frequency-domain residual modeling contribute substantially to performance gains. These findings indicate that joint temporal–spectral modeling combined with static contextual fusion provides a robust and effective solution for complex smart-city electricity forecasting tasks.

Keywords: Smart city energy forecasting; electricity load prediction; temporal attention; Fourier transform; hybrid deep learning; frequency-domain learning; static feature fusion; renewable-aware forecasting; urban energy analytics

1 Introduction

Intelligent energy-management infrastructures have become a fundamental requirement of modern smart cities. Smart cities rely heavily on services like smart meters, Internet of Things (IoT) sensors, renewable energy plants, electric mobility systems, etc. Such systems are interconnected and generate high-frequency multivariate data streams capturing consumption patterns, weather conditions, human mobility, building occupancy, and renewable energy fluctuations. Accurate electricity load forecasting plays a critical role in ensuring grid reliability, operational efficiency, and renewable integration within such complex environments. However, the increasing heterogeneity and dynamic nature of these data sources make forecasting both critically important and technically challenging. In addition, features like improved operational costs, renewable integration, demand-response programs, and managing overload and failure in urban power networks have become inevitable [1–3].

Traditional statistical forecasting methods such as autoregressive integrated moving average (ARIMA), seasonal ARIMA, and vector autoregression (VAR) are no longer adequate for capturing the nonlinear, dynamic, and high-dimensional nature of modern electricity demand. The growing integration of distributed renewable energy sources, including solar and wind, adds further uncertainty, which limits the reliability of purely time series-based models. Recent reviews show that deep learning methods can model complex, nonlinear relationships in high-dimensional smart-city datasets more effectively than classical approaches [4, 5].

Recurrent neural networks (RNNs), especially Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) models, became dominant techniques for load forecasting because of their ability to capture temporal dependencies [6, 7]. Despite these advantages, RNNs face challenges when dealing with long input sequences, multiscale periodicity, and high-frequency fluctuations. Moreover, smart-city electricity load is influenced by both short-term dynamics (e.g., hourly behavior, occupancy changes) and long-term seasonal structures (e.g., daily/weekly cycles), which have been difficult to represent in pure time-domain models.

Transformer-based architectures have recently achieved state-of-the-art performance in many time-series tasks. Models such as Informer, Autoformer, and FEDformer introduce sparse self-attention, seasonal-trend decomposition, and frequency-domain modeling to enhance long-range forecasting efficiency [8–11]. While several transformer-based frameworks incorporate mechanisms for handling static covariates (e.g., via dedicated encoders or cross-attention modules), their performance in highly noisy and heterogeneous smart-city energy datasets remains sensitive to data characteristics and hyperparameter configuration. Moreover, transformer architectures are typically computationally intensive, which may limit their practicality for large-scale real-time deployment in smart-grid environments.

Smart-city electricity datasets present additional challenges not fully addressed by existing models:

- **Heterogeneous multi-source variables:** Real-world electricity load is jointly influenced by weather conditions, mobility, occupancy, energy infrastructure, and renewable generation. Purely sequential models fail to capture such mixed-domain interactions [2].
- **High-dimensional categorical attributes:** Region ID, weather categories, building type, and load sectorhave been typically encoded using one-hot vectors, causing extremely high-dimensional inputs that degrade learning.
- **Complex periodicity and abrupt fluctuations:** Load profiles contain strong daily/weekly cycles, combined with sudden irregular changes due to EV charging, renewable variability, or demand-response actions [4].
- **Incomplete integration of temporal, spectral, and contextual information:** Although hybrid forecasting models have become increasingly common, many existing frameworks do not tightly integrate temporal modeling, frequency-domain learning, and embedding-based static contextual representation within a single end-to-end architecture [6, 12].

These challenges highlight the need for forecasting models capable of jointly addressing multi-source heterogeneity, complex periodicity, and contextual feature fusion. Although prior hybrid architectures have improved specific aspects of forecasting performance, the tight integration of temporal modeling, frequency-domain learning, and embedding-based contextual representation within a unified end-to-end trainable architecture remains insufficiently explored for smart-city electricity forecasting.

To address these limitations, recent literature emphasises hybrid neural architectures that leverage complementary strengths of different modeling paradigms. CNN–LSTM hybrids have been shown to enhance short-term feature extraction [6]. Attention mechanisms further improve interpretability by assigning dynamic importance weights to time steps. Frequency-domain deep models (Fourier or wavelet-based) capture long-term periodicity that cannot be detected in raw temporal signals [4]. Embedding-based representations for categorical features have also proven superior to one-hot encodings in reducing dimensionality and capturing semantic relationships [13].

Motivated by these observations, this study proposes a unified hybrid model for smart-city electricity forecasting that integrates:

1. Bidirectional LSTM (BiLSTM) for contextual temporal modeling,
2. temporal attention for adaptive time-step weighting,
3. FFT-based residual extraction for long-term periodicity modeling, and
4. embedding-based static feature encoding for high-dimensional categorical attributes.

The proposed framework jointly captures multi-scale temporal behavior, long-range seasonal patterns, and contextual non-temporal dependencies within a single forecasting architecture.

1.1 Research Gaps

Although numerous deep-learning approaches have been proposed for electricity load forecasting, many prior studies treat temporal modeling, frequency decomposition, and contextual feature representation as partially independent design choices rather than tightly integrated components of a unified forecasting framework. Furthermore, the literature lacks sufficiently controlled benchmarking studies comparing recurrent, convolutional, attention-based, and transformer-based forecasting architectures under identical experimental settings on the same smart-city energy dataset.

1.2 Problem Statement

Smart-city electricity load forecasting requires models capable of capturing nonlinear temporal patterns, multi-scale periodicity, renewable-driven variability, and heterogeneous contextual features. However, many existing forecasting architectures address these aspects only partially or through loosely coupled modules, which may limit representational synergy and predictive performance. Therefore, this work aims to develop a unified hybrid deep-learning architecture that jointly incorporates temporal dynamics, attention-based temporal weighting, frequency-domain periodicity modeling, and static feature embeddings for robust electricity load forecasting in smart urban environments.

1.3 Contributions

The key contributions of this study are summarized as follows:

- An end-to-end trainable hybrid architecture is proposed for smart-city electricity forecasting. The framework integrates BiLSTM, temporal attention, FFT-based residual learning, and embedding-based static feature fusion within a unified model to jointly capture temporal, spectral, and contextual patterns.
- An embedding-enhanced static feature fusion strategy is introduced to efficiently encode high-dimensional categorical contextual variables into dense representations, improving the interaction between static metadata and dynamic temporal-spectral features.

- Extensive experiments on a real-world smart-city energy dataset validate the complementary roles of temporal, spectral, and contextual components, demonstrating that their joint optimization yields substantial forecasting gains over representative baseline models.

The rest of the article has been organized as follows. Section 2 contain the comprehensive analysis of previous work. Section 3 contained the detail about dataset used in this study along with the clear information about the proposed scheme, while Section 4 depicts the performance comparison of proposed model with other models. It also contain the discussion about the results of the proposed scheme. Finally, the conclusion and future work have been given in Section 5.

2 Related Work

Recent advances in deep learning have significantly accelerated progress in electricity load forecasting, particularly within smart-city ecosystems characterized by heterogeneous, high-frequency data streams from IoT sensors, smart meters, renewable generation units, and socio-environmental indicators. Traditional approaches including ARIMA, SARIMA, SVR, and random forests have gradually been replaced by neural architectures capable of modeling nonlinear multivariate dependencies and long-range temporal variations [14]. Deep recurrent networks remain among the most widely used models in electricity forecasting. Zhang et al. [15] demonstrated that LSTM networks outperform classical statistical models for residential and utility-level forecasting by effectively capturing nonlinear temporal behavior. Similarly, Li et al. [16] introduced a hybrid LSTM framework enhanced with exogenous weather and renewable indicators, showing improved performance under volatile generation conditions.

While recurrent architectures are effective in modeling sequential dependencies, they may face challenges in capturing multi-scale periodic structures and long-term seasonal patterns present in complex smart-city environments.

