Journal of Mechanical, Civil and Industrial Engineering
ISSN: 2709-0809
DOI: 10.32996/jmcie
Journal Homepage: www.al-kindipublisher.com/index.php/jmcie

JMCIE
AL-KINDI CENTER FOR RESEARCH AND DEVELOPMENT

| RESEARCH ARTICLE

Industrial Power Load Forecasting for Grid Operation Using a CNN-Transformer-BiLSTM Model

Remon Das[1]✉, Md Abdul Ahad Juel[2], and Md. Raisul Islam[3]
[1]*MS in Engineering Technology, Western Carolina University, North Carolina, USA*
[2]*University at Buffalo, USA, abdulahadjuel@gmail.com*
[3]*Western Illinois University, USA, raisulislamjony@gmail.com*
Corresponding Author: Remon Das **E-mail**: rdas1@catamount.wcu.edu/remondas4@gmail.com

| ABSTRACT

In the time of AI era, Industrial power load is gradually rising due to the rapid expansion of the chip manufacturing facilities. So that accurate forecasting of industrial power load is important to achieve efficient grid planning and overall energy management. But, due to the nonlinear, volatile and multi scale nature of industrial power load data, the conventional statistical model face challenges in forecasting efficiently. To address these challenges, a novel hybrid deep learning model, CNN-Transformer-BiLSTM has been proposed that integrates the feature extraction capacity of convolutional neural networks (CNN), the long-range dependency modeling of the transformer architecture and the sequential learning strength of bidirectional long, short-term memory (BiLSTM) networks. The CNN layers efficiently capture the local temporal patterns and feature correlations within the load data sets, Transformer layers employs self-attention mechanisms to model complex long-term dependencies and contextual relationships. The BiLSTM layer further enhances temporal representation by learning bidirectional dependencies, thus improving the overall prediction accuracy. Historical monthly industrial electricity load data from the U.S. Energy Information Administration (EIA) spanning over two decades are used to train and evaluate the model. The proposed model output has been compared with other standalone and hybrid deep learning models. The proposed CNN-Transformer-BiLSTM achieves superior forecasting accuracy with Mean Absolute Percentage Error (MAPE) of 1.23%, Root Mean Square Error (RMSE) of 1,276 MWh and Mean Absolute Error (MAE) of 1,040 MWh.

| KEYWORDS

Industrial Electricity Forecasting, Deep Learning, CNN, BiLSTM, Transformer, Time-Series Forecasting

| ARTICLE INFORMATION

ACCEPTED: 01 January 2026 **PUBLISHED:** 22 January 2026 **DOI:** 10.32996/jmcie.2026.7.1.3

1. Introduction

The efficient and reliable operation of modern power grids critically depends on accurate electricity demand forecasting, which serves as a fundamental component of operational planning, grid stability assessment, and strategic financial decision-making [1]. Among all consumer categories, the industrial sector poses one of the most complex yet impactful forecasting challenges due to its high load magnitude, volatility, and sensitivity to economic and operational factors. In 2021, industrial electricity consumption accounted for approximately 33% of total U.S. energy use, underscoring its central role in grid operation and planning [2]. For nearly two decades, U.S. electricity demand remained relatively stable, as efficiency improvements and structural shifts from manufacturing to service-oriented economies offset growth driven by population expansion and economic activity. However, this long-standing trend has changed markedly since 2020. Accordingly, to the U.S. Energy Information Administration (EIA), total electricity consumption is projected to grow at an average annual rate of 1.7% through 2026, with the commercial and industrial sectors experiencing even higher growth rates of 2.6% and 2.1%, respectively [3]. More notably, industrial electricity demand is expected to increase by approximately 43% between 2024 and 2050, driven primarily by the electrification of industrial processes and the rapid expansion of energy -intensive data centers [4]. This accelerating and increasingly dynamic demand profile highlights the urgent need for more accurate, adaptive, and robust load forecasting methodologies tailored to industrial gird operation.

Traditional electrical power load forecasting has largely relied on statistical techniques such as autoregressive integrated moving average (ARIMA), Holt Winters exponential smoothing, and regression-based models [5]. While these methods offer interpretability and ease of implementation, they often struggle to capture the nonlinear relationships, non-stationary behavior, and long-term temporal dependencies inherent in real world energy consumption data [5]. To address these limitations, various advanced forecasting approaches have been proposed, including fuzzy neural networks [6], gray algorithms [7], gray Markov model [8] and support vector regression techniques [9]. However, many of these methods exhibit limited scalability and reduced effectiveness when applied to large scale, high dimensional or highly volatile datasets typical of industrial load profiles [10-13]. In related renewable energy forecasting applications, such as photovoltaic (PV) power prediction, Artificial Neural Networks (ANNs) and Support Vector Machines (SVMs) [14-15] have been widely adopted. Although ANNs generally outperform classical time-series models on nonlinear datasets, they often encounter challenges related to computational complexity, overfitting, and limited performance when handling large datasets or complex feature interactions [15].

More recently, deep learning-based approaches have gained significant traction in power load forecasting, particularly Recurrent Neural Networks (RNNs) and their advanced variants, such as Long Short-Term Memory (LSTM) networks [16]. LSTM architectures are well suited for energy forecasting tasks due to their ability to model long term temporal dependencies and mitigate vanishing gradient issues [17]. However conventional LSTM models process sequential data in a unidirectional manner, potentially restricting their ability to fully exploit temporal correlations across past and future observations. Bidirectional LSTM (BiLSTM) networks address this limitation by learning temporal dependencies in both forward and backward directions, thereby enabling a more comprehensive representation of load dynamics [18]. Despite their improved predictive capability, Bi-LSTM models typically require large training datasets and incur substantial computational costs, which can hinder their practical deployment in real-time or large-scale industrial forecasting applications.

In parallel, Transformer-based architectures have emerged as powerful alternatives for time series forecasting, including electricity load prediction, owning to their attention mechanisms and series forecasting, including electricity load prediction, owing to their attention mechanisms and ability to capture long-range dependencies without recurrent structures. Several studies have demonstrated the effectiveness of Transformer variants in energy forecasting. For example, Fourier Transform-based transformers combined with enhanced optimization algorithms have been proposed to achieve high-resolution load forecasting while addressing sequence redundancy and computational efficiency challenges in energy systems [19]. Informer [20], Autoformer [21], SmartFormer [22], Temporal Fusion Transformer (TFT), improved Autoformer [24] have been introduced in different load application. Despite their strong predictive capabilities, transformer-based models exhibit notable limitations, including high data requirements and sensitivity to data quality, which can pose challenges in industrial environments where historical load data and auxiliary covariates may be limited or proprietary [25].

To overcome the limitations of existing machine learning models for industrial power load forecasting, a framework called CNN-Transformer-BiLSTM has been proposed. In data preprocessing begins with seasonal-tread-loess (STL) decomposition, which separates the raw time series into trend seasonal, and residual components. Each component is then processed through a Convolutional Neural Network (CNN) to capture short-term, long temporal dependencies. The outputs of the CNN layers are fed into Transformer block, which efficiently models long range dependencies and global correlations across time. Finally, a Bidirectional LSTM (BiLSTM) layer refines the sequential information in both forward and backward directions, providing a contextual understanding of temporal dynamics that further improves forecasting accuracy. Collectively, the proposed architecture addresses the challenges of noise, multiscale temporal patterns, and long-term dependencies, making it particularly suitable for the complex and irregular patterns observed in the U.S. industrial power load data.