Hybrid CNN–RNN architectures have emerged to address the shortcomings of pure LSTM or GRU models. He et al. [17] investigated CNN–LSTM–Attention architectures that enhance local pattern extraction and dynamic weighting across timesteps. More recently, Shawon et al. [18] proposed a CNN–LSTM hybrid for urban load forecasting in the Chattogram metropolitan region and demonstrated improved short-term accuracy. Although these hybrid models strengthen local feature learning and short-term volatility modeling, they primarily operate in the time domain and may not explicitly model spectral periodicity or systematically incorporate high-dimensional static contextual information such as region metadata and consumer categories.

Transformer-based methods have gained traction due to their ability to model long-range dependencies with self-attention. Informer [8] introduced sparse attention to enable efficient long-sequence processing, while Autoformer [9] incorporated auto-correlation-based decomposition modules for seasonal–trend modeling. FEDformer [10] further improved forecasting by applying Fourier-based decomposition in the frequency domain. These architectures demonstrate strong performance in structured long-horizon forecasting tasks. However, in highly noisy, heterogeneous, and multi-source smart-city datasets, performance may depend strongly on architectural configuration, hyperparameter tuning, and the manner in which contextual variables are incorporated. Additionally, transformer-based models often involve substantial parameter complexity, which may pose practical considerations for large-scale real-time deployment in smart grids.

Frequency-domain approaches have recently gained attention for capturing long-term periodic behavior. Feng and Zhou [19] demonstrated that FFT-based spectral components substantially improve robustness and interpretability in energy-demand modeling. Similarly, recent spectral–hybrid models [20] combined variational mode decomposition (VMD) with xLSTM and Informer, showing that frequency decomposition enhances long-range forecasting. Meanwhile, deep generative and decomposition-based models such as EMD–ISSA–LSTM [21] and GAN-driven architectures have improved performance under noisy and renewable-rich conditions. Nevertheless, many decomposition-based pipelines treat frequency extraction as a separate preprocessing stage rather than integrating it seamlessly with bidirectional sequence learning and adaptive temporal attention within a unified architecture.

A parallel direction involves embedding high-dimensional static metadata such as region ID, building type, and load sector using dense embedding layers instead of one-hot encodings. Liu et al. [22] demonstrated that embedding-based representations can improve forecasting performance across diverse geographic regions by learning compact and semantically meaningful contextual encodings. More recent deep-learning frameworks for smart grids [23] emphasize integrating contextual features with temporal deep models. However, in several architectures, static metadata are appended at later stages without strong interaction with temporal or spectral representations, potentially limiting joint representation learning.

Taken together, the literature reveals several recurring challenges in smart-city electricity forecasting:

i. Difficulty in jointly modeling nonlinear temporal dependencies and multi-scale periodicity.
ii. Sensitivity of attention-based architectures to noise and heterogeneous multi-source inputs in real-world urban datasets.
iii. Limited integration between frequency-domain decomposition and bidirectional temporal encoding.
iv. Insufficient interaction between static contextual embeddings and dynamic temporal–spectral representations.
v. Lack of unified end-to-end architectures that simultaneously address multiscale periodicity, contextual heterogeneity, and bidirectional sequence learning within a single trainable framework.

Table 1 provides a comparison of recent works with the proposed approach to emphasize how existing methods address specific forecasting challenges and where integration across temporal, spectral, and contextual dimensions remains limited.

Motivated by these observations, the proposed model fuses BiLSTM-based temporal modeling, multi-head temporal attention, FFT-driven spectral residual extraction, and embedding-enhanced static feature fusion within a unified architecture. This integrated design aims to robustly capture short-term variations, long-term periodicity, and contextual heterogeneity, three defining characteristics of real-world smart-city electricity load patterns.

Table 1: Comparison of recent related works (2021–2025) and the proposed hybrid model.

Ref.	**Approach**	**Methodology / Model Characteristics**	**Limitations Compared to Proposed Model**
[8]	Informer	Sparse-attention Transformer for long-sequence forecasting.	No explicit FFT residual modeling; limited static contextual feature integration; performance may vary under noisy heterogeneous conditions.
[9]	Autoformer	Decomposition-based Transformer using auto-correlation.	Does not leverage recurrent temporal encoding; computationally intensive; performance may vary under highly heterogeneous urban data.
[10]	FEDformer	Fourier-based frequency decomposition with attention.	Lacks bidirectional sequence encoding; static contextual metadata not explicitly modeled; sensitivity to noisy consumption patterns may occur.
[18]	Hybrid CNN-LSTM	CNN + LSTM hybrid for urban electricity forecasting.	Primarily time-domain focused; does not explicitly model frequency-domain periodicity or static contextual embeddings.
[17]	CNN-LSTM-Attention	Hybrid CNN + LSTM + attention for short-term load volatility.	Limited explicit frequency-domain modeling for long-term periodic structures.
[19]	FFT Models	FFT-based spectral learning for periodic components.	Does not jointly incorporate bidirectional temporal encoding, adaptive temporal attention, and static contextual fusion.
[20]	VMD-KPCA-xLSTM-Informer	Hybrid spectral + LSTM + Informer architecture.	Relies on multi-stage decomposition pipeline; limited unified end-to-end feature fusion.
[22]	Embedding City Forecasting	Entity embeddings for multi-region electricity prediction.	Focuses on contextual embeddings without explicit spectral modeling or temporal attention integration.
Proposed	Hybrid approach	BiLSTM + Attention + FFT residuals + embedding-based static feature fusion.	-

3 Materials and Methods

The overall architectural workflow of the proposed framework in a real-world smart-city environment is illustrated in Figure 1. The system consists of a data aggregation layer that collects multi-source urban signals, followed by an AI-based prediction module implementing the proposed hybrid BiLSTM–Attention–FFT architecture for next-hour electricity load forecasting.

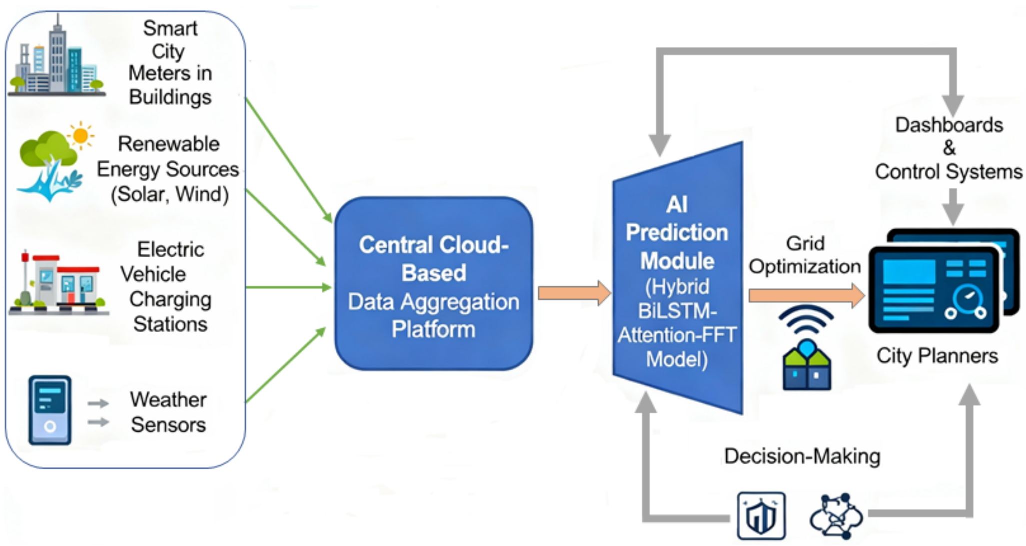

Fig. 1: Architecture of the proposed approach.

3.1 Dataset Description

The experiments in this study have been conducted using the ISO-NE Smart City Energy Dataset, a publicly available large-scale electricity and smart-city analytics dataset hosted on Kaggle.[1]. The dataset contains high-resolution operational, environmental, and socio-technical indicators collected from multiple interconnected smart-city infrastructures within the ISO New England (ISO-NE) region. It is specifically designed to support research in short- and long-term electricity load forecasting, renewable integration, grid stability, and urban analytics.

3.1.1 Structure and Temporal Resolution

The dataset consists of 72,624 time-stamped samples recorded at an hourly temporal resolution. Each observation includes load-related, meteorological, renewable energy, infrastructural, and human-activity variables. Modeling complex electricity consumption patterns in urban environments requires capturing short-term fluctuations, daily and weekly seasonality, and long-range periodic behavior.

3.1.2 Feature Composition

Each sample includes more than 60 numerical features and several categorical descriptors. Broadly, the dataset comprises the following variable groups:

[1] https://www.kaggle.com/datasets/datasetengineer/isone-smart-city-energy-dataset

- **Electricity and Demand Variables:** historical load (kW), peak-load indicators, transformer loading, power factor, voltage and current levels, demand-response events, and substation identifiers.
- **Renewable Energy Indicators:** solar PV output, wind power generation, inverter efficiency, battery state-of-charge (SOC), and renewable forecast error.
- **Meteorological Attributes:** temperature, humidity, rainfall, solar irradiance, cloud cover, snowfall, visibility, dew point, atmospheric pressure, wind speed, and wind direction.
- **Urban Activity and Smart-City Signals:** EV charging load and session count, public-transit operational load, traffic congestion index, human mobility score, and construction or maintenance flags.
- **Spatio-Temporal Metadata:** timestamp, region/zone identifier, latitude, longitude, altitude, and area type.

The target variable for prediction is the 'Electricity Load', representing the instantaneous aggregated grid demand (in kW) within the monitored region.

3.1.3 Relevance to Deep Learning Forecasting

The dataset exhibits several characteristics that make it ideal for advanced time-series forecasting models:

i. **Long Sequences:** Hourly data across multiple years enables long-horizon learning, essential for attention and FFT-based models.
ii. **Multiple Seasonalities:** Daily, weekly, and weather-driven seasonal patterns require models capable of multi-scale temporal representation.
iii. **High Feature Heterogeneity:** The coexistence of numerical and categorical features motivates hybrid deep architectures combining sequence encoders and learned embeddings.
iv. **Signal Noise and Variability:** Real smart-city signals contain operational disturbances, human-behavior fluctuations, and weather uncertainty conditions where BiLSTM and frequency-domain decomposition have been particularly effective.

3.1.4 Preprocessing

Figure 2 illustrates the preprocessing pipeline. The following steps were performed:

- Missing or malformed entries have been removed.
- Categorical variables (e.g., load sector type, weather condition, region ID) have been label-encoded and later transformed via learned embeddings.
- Numerical features have been standardized using Min–Max scaling.
- For sequence models, a fixed input window of 336 hours (two weeks) is used, forming a 3-D tensor of shape $(N, 336, F)$.
- Static features (region/zone type, building category, etc.) have been preserved separately for integration using the static-embedding module of the proposed architecture.

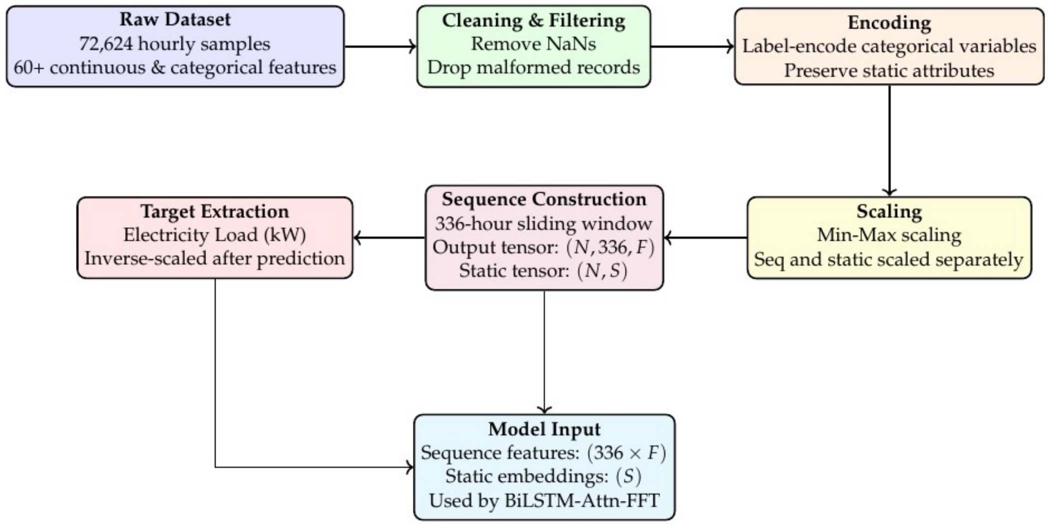

Fig. 2: End-to-end preprocessing pipeline for the ISO-NE Smart City Energy Dataset: cleaning, encoding, scaling, temporal windowing (336 hours), static feature extraction, and preparation for the BiLSTM-Attention-FFT forecasting model.

This rich data representation enables the proposed BiLSTM-Attention-FFT model to jointly exploit: (i) short-term local dynamics, (ii) long-range seasonal periodicity, and (iii) contextual spatial or categorical information leading to a highly expressive and robust load-forecasting framework suitable for real-world smart-city environments.

3.2 Proposed Hybrid Architecture

The objective of this work is to develop a robust deep learning model capable of accurately forecasting hourly electricity load in smart city environments where consumption patterns have been non-linear, periodic, and influenced by diverse exogenous factors. To address these challenges, this study proposes a hybrid architecture, as shown in Figure 3.

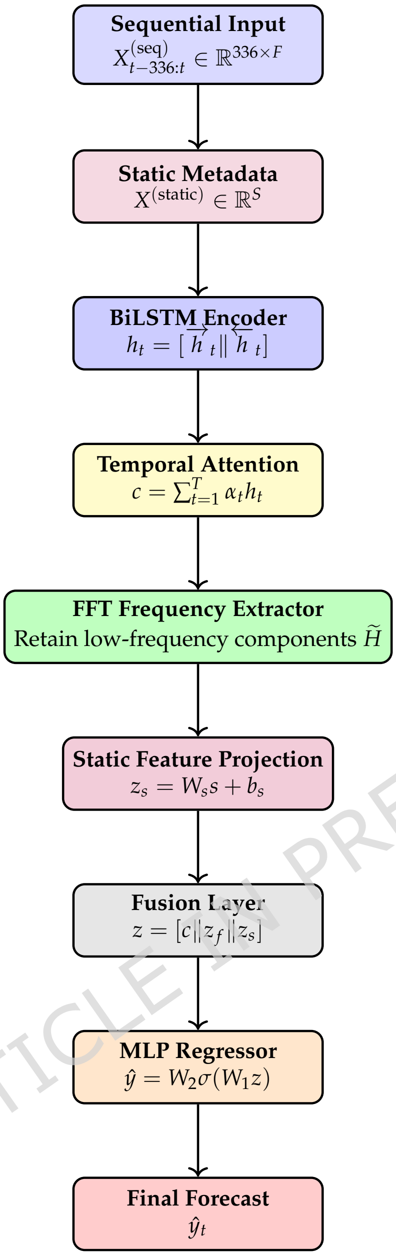

Fig. 3: Proposed hybrid BiLSTM–Attention–FFT–Static fusion model.

It integrates (i) a BiLSTM encoder for capturing temporal dependencies, (ii) an additive temporal attention mechanism to dynamically emphasize important time steps, (iii) a Fourier-based frequency extraction module to explicitly model periodicity, and (iv) a static feature fusion block to incorporate non-temporal categorical characteristics such as region identity, area type, and weather condition codes. The overall design yields a unified sequence-to-one forecasting model denoted as:

$$\widehat{y}_t = \mathcal{F}\big(X^{(\mathrm{seq})}_{t-336:t},\ X^{(\mathrm{static})}\big), \tag{1}$$

where $X^{(\mathrm{seq})}_{t-336:t} \in \mathbb{R}^{336\times F}$ contains 336 hours of historical features and $X^{(\mathrm{static})} \in \mathbb{R}^{S}$ represents static metadata.

Recent studies emphasize the importance of hybrid temporal–frequency models for complex energy forecasting [24–26], while attention mechanisms and multi-source fusion have shown strong improvements in recent load-prediction literature [27, 28]. Motivated by these findings, we develop an end-to-end trainable framework formalized below.