The contributions of this research work are:

- A novel hybrid deep learning-based model **CNN-Transformer-BiLSTM** has been utilized for industrial power load forecasting in the U.S. This approach is designed to compatible the model for feature extraction, long range dependency modeling and the sequential learning strength. To improve the quality of learning and enable the models to handle diverse seasonal patterns and trends in industrial power load requirement, STL (Seasonal-Trend decomposition using Loess) has been introduced as a data preprocessing step.
- This work addresses a notable gap in the existing literature, as there has been no prior study focused on industrial power load forecasting in the United States.

The structure of the paper is organized as follows: Section II describe the related work on industrial power load forecasting models, Section III presents the performance evaluation matrix to evaluate our proposed model, Section IV outlines the data collection and preprocessing, Section V depicts the details about the proposed forecasting model Section VI results and discussion present benchmarking outcomes and insights, and the final section VII concludes with key findings and future research directions.

2. Literature Review on Industrial Load Forecasting

In this section, a range of different previously developed data driven model for industrial power load forecasting has been discussed and highlighting their methodologies, performance, and real-world applicability. A hybrid intelligent forecasting model combining Reinforcement Learning, Particle Swarm Optimization (PSO), and Least Squares Support Vector Machine (LSSVM) for short-term industrial power load forecasting was developed for single industrial power consumer [26]. Experimental results demonstrated that the proposed Q-PSO-LSSVM model achieved significantly lower mean absolute percentage error (MAPE) 3–5%. A hybrid model named TCN-LightGBM [27] combining a Temporal Convolutional Network (TCN) and Light Gradient Boosting Machine (LightGBM) for accurate short-term industrial power load forecasting has been developed. The model uses a fixed length sliding time window to reconstruct electrical features (load, current, power, etc.) along with meteorological and calendar data, enabling the TCN to extract long-term temporal dependencies and hidden patterns. TCN-LightGBM achieves the lowest MAPE around 2-4%. Neural Network Autoregression (NNAR) and Multilayer Perceptron (MLP) model proposed for forecasting monthly industrial electricity consumption in Brazil. The study uses historical data from 1979 to 2020, obtained from the Central Bank of Brazil, divided into training (1979-2018) and testing (2019-2020) datasets. MLP model outperformed others, achieved the lowest MAPE 3.4% [28]. A novel Fourier Transform (FT)-Transformer and Genetic Algorithm (GA) hybrid model [29] for accurate energy forecasting and load optimization has been developed in industrial and commercial energy systems. The FT-Transformer leverages self-attention mechanisms and Fourier-based seasonality encoding to capture long-term dependencies and temporal patterns in large-scale energy demand data. Experimental results demonstrate that the FT-Transformer achieved a Mean Absolute Error (MAE) of 3.03×10^5 kWh and Root Mean Square Error (RMSE) of 3.31×10^5 kWh, outperforming RNN, PSO, and tree-based models by up to 48% in accuracy and reducing computational time by 38%. A hybrid forecasting model DCN–E-TCN [30], that integrated a Data Completion Network (DCN) with an Enhanced Temporal Convolutional Network (E-TCN) for ultra-short-term industrial load prediction. Results showed that the DCN–E-TCN model significantly outperforms conventional TCN, GRU, and LSTM models, achieving up to 35% lower RMSE and 30% higher R^2, leading to highly stable and precise forecasts suitable for industrial demand response, production scheduling, and grid stability applications. A hybrid ensemble forecasting model [31] that combined Long Short-Term Memory (LSTM) networks with multiple machine learning algorithms, Support Vector Regression (SVR), Extreme Gradient Boosting (XGBoost), and Random Forest (RF) to enhance the accuracy and robustness of industrial power demand forecasting. The ensemble used Bayesian optimization for parameter tuning and a stacking strategy to fuse the predictions from base learners, enabling effective capture of both linear and nonlinear temporal dependencies. Experimental results showed that the proposed LSTM-based hybrid ensemble achieved superior performance with a Mean Absolute Percentage Error (MAPE) below 1.5%, outperforming standalone deep learning and traditional models such as GRU, CNN, and ARIMA. A case study of Vietnam introduced a hybrid forecasting model that combines Ensemble Empirical Mode Decomposition (EEMD) with a Long Short-Term Memory (LSTM) [32] network to improve the accuracy of short-term industrial load prediction. The model validated using hourly power load data from a Vietnamese industrial manufacturing plant, collected over approximately 18 months and characterized by its nonlinear and highly fluctuating nature. Experimental results demonstrated that the proposed EEMD-LSTM model achieves superior performance with a Mean Absolute Percentage Error (MAPE) of approximately 1.3% for 1-step forecasting, significantly outperforming standalone models like Linear Regression (LR), Artificial Neural Networks (ANN), and a standard LSTM. A novel deep learning model that leveraged a hybrid ensemble strategy and an error correction mechanism to accurately forecast industrial power demand. The model first employed an ensemble of Gated Recurrent Unit (GRU) networks, where base learners were generated by perturbing GRU parameters using a novel Multi-Objective Molecular Dynamics Theory Optimization Algorithm (MMDTOA) and then integrated via kernel ridge regression stacking. Experimental results demonstrated that the proposed model achieves superior performance with a Normalized Mean Absolute Error (NMAE) as low as 3.684%, outperforming eight other models including CNN and LSTM, thereby highlighting its effectiveness and robustness for complex industrial forecasting tasks [33]. A hybrid machine learning model designed for day-ahead industrial load forecasting utilizes an Extreme Learning Machine (ELM) as its base predictor, with its initial weights and biases optimized by a Firefly Algorithm (FA) to enhance accuracy. The results showed that the proposed LCR-AdaBoost-FA-ELM model achieves a Mean Absolute Percentage Error (MAPE) of 3.01%, significantly outperforming standalone ELM and SVR models [34]. A Typical Load Profile (TLP)-supported Convolutional Neural Network (CNN) framework [35] designed to improve short-term industrial load forecasting accuracy by incorporating typical daily and weekly load patterns into the learning process. Results demonstrated that the proposed TLP-CNN model outperforms baseline models such as LSTM, GRU, and SVR, achieving a Mean Absolute Percentage Error (MAPE) of 2.17%, thereby enhancing forecasting stability and adaptability across varying industrial processes.

3. Performance Evaluation Matrix

3.1 Mean Absolute Error (MAE)

Mean Absolute Error (MAE) is computed by taking the sum of the absolute difference between each predicted value and its true value and then dividing by the total number of data samples [35]. The MAE depicts as:

$$MAE = \frac{1}{N}\sum_{i=1}^{N} |y_i - f_i|$$

where y_i is the actual and f_i is the forecasted value for the power load and N is the number of data samples.