3.2.1 BiLSTM Temporal Encoding

Given the sequential input

$$X = \{x_1, x_2, \ldots, x_T\}, \qquad x_t \in \mathbb{R}^F, \tag{2}$$

A BiLSTM processes the sequence in both forward and backward directions:

$$\overrightarrow{h}_t = \mathrm{LSTM}_f(x_t, \overrightarrow{h}_{t-1}), \qquad \overleftarrow{h}_t = \mathrm{LSTM}_b(x_t, \overleftarrow{h}_{t+1}), \tag{3}$$

and the concatenated representation is

$$h_t = [\overrightarrow{h}_t \,\|\, \overleftarrow{h}_t] \in \mathbb{R}^{2H}. \tag{4}$$

This encoder captures long-term and short-term dependencies as well as bidirectional temporal correlations, which have been essential for electricity load influenced by cyclic and abrupt consumption changes. Empirical studies since 2021 confirm the superiority of BiLSTM architectures for energy and sensor forecasting [30, 32].

Not all past time steps contribute equally to predicting future load; peak consumption hours, sudden weather changes, or EV-charging events have disproportionately strong influence. Hence, we use an additive attention mechanism [11, 29, 31], defined as:

$$e_t = v^\top \tanh(W h_t), \tag{5}$$

$$\alpha_t = \frac{\exp(e_t)}{\sum_{\tau=1}^{T} \exp(e_\tau)}, \tag{6}$$

$$c = \sum_{t=1}^{T} \alpha_t h_t, \tag{7}$$

where α_t denotes the normalized weight over each hour in the look-back window and $c \in \mathbb{R}^{2H}$ is the context vector.

This mechanism adaptively highlights informative temporal regions such as sudden demand spikes or renewable fluctuation periods consistent with recent attention-based load forecasting models [27, 33].

3.2.2 FFT-Based Frequency Extraction

To explicitly capture periodic structures:

$$X_f = \mathrm{rFFT}(H), \quad H = [h_1, \ldots, h_T], \tag{8}$$

$$\widetilde{X}_f[:, 1:K] = X_f[:, 1:K], \qquad \widetilde{X}_f[:, K+1:] = 0, \tag{9}$$

$$\widetilde{H} = \mathrm{irFFT}(\widetilde{X}_f), \tag{10}$$

$$z_f = \mathrm{mean}_t(\widetilde{H}_t). \tag{11}$$

This procedure isolates stable periodic trends while filtering noise, aligning with 2022-2024 deep frequency-domain forecasting approaches [34, 35].

3.2.3 Static Feature Embedding and Fusion

Static categorical attributes such as region ID, grid type, and area class do not vary hourly while strongly modulate consumption behavior. Let:

$$s = X^{(\mathrm{static})} \in \mathbb{R}^S. \tag{12}$$

A linear projection maps these into the same latent space as the sequence features:

$$z_s = W_s s + b_s, \qquad z_s \in \mathbb{R}^{2H}. \tag{13}$$

This enables seamless fusion with temporal and frequency representations, consistent with modern multi-source fusion frameworks [36].

The three complementary representations have been concatenated:

$$z = [c \,\|\, z_f \,\|\, z_s], \tag{14}$$

and passed through a two-layer feed-forward network:

$$\widehat{y} = W_2\,\sigma(W_1 z + b_1) + b_2, \tag{15}$$

where $\sigma(\cdot)$ is ReLU.

The model is trained to minimize the mean squared error (MSE):

$$\mathcal{L} = \frac{1}{N}\sum_{i=1}^{N}(\widehat{y}_i - y_i)^2 \tag{16}$$

Adam optimizer with early stopping ensures stable convergence.

The final fused representation combines temporal context, frequency-domain information, and static contextual embeddings within a unified end-to-end trainable forecasting framework. Algorithm 1 summarizes the complete inference and training procedure of the proposed hybrid BiLSTM-Attention-FFT-Static Fusion model.

Algorithm 1 Hybrid BiLSTM-Attention-FFT-Static fusion forecasting model.

Require: Sequence window $X^{(\text{seq})}_{t-336:t} \in \mathbb{R}^{T\times F}$, Static features $X^{(\text{static})} \in \mathbb{R}^{S}$, Model parameters Θ.
Ensure: Predicted electricity load $\hat{y}_t$

(1) BiLSTM Temporal Encoding
for $t = 1$ to T **do**
    $\overrightarrow{h}_t \leftarrow \text{LSTM}_f(x_t, \overrightarrow{h}_{t-1})$
    $\overleftarrow{h}_t \leftarrow \text{LSTM}_b(x_t, \overleftarrow{h}_{t+1})$
    $h_t \leftarrow [\overrightarrow{h}_t \,\|\, \overleftarrow{h}_t]$
end for

(2) Additive Temporal Attention
for $t = 1$ to T **do**
    $e_t \leftarrow v^{\top} \tanh(W h_t)$
end for
$\alpha_t \leftarrow \exp(e_t) / \sum_{\tau=1}^{T} \exp(e_\tau)$
$c \leftarrow \sum_{t=1}^{T} \alpha_t h_t$

(3) FFT-based Frequency Extraction
$H \leftarrow [h_1, h_2, \ldots, h_T]$
$X_f \leftarrow \text{rFFT}(H)$
Retain first K frequencies: $\widetilde{X}_f[:, 1:K] \leftarrow X_f[:, 1:K]$
Zero out high frequencies: $\widetilde{X}_f[:, K+1:] \leftarrow 0$
$\widetilde{H} \leftarrow \text{irFFT}(\widetilde{X}_f)$
$z_f \leftarrow \text{mean}_t(\widetilde{H}_t)$

(4) Static Feature Projection
$z_s \leftarrow W_s X^{(\text{static})} + b_s$

(5) Fusion and Regression
$z \leftarrow [c \,\|\, z_f \,\|\, z_s]$
$h \leftarrow \sigma(W_1 z + b_1)$
$\hat{y}_t \leftarrow W_2 h + b_2$

(6) Optimization
$\mathcal{L} = \frac{1}{N} \sum_{i=1}^{N} (\hat{y}_i - y_i)^2$
Update Θ using Adam optimizer
Return $\hat{y}_t$

3.3 Hyperparameter Configuration

All models were trained under identical experimental conditions to ensure fair comparison. A sliding window of 336 hours (two weeks) was used as input sequence length for all architectures. Recurrent and convolutional models used a hidden dimension of 128, while transformer-based models followed standard configurations with $d_{model} = 256$

and 4 attention heads. Dropout was fixed at 0.3 to reduce overfitting across all experiments. The Adam optimizer was employed for training. Learning rate was set to 1×10^{-3} for recurrent-based models and 1×10^{-4} for transformer-based architectures, as recommended in recent time-series literature. Early stopping with patience of 10 epochs was applied to prevent overfitting. All models were trained for a maximum of 40 epochs using a batch size of 128. Table 2 shows the configuration of all models concerning hyperparameters.

Table 2: Hyperparameter configuration for all baseline and proposed models.

Model	Seq Len	Hidden Dim	Dropout	Batch	LR	Optimizer
LSTM	336	128	0.3	128	1e-3	Adam
BiLSTM	336	128×2	0.3	128	1e-3	Adam
GRU	336	128	0.3	128	1e-3	Adam
Transformer	336	256	0.3	128	1e-4	Adam
Autoformer	336	256	0.3	128	1e-4	Adam
FEDformer	336	256	0.3	128	1e-4	Adam
TFT	336	256	0.3	128	1e-4	Adam
Pro_BiLSTM_Attn	336	128×2	0.3	128	1e-3	Adam
Proposed_Full	336	128×2	0.3	128	1e-3	Adam

4 Results and Comparative Evaluation

A comprehensive evaluation of the proposed model is carried out and results are presented here in comparison to state-of-the-art models. The comparative analysis is done using several evaluation metrics such as mean squared error (MSE), RMSE, mean absolute error (MAE), mean absolute percentage error (MAPE), symmetric MAPE (SMAPE), R^2, and Nash-Sutcliffe efficiency (NSE). All experiments involve the same dataset and identical training strategy to ensure fair comparison.

4.1 Evaluation Metrics Overview

These evaluation matrices have been adopted for performance comparison:

- **Mean Squared Error:** MSE measures squared deviations between predictions and original values. It penalizes with respect to forecasting mistakes and has been widely used for grid operators where extreme peaks carry financial and operational risks.
- **Root Mean Squared Error:** RMSE is a more interpretable version of MSE and the same unit of kW is use. Lower RMSE refers to a more stable energy planning and reduced uncertainty.
- **Mean Absolute Error:** MAE is a linear measure of average forecasting deviations. Essential for evaluating day-to-day operational accuracy.
- **Mean Absolute Percentage Error:** Evaluates the relative forecast accuracy. Particularly relevant when utilities require proportional error understanding for dynamic tariffs and demand response.