3.2 Root Mean Squared Error (RMSE)

Root Mean Squared Error (RMSE) defines as the square root of the average squared difference of actual value and prediction value, in other words, the square root of Mean Squared Error (MSE) [35]. The RMSE is depicted as:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N} (y_i - f_i)^2}$$

where y_i is the actual and f_i is the forecasted value for the power load and N is the number of data samples.

3.3 Mean Absolute Percentage Error (MAPE)

Mean Absolute Percentage Error (MAPE) calculates the average of the absolute percentage errors between actual and predicted values. [36] The MAPE is defined as:

$$MAPE = \frac{100}{N}\sum_{i=1}^{N} |\frac{y_i - f_i}{y_i}|$$

where y_i is the actual value and f_i is the forecasted value for the power load, and N is the number of data samples

3. Performance Evaluation Matrix

In this section, the detailed process for collecting and pre-processing industrial power load data has been discussed

3.1 Data Collection

The historical industrial power load dataset obtained from the U.S. Energy information administration (EIA), as shown in Figure 1 as curve. Datasets are arranged in monthly periods and contain 159 observations from January 2012 to May 2025. From the curve, we have seen that, the total industrial power load has increased gradually, with a notable power requirement drop in 2020.

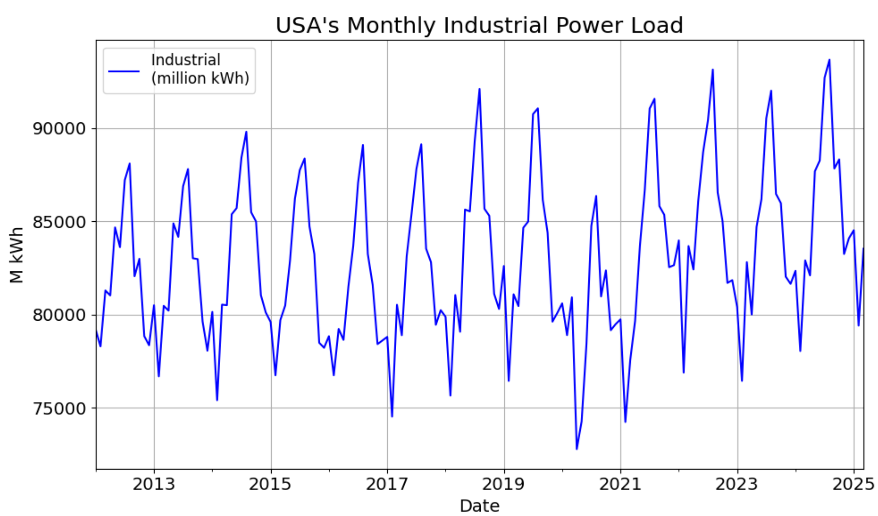


Figure 1: U.S. Monthly Industrial Power Load Data from January 2012 to May 2025

3.2 Data Preprocessing by Decomposition Technique

To enhance the quality of forecasting and enable the model to learn distinct temporal patterns more effectively, we apply a decomposition-based data preprocessing approach prior to model training. Specifically, we use the Seasonal-Trend decomposition using Loess (STL) technique to disaggregate the original time series into three fundamental components: trend, seasonal, and residual as shown in Figure 2. This decomposition enables the separation of systematic patterns in the data, thereby simplifying the complexity of the forecasting task and improving model interpretability.

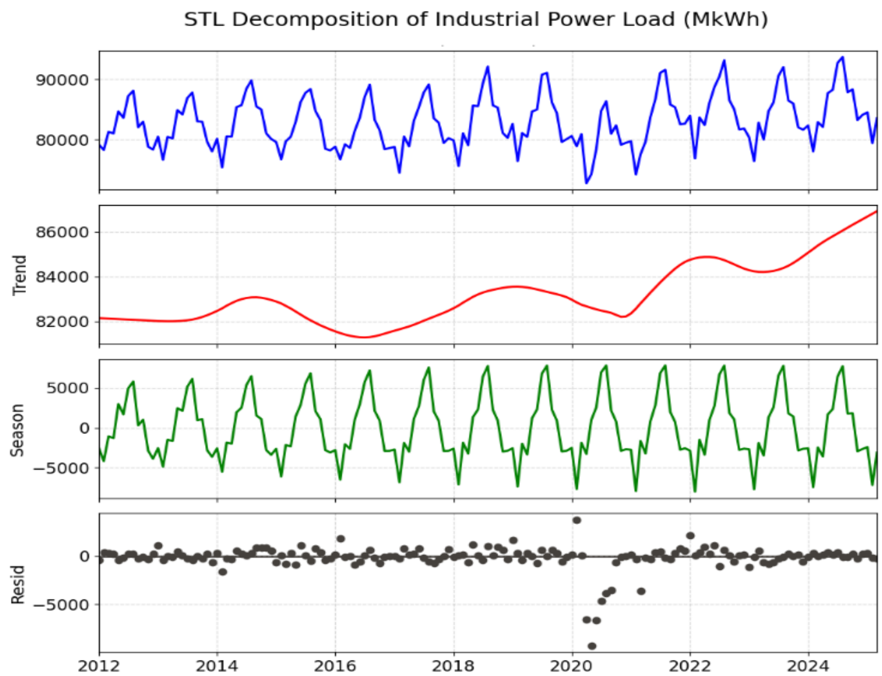


Figure 2: Seasonal-Trend decomposition using Loess (STL) technique

Mathematically, the original time series Y(t) is expressed as the sum of its components:

$$\boldsymbol{Y(t) = T(t) + S(t) + R(t)}$$

Where, T(t) represents the trend component, S(t) denotes the seasonal component, and R(t) is the residual component.

After decomposition, each component T(t), S(t) and R(t) is individually transferred as a separate input feature to the proposed hybrid deep learning architecture. This enables the model to learn unique temporal dependencies from each component without interference. The trend component contributes to long-term memory learning, the seasonal part provides information about periodic patterns and the residual highlights anomalies or short-term volatility. By integrating STL decomposition into the data preprocessing pipeline, the model benefits from a more structured and noise reduced input space, ultimately targeting to improve forecasting accuracy.

3. Proposed Model for Industrial Power Load Forecasting

In this study, seasonal trend decomposition using loess (STL) techniques has been utilized in the proposed hybrid model, to achieve high precision industrial load forecasting. This section includes the details of the forecasting methodology to generate final load predictions, the architecture design of the proposed hybrid model, and the parameter configuration by hyperparameter tuning process.

3.1 Flowchart of Forecasting Method by Proposed Model

The flowchart of the proposed methodology for industrial power load forecasting is outline in figure 3. The process starts with the collection of industrial power load data, which serves as the input time series for the forecasting model. This dataset contains temporal load information reflecting industrial power consumption patterns influenced by seasonality, trends, and irregular fluctuations.