- **Symmetric MAPE:** Addresses MAPE's asymmetry and stabilizes low-load periods. Critical for renewable-based grids where nighttime loads drop significantly.
- R^2 **(Coefficient of Determination):** Measures how much of the variance has been explained by the model. Higher means stronger predictive structure learning.
- **Nash–Sutcliffe Efficiency:** Extremely important for hydrological and energy systems forecasting. A value close to 1 indicates near-perfect predictive skill.

Together, these metrics ensure a holistic and unbiased evaluation of each model from multiple perspectives absolute accuracy, relative accuracy, variance explanation, and stability under extreme variations.

4.2 Comparison with Baseline and SOTA Models

To assess model efficiency and architectural complexity, Table 3 reports a parameter-level comparison between baseline models and the proposed architectures.

Table 3: Architectural and parameter-level comparison between baseline models and the proposed hybrid model.

Model	#Layers	Attention	FFT / Spectral	Static Embed	Params (M)	Cost
LSTM	2	No	No	No	0.52	Low
BiLSTM	2	No	No	No	0.94	Low
GRU	2	No	No	No	0.48	Low
TCN	6	No	No	No	0.81	Medium
Transformer	4 Enc + 4 Dec	Self-Attn	No	No	7.8	Very High
Autoformer	4 Enc + 4 Dec	Auto-Corr	Trend/ Season	No	8.2	Very High
FEDformer	4 Enc + 4 Dec	Self-Attn	Fourier	No	8.5	Very High
TFT	4	Gated Attn	No	Yes	6.9	High
Prop(BiLSTM-Attn)	2	Temporal	No	Yes	1.35	Low-Medium
Proposed Full	2	Temporal	FFT	Yes	1.48	Low-Medium

Table 3 compares the architectural complexity and trainable parameter counts of baseline models and the proposed architectures. Transformer-based models contain more than 7M parameters due to stacked self-attention layers and large feed-forward blocks, whereas classical recurrent models remain below 1M parameters. The proposed full hybrid model contains approximately 1.48M parameters, which is only marginally higher than the BiLSTM baseline, despite integrating temporal attention, FFT-based spectral modeling, and static feature embeddings. This demonstrates that the substantial accuracy gains are achieved without introducing excessive model size or computational burden. Compared to Transformer, Autoformer, FEDformer, and TFT, the proposed model uses substantially fewer trainable parameters while achieving lower forecasting error under the evaluated settings. These results suggest that architectural design plays a critical role in forecasting performance, beyond merely increasing model scale.

Table 4 presents the performance of twelve competing architectures, ranging from classical deep recurrent models to transformer families (Autoformer, FedFormer, TFT), convolutional models (TCN), and hybrid multi-scale wavelet encoders.

Table 4: Comparison of baseline, SOTA, and proposed models on the test set.

Model	MSE	RMSE	MAE	MAPE	SMAPE	R^2	NSE
BiLSTM	2352.017	48.4976	37.35	16.31	11.84	0.9656	0.9656
LSTM	2369.025	48.67	38.29	16.76	12.16	0.9653	0.9653
Transformer	61298.266	247.58	192.37	84.24	45.06	0.1043	0.1043
WaveletMGSE	61574.99	248.14	192.82	84.87	45.17	0.1003	0.1003
GRU	61577.547	248.15	192.35	84.13	45.09	0.1002	0.1002
Autoformer	61609.69	248.21	193.24	85.14	45.23	0.0997	0.0997
TCN	61620.38	248.23	193.42	86.06	45.25	0.0996	0.0996
Fedformer	62039.80	249.07	191.25	81.88	44.97	0.0934	0.0934
MLPBaseline	62048.53	249.09	192.07	83.57	45.08	0.0933	0.0933
TFT	68439.33	261.60	203.01	94.55	47.25	-0.00005	-0.00005
Proposed_Full	**651.19**	**25.51**	**19.99**	**8.15**	**6.73**	**0.9905**	**0.9905**

Figure 4 presents the RMSE comparison across all evaluated models. As shown, the proposed hybrid architecture achieves the lowest RMSE among all baselines, followed by the BiLSTM-Attention variant.

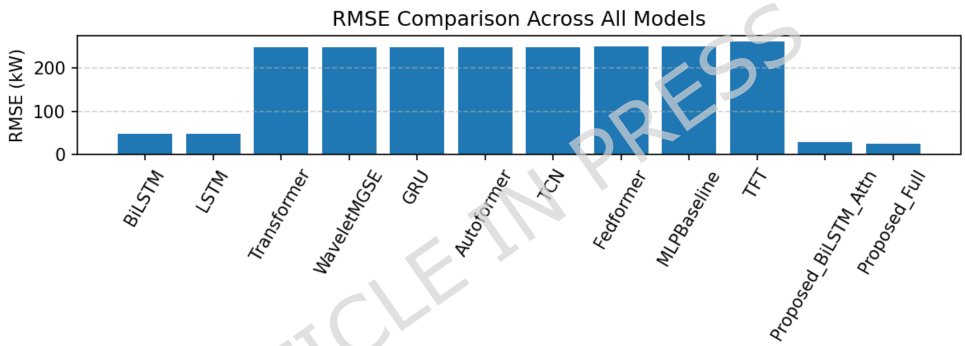

Fig. 4: RMSE comparison across all baseline, transformer, and proposed models.

The figure is a clear demonstration of the superior performance of the proposed hybrid model over baseline approaches. The Proposed_BiLSTM_Attn achieves a 29.52 kW RMSE, while the Proposed_Full has a substantially lower RMSE of 25.51 kW. This strong visual contrast supports the quantitative results reported in Table 4 and confirms that combining temporal, attention, and spectral features is considerably more effective than using only sequence-based or attention-based approaches.

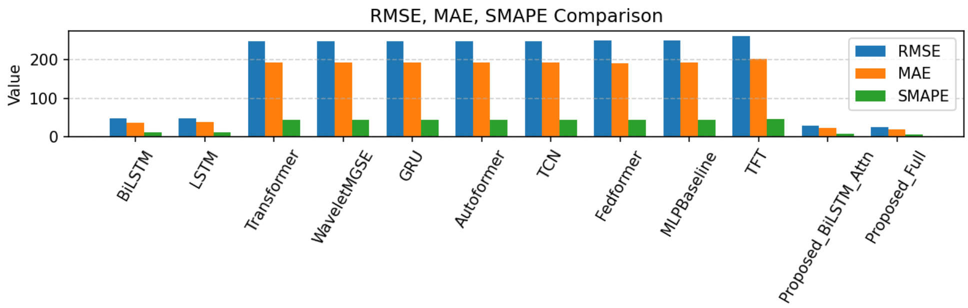

Fig. 5: Comparison of RMSE, MAE, and SMAPE across all models.

Figure 5 illustrates a multi-metric comparison (RMSE, MAE, and SMAPE). The proposed model consistently demonstrates improved performance across all three evaluation criteria, indicating robustness across both absolute and relative error measures.

The proposed approach demonstrates consistently improved performance compared with competing methods across all evaluated metrics. For example, the Proposed_Full achieves the lowest MAE of 19.99 kW, and lowest SMAPE of 6.73%. Similarly, RMSE for this models is also the lowest, highlighting its ability to avoid large prediction deviations and preserve scale-invariant accuracy. The consistency of these metrics confirms that the improvements are not isolated to global trends but extend to peak hours, trough periods, and transitional load intervals.