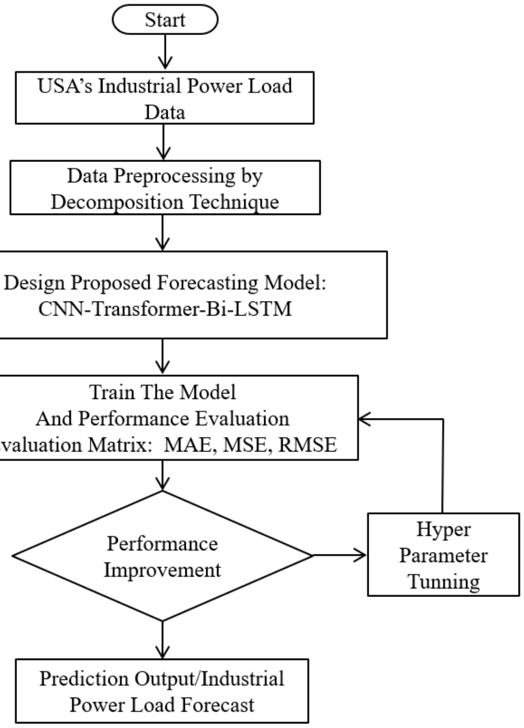


Figure 3: Flow diagram for the Methodology of Industrial Power Load forecasting by CNN-Transformer-BiLSTM model

In the data preprocessing phase, a decomposition technique such as seasonal trend decomposition technique using loess (STL) has been applied to separate the original load data into three distinct components: trend, seasonal, and residual. This step helps the model capture diverse temporal behavior more effectively and reduces the complexity of the forecasting task.

The proposed hybrid model integrates three deep learning architecture: convolution neural network (CNN), a transformer encoder and a bidirectional long-short term memory (Bi-LSTM). The CNN extract the feature to identify local temporal dependencies and short-term variations within the decomposed sequences. The transformer encoder is then employed to model long-range temporal dependencies through multi-head self-attention mechanisms, enhancing the representation of global patterns. In last stage, the Bi-LSTM captures bidirectional temporal dynamics, enhance the model to learn from both past and future dependencies in the time series.

In the final phase, 80% of the data was used to train the proposed CNN-Transformer-BiLSTM model and 20% data was used for testing. Model performance is evaluated using mean absolute error, mean squared error and root mean square error to ensure the forecasting accuracy and robustness. To locate the optimum parameter combination, hyperparameter tuning has been performed. By these techniques, the model has been optimized by tuning different parameter such as learning rate, number of filters, hidden units and attention heads to improve the overall predictive performance.

3.2 Architecture and Mathemathical Modeling of Proposed Model

The proposed CNN-Transformer-BiLSTM architecture integrates deep learning techniques to address the complex temporal variability, nonlinear dependencies, and multi-scale patterns inherent in industrial electricity load data. A details architecture of the CNN-Transformer-BiLSTM model has been shown in Figure 4.

In the first layer, the three decomposed components are concatenated to form a multi-feature input matrix, which is then fed into a Convolutional Neural Network (CNN). The CNN layer employs a 1-D convolutional filters to extract high level spatial-temporal features from the decomposed signals, enhancing local patten recognition. The effectiveness in time series analysis comes from their ability to extract meaningful local patterns over time. One dimensional (1D) convolutional kernel that move along the sequence, applying convolutional operations to small and overlapping segments. Each segment is transformed into an embedding vector, that is called "token", which captures important short-term features and trends within the time series [23]. Mathematically, the convolutional layer represent as:

$$y_t = \sigma\left(\sum_{i=0}^{k-1} (w_i . x_{t+i} + b)\right)$$

where, y_t = Output of convolutional operation, W_i = 1D convolutional filter (kernel) of size K, x_{t+i} = input time series of length T, $\boldsymbol{\sigma}$ = Activation function, b = Bias of output map

In the second layer, the feature maps extracted by the CNN then processed by a transformer encoder block, which is effectively captures long-range temporal dependencies and contextual relationships among features. The transformer architecture functions based on self-attention mechanism, allowing the model to assess the relative importance of each time step with respect to others in the sequence. By this mechanism, the network dynamically learns the global temporal interactions without relying on sequential recurrence.

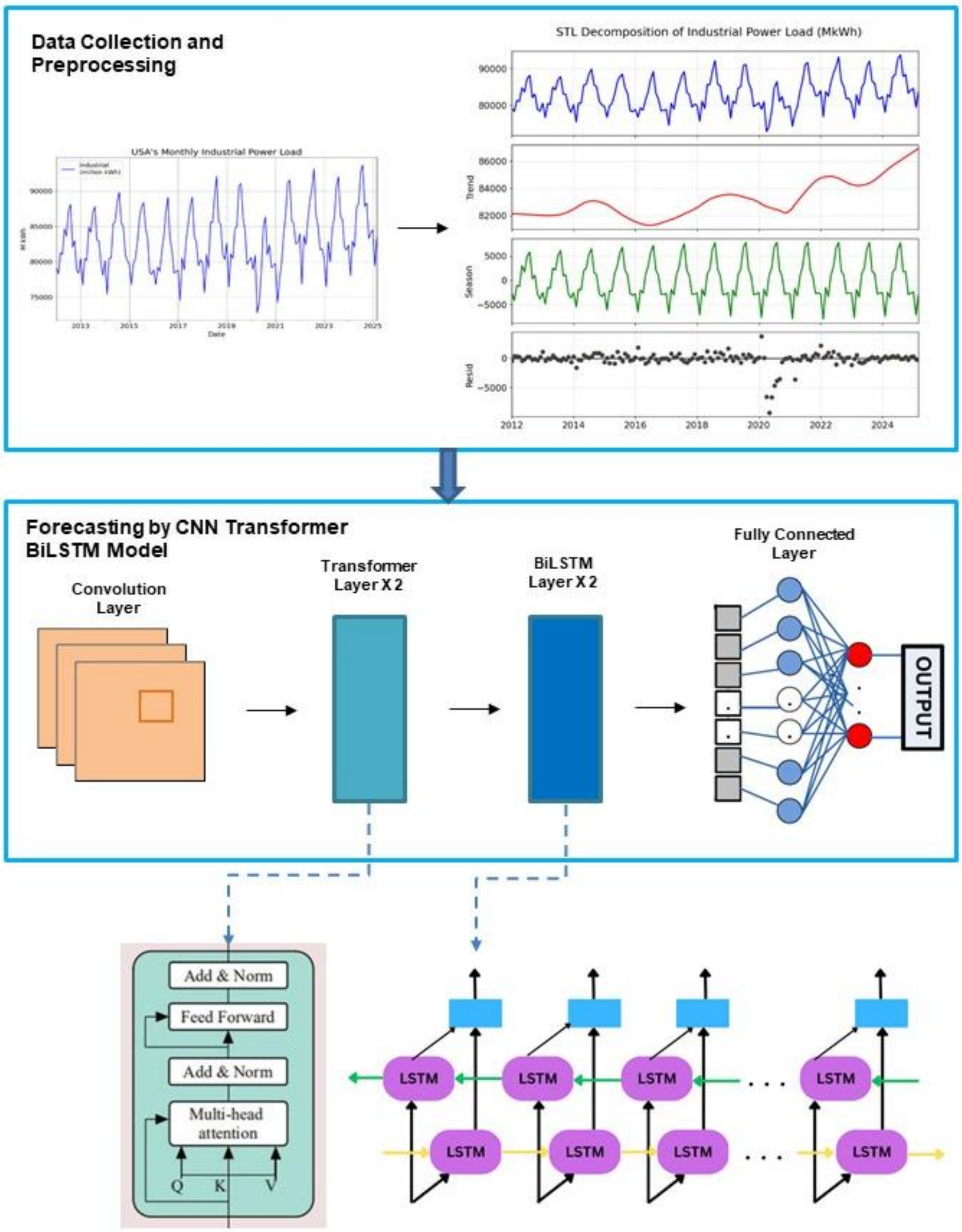


Figure 4: Seasonal-Trend decomposition using Loess (STL) technique

Mathematically, given an input sequence $U \in R^{T \times d}$ (where T denotes the number of time steps and d represents the feature dimension, with **U**= y_t being the CNN output), the Transformer encoder computes query, key, and value matrices for each attention head as $Q^{(h)} = UW_q^{(h)}$, $K^{(h)} = UW_k^{(h)}$, $V^{(h)} = UW_v^{(h)}$. The scaled dot-product attention is then applied to model inter-temporal relationships:

$$Attn^{(h)}(U) = softmax\left(\frac{Q^{(h)}K^{(h)T}}{\sqrt{d_k}}\right)V^{(h)}$$

The outputs of all attention heads are concatenated and projected using a linear transformation as

$$MHA(U) = [Attn^{(1)}, Attn^{(2)}, \ldots \ldots Attn^{(H)}]\, W_0$$

To improve training stability and prevent overfitting, a residual connection with dropout and layer normalization is applied, resulting in

$$U' = LayerNorm\,(U + Dropout\,(MHA\,(U)))$$

The resulting attention-refined features are then passed through a position-wise feed-forward network (FFN), defined as

$$FFN(U') = W_2\left(ReLU\,(W_1U' + b_1)\right) + b_2$$

Followed by another residual connection and normalization step to produce the final encoded representation:

$$TransformerBlock(U) = LayerNorm\,(U' + Dropout\,(FFN\,(U')))$$

Here, W_q,W_k,W_v,W_o,W_1,W_2 represent the learned weight matrices, b_1 and b_2 are bias terms, and d_k denotes the attention dimension. Through this hierarchical mechanism, the Transformer encoder effectively integrates both the local feature dependencies captured by the CNN and the global temporal relationships across the sequence. This attention-enhanced representation is then forwarded to the BiLSTM layer.

In the third layer, the attention refined feature sequence is passed through a bidirectional long short-term memory (BiLSTM) network. By processing data in both forward and backward temporal directions, temporal directions, the BiLSTM captures contextual dependencies from past and future time steps, improving temporal coherence in forecasting. By this mechanism, BiLSTMs enable the maintenance of important context across long sequences.

The mathematical operations governing the LSTM unit are as follows:

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f)$$
$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i)$$
$$C_t = \sigma(W_c x_t + U_c h_{t-1} + b_c)$$
$$c_t = f_t . c_{t-1} + i_i . C_t$$
$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_0)$$
$$h_t = o_t tanh(c_t)$$

where, σ = Activation function, f_t = Forget gate, i_t = Input gate, c_t = Cell State Update, o_t = Output Gate, $\tilde{C}_t$ = Cell state, x_t = Input vector at time t, h_{t-1} = Hidden state vector from the previous time step, c_{t-1} = Cell state vector from the previous time step, W and U = Learned weight matrices, b = Learned bias terms.

The output from the LSTM layers is concatenated to form the final representation:

$$h_t = [\overrightarrow{h_t}, \overleftarrow{h_t}]$$

where, $\overrightarrow{h_t}, \overleftarrow{h_t}$ denote the hidden states of the forward and backward LSTMs, respectively.

The BiLSTM output is then flattened and connected to a fully dense layer with linear activation to generate the final forecasted load values.

4. Simulation Results

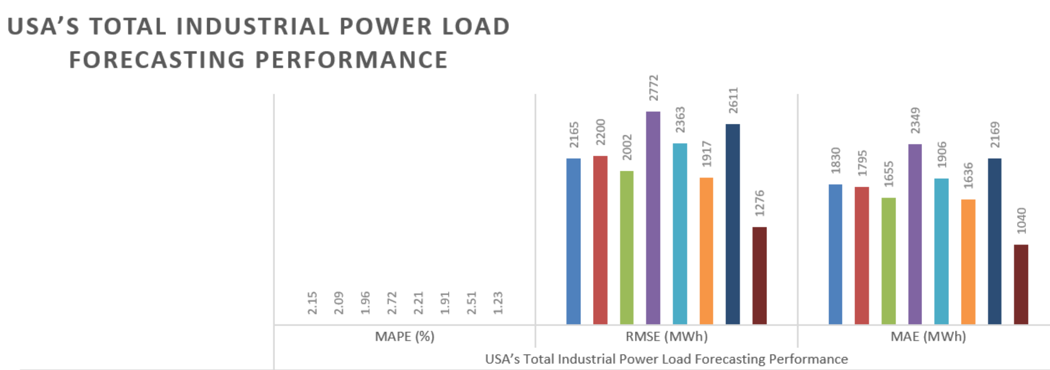
This section presents a comprehensive evaluation of the proposed CNN-Transformer-BiLSTM model to assess its forecasting capability and performance. A comparison analysis has been done between other deep learning model such as CNN, LSTM, BiLSTM, CNN-LSTM, with the proposed approach. To compare and measure forecasting performance across different models and scenarios, the evaluation metrics such as mean absolute percentage error (MAPE), root mean square error (RMSE) and mean absolute error (MAE) are employed.

4.1 Result Analysis of Proposed Model

The comparative forecasting performance of various deep learning models for total industrial power load is illustrated in Fig. 5. The results demonstrate that the proposed CNN-Transformer-BiLSTM hybrid model significantly outperformed all other deep learning and hybrid models in terms of all three metrics.

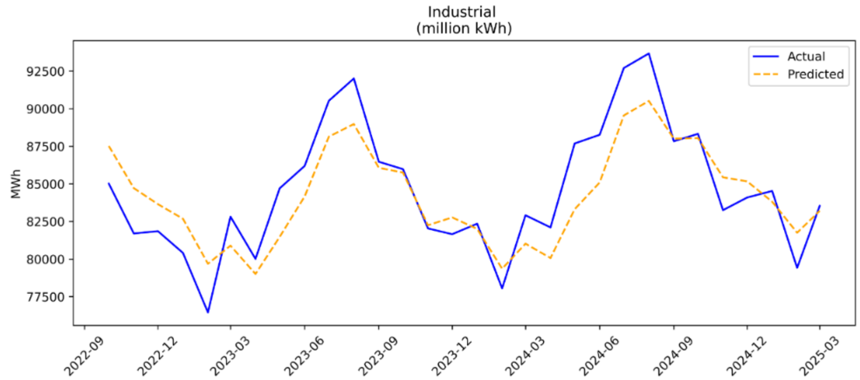
Traditional standalone models such as CNN, LSTM, and Bi-LSTM achieved moderate forecasting accuracy, with MAPE values of 2.15 %, 2.09 %, and 1.96 %, respectively. While hybrid models such as CNN-LSTM and CNN-BiLSTM improved temporal feature extraction, they still suffered from higher RMSE and MAE values, notably with the CNN-LSTM model showing an RMSE of 2772 MWh and an MAE of 2349 MWh. Incorporating bidirectional learning in CNN-BiLSTM-CNN and self-attention mechanisms in the Transformer-BiLSTM network improved learning of long-term dependencies, reducing RMSE to 1917 MWh and 2611 MWh, respectively.