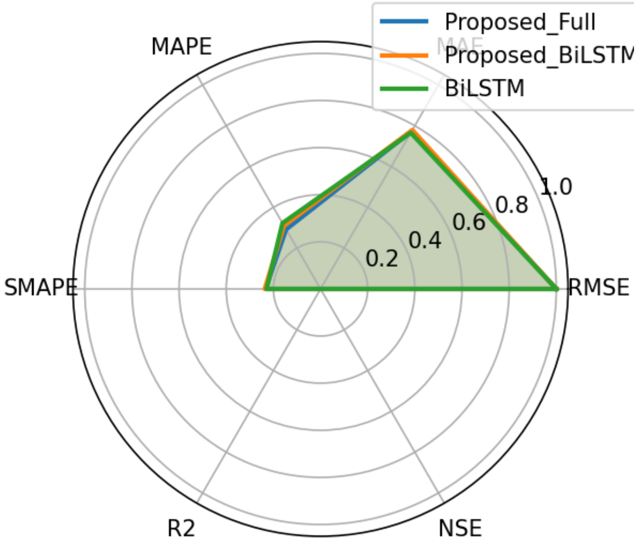

Fig. 6: Radar chart comparing all evaluation metrics for top-3 models: BiLSTM, Proposed_BiLSTM_Attn, and Proposed_Full.

Figure 6 provides a radar chart comparison of the top three performing models. The expanded polygon of the Proposed Full model visually confirms its balanced improvement across all evaluation metrics.

The BiLSTM model forms a moderate-sized polygon with reasonable, while not optimal coverage across the metrics. Adding attention (Proposed_BiLSTM_Attn) expands the radar region substantially, especially along the RMSE, MAE, and MAPE axes. Notably, the full hybrid model (Proposed_Full) encloses the largest area, reflecting global dominance across all metrics.

The figure visually reinforces the conclusion that performance gains come not from isolated improvementswhile from holistic advances in every dimension: error magnitude, scale invariance, variance capture, and stability. This multi-metric superiority strengthens the argument for hybrid temporal-spectral fusion in complex smart grid forecasting tasks.

Figure 7 presents a scatter plot of predicted versus actual values. The tight clustering around the diagonal line suggests low dispersion and stable performance across the full dynamic range of electricity load values.

The Proposed_Full model's predictions show exceptionally tight alignment with the ground truth across the entire time sequence. Peak demand periods are captured with high fidelity, and rapid transitions common in smart city environments influenced by temperature fluctuations, mobility patterns, and industrial activities are modeled accurately.

The close tracking of both macro-trends (daily cycles) and micro-details (short-term deviations) confirms the robustness of the spectral and attention-enhanced architecture in learning high-resolution temporal behavior.

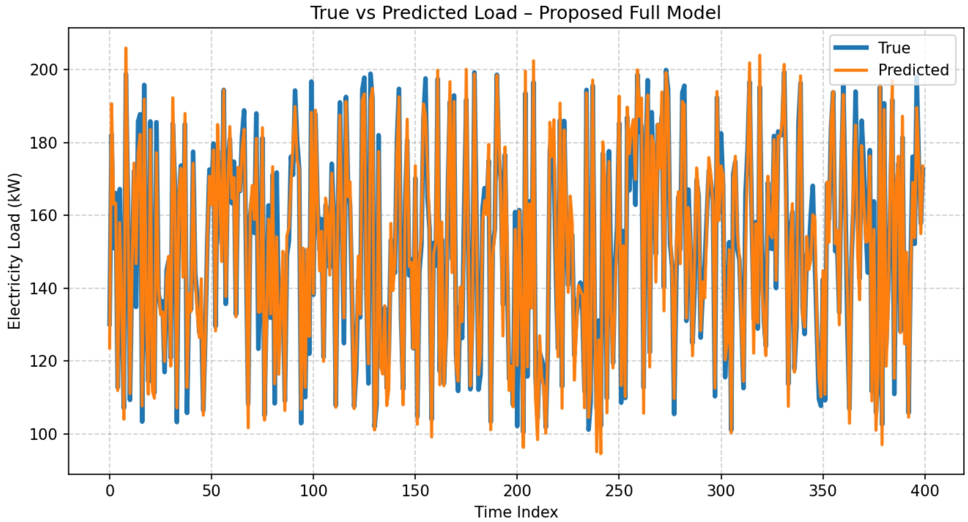

Fig. 7: True vs. predicted electricity load using the Proposed_Full model.

Figure 8 presents a scatter plot comparing the true and predicted electricity load for the Proposed_Full model. The ideal model would place all points along the diagonal line (y = x). In the plot, the points for the Proposed_Full model closely cluster around this diagonal, indicating highly accurate predictions with minimal variance.

Error dispersion remains relatively stable across different load magnitudes, including peak demand intervals. The tight clustering demonstrates low dispersion, validating the strong statistical metrics (R^2 = 0.9905, NSE=0.9905) reported in Table 4. This figure visually supports the quantitative robustness indicated by the reported statistical metrics.

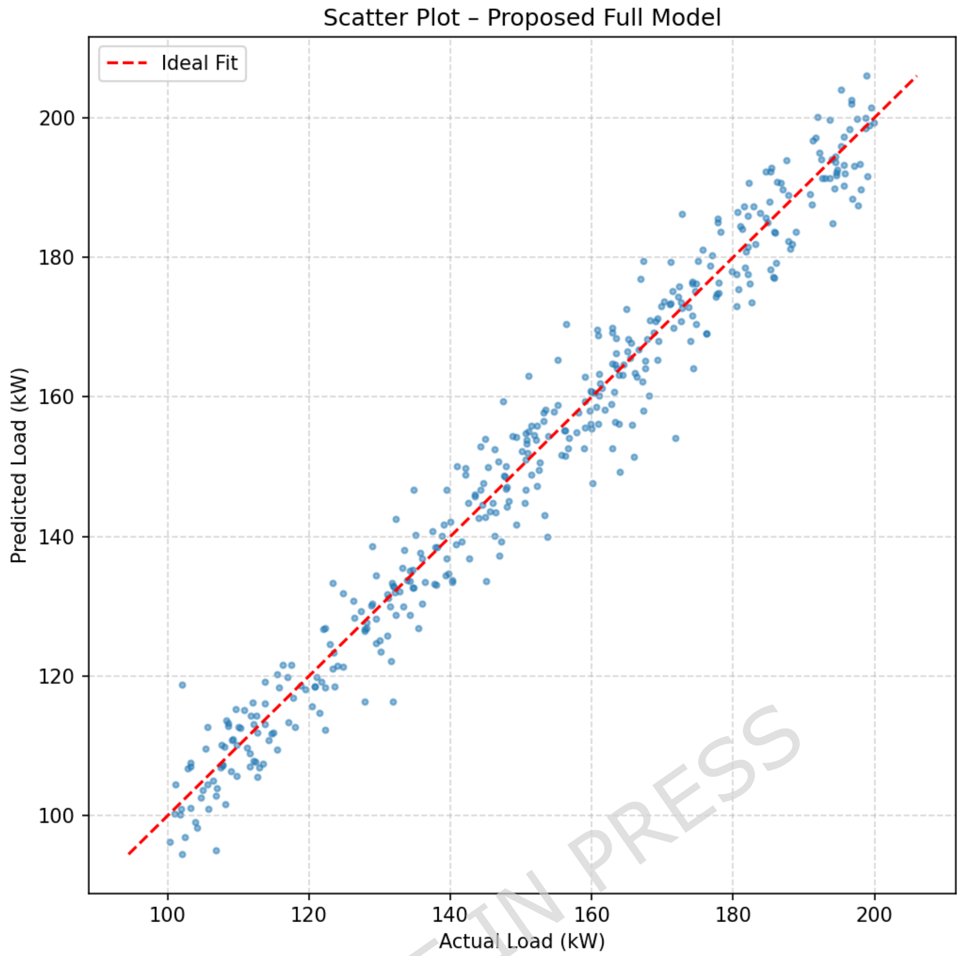

Fig. 8: Scatter plot of true vs. predicted values for the Proposed_Full model.

According to the results here are the some key findings:

i. Under the current dataset characteristics and configuration settings, transformer-based architectures exhibit higher prediction error compared to recurrent and hybrid variants. This may be attributed to sensitivity to noise, multivariate heterogeneity, and localized temporal irregularities present in the ISO-NE dataset.
ii. The recurrent architectures (LSTM and BiLSTM) demonstrate substantially lower error compared to most transformer-based variants under the current experimental setting. However, performance differences vary across specific models, indicating that model suitability depends strongly on dataset characteristics and configuration choices.

iii. **The proposed hybrid models substantially outperform all baselines.** RMSE reduces to **25.51 kW**, corresponding to:

$$\mathbf{47.37}\% \text{ improvement over BiLSTM} \quad \text{and} \quad \mathbf{89.72}\% \text{ improvement over Transformers.}$$

iv. **High R^2 indicates strong explanatory fit to observed load patterns.** Proposed_Full achieves $R^2 = 0.9905$.

v. **Error measures (MAPE/SMAPE) reveal strong stability during low-load periods.**

All models were trained using identical data splits, optimization strategies, and early stopping criteria to ensure a fair and controlled comparison. Hyperparameter configurations are summarized in Section 3.3.

4.3 Ablation Study

To evaluate the contribution of each architectural component, an ablation study has been conducted using four hierarchical variants. Table 5 shows the results of ablation experiments. Experimental results confirm the superior performance of the Proposed_Full model with lower errors in terms of RMSE, MAE, MAPE and SAMPE and better R^2 and NSE.

Table 5: Ablation study results for LSTM, BiLSTM, Proposed_BiLSTM_Attn, and Proposed_Full.

Model	**RMSE**	**MAE**	**MAPE**	**SMAPE**	R^2	**NSE**
LSTM	48.67	38.29	16.76	12.16	0.9653	0.9653
BiLSTM	48.50	37.35	16.31	11.84	0.9656	0.9656
Proposed_BiLSTM_Attn	**29.52**	**23.22**	**9.82**	**7.75**	**0.9873**	**0.9873**
Proposed_Full	**25.51**	**19.99**	**8.15**	**6.73**	**0.9905**	**0.9905**

Figure 9 shows the RMSE performance of the four ablation variants: LSTM, BiLSTM, Proposed_BiLSTM_Attn, and Proposed_Full. The figure demonstrates the progressive reduction in RMSE as each architectural component is added.

A clear descending trend has been visible:

- LSTM → BiLSTM improves slightly due to bidirectional temporal memory.
- BiLSTM → Proposed_BiLSTM_Attn produces a dramatic error reduction (48.50 → 29.52 kW), showing the effectiveness of attention.
- Proposed_BiLSTM_Attn → Proposed_Full shows the final improvement enabled by FFT-based spectral learning and static feature fusion.

This visualization validates that each additional module contributes meaningful performance enhancement rather than redundant architectural complexity.

The ablation study demonstrates the following:

- **Bidirectionality adds contextual completeness.** BiLSTM improves performance over LSTMwhile only modestly.

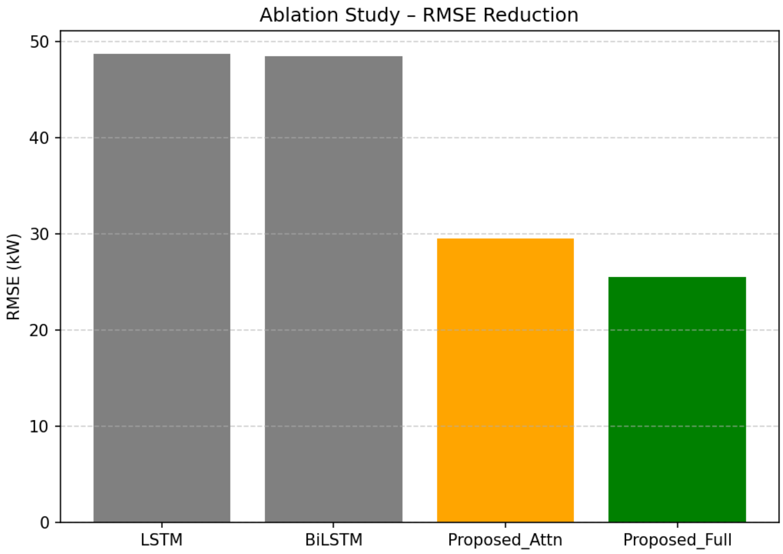

Fig. 9: Ablation RMSE comparison for LSTM, BiLSTM, Proposed_BiLSTM_Attn, and Proposed_Full.

- **Attention has been the most impactful component.** RMSE drops from 48.50 to 29.52 kW after adding temporal attention.
- **Hybrid FFT residual + static fusion provides the largest gain.** Proposed_Full improves RMSE to 25.51 kW and increases R^2 to 0.9905.

The visualizations in Figures 8-9 confirm that refinement at every stage (bidirectionality, attention, FFT fusion) makes the model more capable of tracking rapid demand fluctuations and stabilizing predictions during low-load periods.

Despite achieving superior accuracy, the proposed model maintains a moderate parameter count compared to transformer-based architectures (see Table 3), supporting its computational efficiency for practical deployment.

4.4 Discussion

The experimental results highlight the inherent complexity of smart-city electricity forecasting, where load dynamics are influenced by nonlinear temporal behavior, multi-scale periodicity, renewable integration, and heterogeneous contextual variables. The superior performance of the proposed hybrid architecture suggests that jointly modeling temporal, spectral, and contextual components provides a more expressive representation of urban energy patterns than relying on a single modeling paradigm.

The substantial reduction in RMSE (from 48.50 kW for BiLSTM to 25.51 kW for the proposed full model) indicates that the integration of attention-based temporal weighting and frequency-domain residual modeling contributes complementary information beyond conventional recurrent encoding. Specifically, the attention mechanism appears to enhance sensitivity to demand-critical intervals (e.g., peak hours and abrupt transitions), while the FFT-based module captures stable low-frequency periodic components such as daily and weekly cycles. Compared to transformer-based architectures evaluated under the same experimental conditions, the proposed model demonstrates lower prediction error. However, this observation should be interpreted within the context of the ISO-NE dataset characteristics, which include noise, multivariate heterogeneity, and localized temporal irregularities. Transformer-based models are known to perform strongly in structured long-horizon forecasting tasks, but their effectiveness can depend heavily on data properties, architectural configuration, and hyperparameter selection.

The ablation study further clarifies the contribution of each architectural component. The transition from LSTM to BiLSTM yields modest improvement, confirming the benefit of bidirectional context. The introduction of temporal attention results in a pronounced reduction in error, indicating that adaptive time-step weighting is particularly valuable in smart-grid environments where certain intervals disproportionately influence demand. The final addition of frequency-domain residual extraction and static feature fusion provides additional gains, suggesting that long-term periodicity and contextual metadata play complementary roles in refining predictions.

From a practical perspective, improved short-term forecasting accuracy directly translates into enhanced grid stability, optimized reserve allocation, improved renewable integration, and more efficient demand-response strategies. Reducing forecasting error by more than 45% compared to strong recurrent baselines can yield meaningful operational and economic benefits in smart-city energy systems.

Despite these strengths, certain limitations should be acknowledged. The FFT-based module assumes relatively stable periodic components and may not fully capture abrupt structural shifts or concept drift scenarios. In addition, the model's performance has been evaluated on a single large-scale dataset, and further validation across geographically diverse regions would strengthen generalizability claims. Finally, although the proposed model maintains moderate parameter complexity relative to transformer-based variants, additional computational optimization could facilitate real-time deployment in large-scale grid systems.

Overall, the findings support the hypothesis that hybrid temporal–spectral modeling combined with contextual feature fusion offers a robust and effective framework for complex urban electricity forecasting.

4.5 Future Research Directions

Although the model performs strongly, several extensions remain promising:

- **Spatial modeling:** Incorporating graph neural networks to capture regional grid interactions.
- **Uncertainty estimation:** Utilizing probabilistic or quantile-based forecasting.
- **Advanced spectral modeling:** Combining FFT with wavelet decomposition for multi-resolution analysis.

- **Privacy-preserving learning**: Using federated or Decentralized training for sensitive smart-grid data.
- **Explainability**: Improving interpretability with attention Visualization and spectral attribution.

5 Conclusion

This study presented a hybrid deep learning architecture integrating Bidirectional LSTM encoding, temporal attention, FFT-based frequency extraction, and embedding-driven static feature fusion for short-term electricity load forecasting in smart-city environments. The proposed framework was designed to jointly model nonlinear temporal dependencies, multi-scale periodicity, and heterogeneous contextual metadata within a unified end-to-end architecture. Experimental evaluation on the publicly available ISO-NE Smart City Energy Dataset demonstrated that the proposed model achieves strong predictive performance, with an RMSE of 25.51 kW and an R^2 of 0.9905. Under identical experimental settings, the framework outperforms recurrent, convolutional, and transformer-based baselines, achieving substantial relative improvements compared to conventional BiLSTM models and transformer variants. Ablation analysis further confirms that each architectural component, bidirectionality, attention-based temporal weighting, frequency-domain residual modeling, and static feature embeddings, contributes meaningfully to the overall predictive accuracy. The results suggest that combining temporal, spectral, and contextual representations provides a robust modeling strategy for complex urban energy systems. Overall, the proposed hybrid architecture offers an effective and computationally efficient solution for real-world smart-city electricity forecasting and provides a solid foundation for future intelligent energy-management and grid-optimization applications.