However, the proposed CNN-Transformer-BiLSTM model achieved the lowest MAPE (1.23 %), RMSE (1276 MWh), and MAE (1040 MWh), indicating superior forecasting accuracy and generalization capability. This improvement can be attributed to the synergistic integration of convolutional layers for spatial–temporal pattern extraction, transformer layers for global attention-based feature refinement, and bidirectional LSTM layers for capturing long-range temporal dependencies in both forward and backward directions. These combined mechanisms enable the proposed model to achieve smoother convergence and higher predictive stability across dynamic load variations in industrial energy demand.

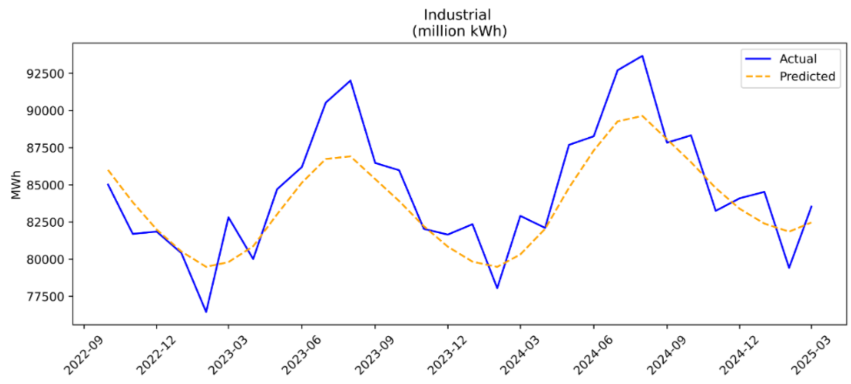


	MAPE (%)	RMSE (MWh)	MAE (MWh)
CNN	2.15	2165	1830
LSTM	2.09	2200	1795
Bi-LSTM	1.96	2002	1655
CNN-LSTM	2.72	2772	2349
CNN Bi-LSTM	2.21	2363	1906
CNN Bi-LSTM CNN	1.91	1917	1636
Transformer -BiLSTM	2.51	2611	2169
CNN-Transformer-BiLSTM (Proposed)	1.23	1276	1040

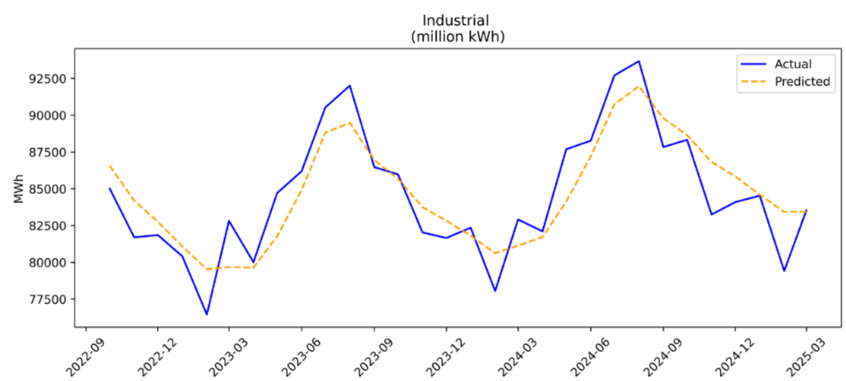
Figures 5 and 6 illustrate the comparison between the actual and predicted industrial power load using the proposed CNN-Transformer-BiLSTM model. The blue solid line represents the actual recorded values, while the orange dashed line indicates the predicted outputs. The close alignment between the two curves demonstrates that the proposed model effectively captures both the short-term fluctuations and long-term seasonal variations in industrial energy consumption. The prediction trend closely follows the actual data, indicating that the hybrid architecture successfully integrates local feature extraction (via CNN), temporal dependency modeling (via BiLSTM), and long-range correlation learning (via the Transformer). This strong agreement validates the model's superior forecasting accuracy and robustness in handling nonlinear and dynamic industrial power load patterns.



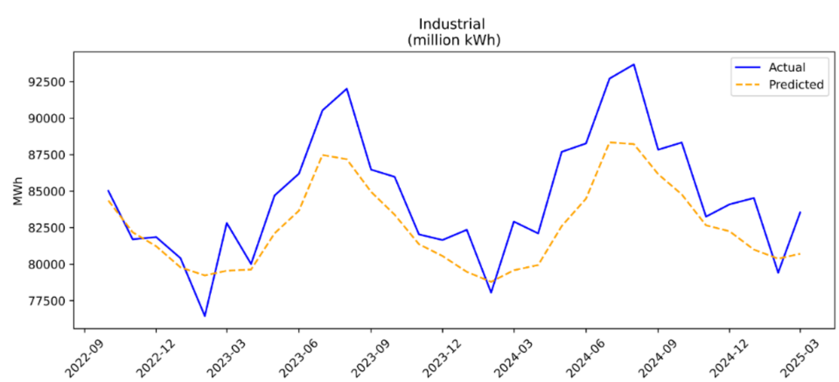
(a)



(b)

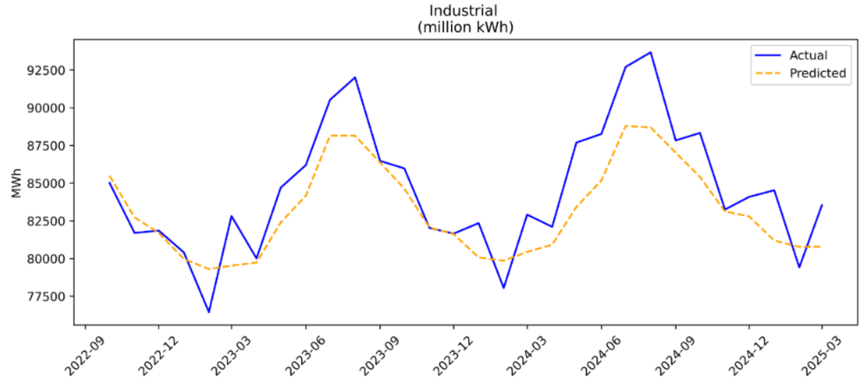


(c)

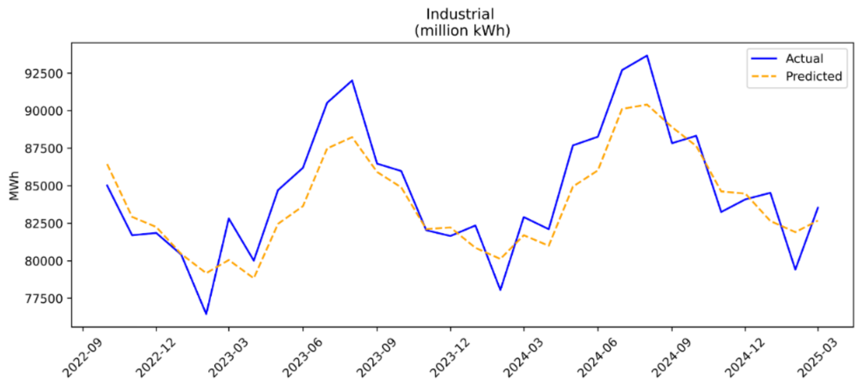


(d)

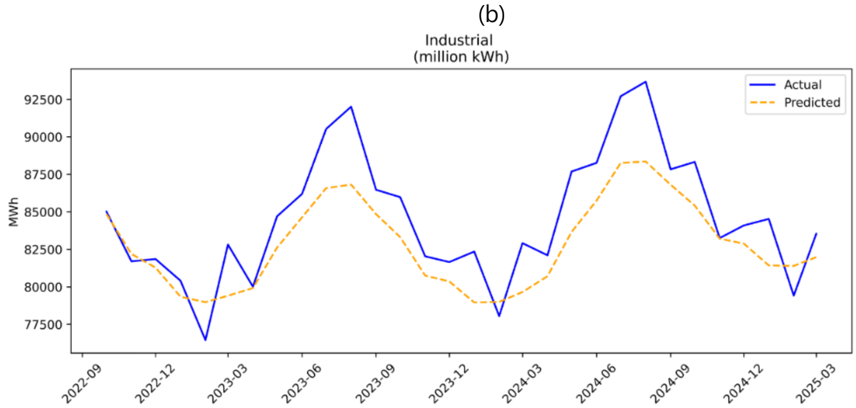
Figure 5: Actual Vs Predicted Curve by using model (a) CNN, (b) LSTM, (c) BiLSTM (d) CNN LSTM



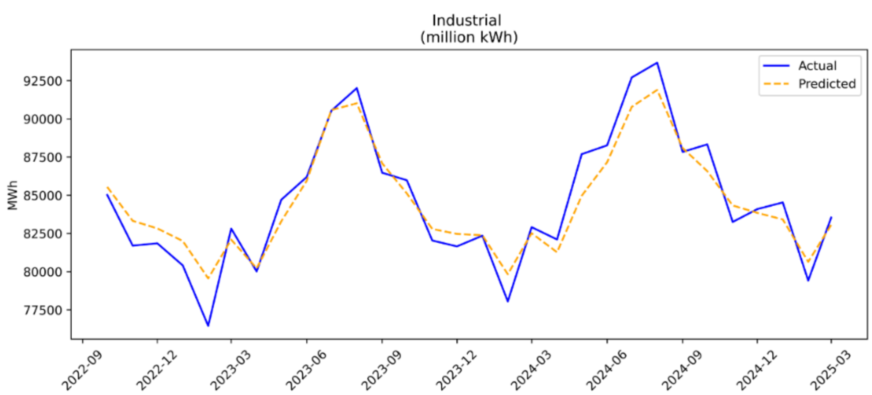
(a)



(b)



(c)



(d)

Figure 6: Actual Vs Predicted Curve by using model (a) CNN-BiLSTM, (b) CNN-BiLSTM-CNN, (c) Transformer-BiLSTM (d) CNN-Transformer-BiLSTM

4.2 Comparison with Previous Works in Application Point of View

In this section, a comprehensive comparison of hybrid deep learning based forecasting model for industrial power load application is presented. Recent research has explored diverse combinations of statistical model, neural networks, and optimization algorithms to enhance prediction reliability in complex industrial load environments. The Q-PSO-LSSVM [26] model, which integrates reinforcement learning with support vector machine, with a reported MEAN of 3% for short-term industrial load forecasting. Similarly, TCN-LightGBM [27] framework achieved a MAPE around 2-4% by combining convolutional networks with gradient boosting for feature extraction and temporal correlation modeling.

Table 1: Comparisons with the Previous Works

Model / Reference	Input Data Type	Application Domain	Reported Performance
Q-PSO-LSSVM [26]	15-min interval industrial load data (multiple industries, China)	Short-term industrial power load forecasting	Mean Absolute Percentage Error (MAPE) ≈ 3-5%
TCN-LightGBM [27]	Smart meter data with electrical, meteorological, and calendar features	Industrial load forecasting across multiple regions	Mean Absolute Percentage Error (MAPE) around 2-4%
MLP vs. Statistical Models [28]	Monthly industrial electricity data (Brazil, 1979–2020)	Industrial electricity demand forecasting	Mean Absolute Percentage Error (MAPE) ≈ 3.4%
FT-Transformer + GA [29]	Monthly industrial energy data (Botswana, Jwaneng Mine)	Industrial and commercial energy forecasting and optimization	Mean Absolute Error (MAE) of 3.03×10^5 kWh and Root Mean Square Error (RMSE) of 3.31×10^5 kWh
DCN–E-TCN [30]	15-second industrial load data (China)	Ultra-short-term industrial load prediction	35% lower RMSE and 30% higher R^2 and the maximum error is 9.4%
LSTM + SVR + XGBoost + RF [31]	Minute-level industrial load data (steel plant, China)	Industrial power demand forecasting	Mean Absolute Percentage Error (MAPE) below 1.5%
EEMD-LSTM [32]	Hourly industrial load data (Vietnam)	Short-term industrial load forecasting	Mean Absolute Percentage Error (MAPE) of approximately 1.3% for 1-step forecasting
MMDTOA-GRU Ensemble [33]	15-min interval industrial load data (steel plant, South Korea)	Industrial power demand forecasting	Normalized Mean Absolute Error (NMAE) as low as 3.684%
LCR-AdaBoost-FA-ELM [34]	Hourly industrial load data (furniture factory, China)	Day-ahead industrial load forecasting	Mean Absolute Percentage Error (MAPE) of 3.01%
TLP-CNN [35]	30-min interval industrial load data (China)	Short-term industrial load forecasting	Mean Absolute Percentage Error (MAPE) of 2.17%
Proposed STL+CNN-Transformer-BiLSTM	**USA's Industrial load data from EIA**	**USA's Industrial power demand forecasting**	**MAPE of 1.23%, RMSE of 1276 MWh, and MAE of 1040 MWh.**

A multi-layer perceptron (MLP) [28], achieved MAPE values near 3.4%, while the FT-Transformer-GA [29] and DCN-E-TCN [30] frameworks combine temporal attention and data completion networks to capture fine grained temporal dependencies, reporting Mean Absolute Error (MAE) of 3.03×10^3 kWh and Root Mean Square Error (RMSE) of 3.31×10^3 kWh, respectively. Hybrid ensemble methods such as LSTM+SVR+XGBoost [31] and EEMD-LSTM [32] achieved superior accuracy with MAPE 1.5% and 1.3% respectively. Other approaches, such as LCR-AdaBoost-FA-ELM [34], further optimized forecasting accuracy through adaptive feature learning, achieving exceptionally low MAPE values of approximately 0.31%. While these models collectively signify significant progress in industrial load forecasting, they focus on short-term or domain-constrained datasets such as steel plant or furniture factory loads.