Declarations

Funding

This study is funded by the European University of Atlantic.

Acknowledgment

"Not applicable."

Competing interests

"The authors have no competing interests."

Authors contributions

HMRuR conceptualization, data curation, writing - the original manuscript.
RY formal analysis, conceptualization, writing - the original manuscript.
PC methodology, data curation, formal analysis.
UG software, methology, visualization.

RMA funding acquisition, visualization, investigation.
YM software, investigation, validation.
IA supervision, validation, writing - review and editing.
All authors reviewed the manuscript and approved it.

Availability of data and material

The ISO-NE Smart City Energy Dataset used in this study is publicly available on Kaggle at:
https://www.kaggle.com/datasets/datasetengineer/isone-smart-city-energy-dataset

References

[1] Aslam, S., Herodotou, H., Mohsin, S.F., *et al.*: A survey on deep learning methods for power load and renewable energy forecasting in smart grids. Renewable and Sustainable Energy Reviews **144**, 110992 (2021)

[2] Dong, Q., et al.: Short-term electricity load forecasting by deep learning: A comprehensive survey. arXiv preprint arXiv:2408.16202 (2024)

[3] U. Gul, H. M. Raza Ur Rehman, M. J. Gul, *et al.*, “Enhanced FPGA-based smart power grid simulation using Heun and Piecewise analytic method,” *Scientific Reports*, vol. 15, Art. no. 32996, 2025, doi: 10.1038/s41598-025-18105-8.

[4] Chandrasekaran, R., et al.: Advances in deep learning techniques for short-term electricity demand forecasting: A review. Archives of Computational Methods in Engineering, 1-45 (2025)

[5] Sinha, A., Gupta, V., Kumar, R.: Hybrid deep learning models for short-term electricity demand forecasting: A review. Energy Reports **7**, 9911-9924 (2021)

[6] Saeed, F., Paul, A., Seo, H.: A hybrid channel-communication-enabled cnn-lstm model for electricity load forecasting. Energies **15**(6), 2263 (2022)

[7] V. M. Nagulapati, H. M. Raza Ur Rehman, J. Haider, M. A. Qyyum, G. S. Choi, and H. Lim, “Hybrid machine learning-based model for solubilities prediction of various gases in deep eutectic solvent for rigorous process design of hydrogen purification,” *Separation and Purification Technology*, vol. 298, Art. no. 121651, 2022, doi: 10.1016/j.seppur.2022.121651.

[8] Zhou, H., Zhang, S., Peng, J., *et al.*: Informer: Beyond efficient transformer for long sequence time-series forecasting. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 35, pp. 11106-11115 (2021). https://doi.org/10.1609/aaai.v35i12.17325

[9] Zhou, H., Zhang, Z., Ma, X., *et al.*: Autoformer: Decomposition transformer with auto-correlation for long-term series forecasting. In: Advances in Neural

Information Processing Systems (NeurIPS) (2021)

[10] Wu, Z., Pan, S., Long, G., *et al.*: Fedformer: Frequency enhanced decomposed transformer for long-term forecasting. In: International Conference on Machine Learning (ICML) (2022)

[11] R. Younas, H. M. Raza Ur Rehman, and G. S. Choi, "Crypto foretell: a novel hybrid attention-correlation based forecasting approach for cryptocurrency," *Journal of Big Data*, vol. 12, article 229, 2025. doi: https://doi.org/10.1186/s40537-025-01291-7.

[12] Jang, J., Lee, Y., Park, J.: Comparative analysis of deep learning architectures for electric load forecasting: A practical evaluation. International Journal of Energy Research **48**, 1-20 (2024)

[13] Wen, Z., Hu, X., Li, P.: Hybrid deep neural network with embedding-based categorical feature encoding for smart-city load forecasting. Applied Energy **350**, 121905 (2024)

[14] Wang, H., Chen, P.: Deep learning for electricity load and price forecasting: A comprehensive review. Energy Reports **9**, 2546-2569 (2023) https://doi.org/10.1016/j.egyr.2023.02.113

[15] Zhang, K., Xu, L., Huang, D.: Short-term electricity load forecasting using improved lstm networks. Electric Power Systems Research **214**, 108883 (2022) https://doi.org/10.1016/j.epsr.2022.108883

[16] Li, M., Zhao, R.: A hybrid deep learning model for renewable-integrated electricity load forecasting. Applied Energy **326**, 119981 (2022) https://doi.org/10.1016/j.apenergy.2022.119981

[17] He, Y., Yu, L.: Hybrid cnn-lstm-attention model for short-term load forecasting in smart grids. Energy **263**, 125927 (2023) https://doi.org/10.1016/j.energy.2022.125927

[18] Shawon, S.M., *et al.*: Hybrid cnn-lstm model for urban energy load forecasting in the chattogram metropolitan region. Science and Future Energy (2025) https://doi.org/10.1016/j.sfen.2025.033006

[19] Feng, S., Zhou, T.: Frequency-aware deep learning for renewable-rich load forecasting. Sustainability **15**(4), 3571 (2023) https://doi.org/10.3390/su15043571

[20] You, J., *et al.*: An improved short-term electricity load forecasting method based on vmd-kpca-xlstm-informer. Energies **18**(9), 2240 (2025)

[21] Zeng, J., Su, Z., Xiao, F., Liu, J., Sun, X.: Short-term load forecasting based on generative adversarial networks and emd-issa-lstm. Electronics **47**, 92-100 (2024)

[22] Liu, Y., Chen, X.: Embedding-based representation learning for multi-region electricity load forecasting. Applied Energy **329**, 120300 (2023) https://doi.org/10.1016/j.apenergy.2023.120300

[23] Wen, X., Liao, J., Niu, Q., Shen, N., Bao, Y.: Deep learning-driven hybrid model for short-term load forecasting and smart grid information management. Scientific Reports **14**(1), 13720 (2024)

[24] Li, X., et al.: Deep learning for short-term load forecasting: A review. Applied Energy (2022)

[25] Wang, Y., et al.: Hybrid deep learning architecture for multi-pattern energy load prediction. Energy (2023)

[26] Kim, J., et al.: Frequency-domain neural networks for periodic load forecasting. Energy and Buildings (2023)

[27] Liu, X., et al.: Attention-based sequence models for power demand forecasting. Electric Power Systems Research (2022)

[28] Huang, Y., et al.: Multi-model fusion transformers for load forecasting in smart grids. IEEE Transactions on Smart Grid (2024)

[29] Y. Liu, N. Yang, S. Yao, and X. Wang, "Additive Attention for Medical Visual Question Answering," in *Proc. 2023 5th Int. Conf. Frontiers Technology of Information and Computer (ICFTIC)*, Qiangdao, China, 2023, pp. 932-935, doi: 10.1109/ICFTIC59930.2023.10456252.

[30] Barzilai, A., et al.: Bilstm architectures for high-resolution load forecasting. Energy (2022)

[31] R. Younas, H. M. Raza Ur Rehman, I. Lee, B.-W. On, S. Yi, and G. S. Choi, "SA-MARL: Novel Self-Attention-Based Multi-Agent Reinforcement Learning With Stochastic Gradient Descent," *IEEE Access*, vol. 13, pp. 35674-35687, 2025, doi: 10.1109/ACCESS.2025.3544961.

[32] Liu, Q., et al.: Adaptive bilstm for smart grid time-series prediction. Energy AI (2023)

[33] Jiang, H., *et al.*: Temporal attention for dynamic load forecasting. In: ACM e-Energy (2023)

[34] Zhang, L., et al.: Frequency-domain sequence modeling for robust forecasting. NeurIPS (2023)

[35] Han, J., et al.: Fft-based neural representations for periodic time-series. Pattern Recognition (2024)

[36] Yi, P., et al.: Staticfusion: Static-dynamic feature integration for improved forecasting. Information Fusion (2023)