In contrast, the proposed CNN-Transformer-BiLSTM framework introduces a generalized and multi-stage hybrid learning architecture designed to overcome these limitations. This integrated design enables the proposed model to provide robust and scalable forecasting across various industrial sectors and temporal horizons. When evaluated using USA's industrial load data from the EIA, the model achieved a MAPE of 1.23%, RMSE of 1276 MWh, and MAE of 1040 MWh, demonstrating its superior performance and generalization ability compared to existing state-of-the-art models.

5. Conclusion

The proposed CNN-Transformer-BiLSTM, hybrid deep learning model integrates convolutional layers for extracting local temporal patterns, transformer layers for capturing long range dependencies and bidirectional LSTM layers for capturing bidirectional temporal correlations. The proposed architecture combines the strengths of each component, enabling the model superiority and forecasting stability. According to the experimental evaluation, the model outperforms among all other deep learning and hybrid model compared, with a Mean Absolute Percentage Error (MAPE) of 1.23%, Root Mean Square Error (RMSE) of 1,276 MWh and Mean Absolute Error (MAE) of 1,040 MWh. The results represent a substantial improvement and highlighting the model's capability to capture complex nonlinear dependencies and seasonal variation in industrial power demand. The close alignment between the predicted and actual load profile curves validates the accuracy of the proposed approach. In future research, this framework will be extended to incorporate additional impacting factors such as weather and industrial operational hours data to enhance energy management and grid reliability.

Funding: This research received no external funding.
Conflicts of Interest: The authors declare no conflict of interest.
ORCID iD https://orcid.org/0009-0009-4052-9279

References

[1] U.S. Energy Information Administration (EIA), "Electricity Explained: Electricity in the United States," [Online] Available: https://www.eia.gov/energyexplained/electricity/

[2] Energy in the United States - Wikipedia, accessed September 22, 2025, Available: https://en.wikipedia.org/wiki/Energy_in_the_United_States

[3] U.S. Energy Information Administration (EIA), "After more than a decade of little change, U.S. electricity consumption is rising again" Available: https://www.eia.gov/todayinenergy/detail.php?id=65264

[4] Which Sources Will Dominate U.S. Electricity Use? (2024-2050) - Visual Capitalist, accessed September 22, 2025, Available: https://www.visualcapitalist.com/sp/tema-01-u-s-electricity-demand-by-source-2024-2050/

[5] G. P. Zhang, "Time series forecasting using a hybrid ARIMA and neural network model," Neurocomputing, vol. 50, pp. 159–175, 2003.

[6] P.C. Chang, C.Y. Fan, J.C. Hsieh, "A weighted evolving fuzzy neural network for electricity demand forecasting", 2009 First Asian Conference on Intelligent Information and Database System, vol.9, 2009, pp.330–350.

[7] Y. Xie, M. Li, "Application of gray forecasting model optimized by genetic algorithm in electricity demand forecasting", Second International Conference on Computer Modeling and Simulation, vol.10, 2010, pp.275–277.

[8] D.X. Niu, Y. Wei, J. Li, J. Wu, "Analysis of electricity demand forecasting in In-ner Mongolia based on gray Markov model", International Conference on E-Business and E-Government, vol.10, 2010, pp.5082–5085.

[9] A. Setaiwan, I. Koprinska, V.G. Agelidis, "Very short-term electricity load demand forecasting using support vector regression", International Joint Conference on Neural Network, Atlanta, Georgia, USA, vol.1, 2009, pp.2888–2894.

[10] Gigoni L, Betti A, Crisostomi E, Franco A, Tucci M, Bizzarri F, Mucci D (2017) "Day-ahead hourly forecasting of power generation from photovoltaic plants" IEEE Trans Sustain Energy 9(2):831–842

[11] Liu L, Zhan M, Bai Y (2019) "A recursive ensemble model for forecasting the power output of photovoltaic systems" Solar Energy 189:291–298

[12] Visser L, AlSkaif T, Van Sark W (2019) "Benchmark analysis of day ahead solar power forecasting techniques using weather predictions", 2019 IEEE 46th photovoltaic specialists conference (PVSC), pp 2111–2116

[13] H. Kong, D. Deng, W. Wu, "Short-term photovoltaic power forecasting based on LSTM neural network," in Proc. IEEE PES General Meeting, 2017, pp. 1–5.

[14] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997.

[15] A. Graves and J. Schmidhuber, "Framewise phoneme classification with bidirectional LSTM networks," in Proc. IEEE Int. Joint Conf. Neural Networks (IJCNN), 2005, pp. 2047–2052.

[16] H. Kong, D. Deng, W. Wu, "Short-term photovoltaic power forecasting based on LSTM neural network," in Proc. IEEE PES General Meeting, 2017, pp. 1–5.

[17] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997.

[18] A. Graves and J. Schmidhuber, "Framewise phoneme classification with bidirectional LSTM networks," in Proc. IEEE Int. Joint Conf. Neural Networks (IJCNN), 2005, pp. 2047–2052.

[19] Parvathareddy, S., Yahya, A., Amuhaya, L., Samikannu, R., & Suglo, R. S. (2025). A hybrid machine learning and optimization framework for energy forecasting and management. Energy Reports, 26, 103456.

[20] Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., & Zhang, W. (2021). Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting. Proceedings of the AAAI Conference on Artificial Intelligence, 35(12),11106-11115.

[21] Litao TANG, Zhiyong ZHANG, Jun CHEN, Linna XU, Jiachen ZHONG, Peisen YUAN. Research on Autoformer-based electricity load forecasting and analysis[J]. Journal of East China Normal University (Natural Science), 2023, 2023(5): 135-146.

[22] F. Saeed, A. Rehman, H. A. Shah, M. Diyan, J. Chen, and J. M. Kang, "SmartFormer: Graph-based transformer model for energy load forecasting," Sustainable Energy Technologies and Assessments, vol. 73, p. 104133, Jan. 2025, doi: 10.1016/J.SETA.2024.104133.

[23] A. B. A. Ferreira, J. B. Leite, and D. H. P. Salvadeo, "Power substation load forecasting using interpretable transformer-based temporal fusion neural networks," Electric Power Systems Research, vol. 238, p. 111169, Jan. 2025, doi: 10.1016/J.EPSR.2024.111169.

[24] Laiqi Zhao, Zhifeng Wu, Maoyu Du, "Improved Autuoformer Electricity load forecasting based on model fusion" Advances in Engineering Technology Research, Published: 2024-05-08, DOI: https://doi.org/10.56028/aetr.11.1.8.2024

[25] B. Lim, S. Arık, N. Loeff, and T. Pfister, "Temporal Fusion Transformers for interpretable multi-horizon time series forecasting," International Journal of Forecasting, vol. 37, no. 4, pp. 1748–1764, Oct. 2021, doi: 10.1016/J.IJFORECAST.2021.03.012.

[26] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, "Time Series Analysis: Forecasting and Control". in Wiley Series in Probability and Statistics. Wiley, 2015.

[27] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," Nature, vol. 323, no. 6088, pp. 533–536, Oct. 1986, doi: 10.1038/323533a0.

[28] J. Schmidhuber, "Deep learning in neural networks: An overview," Neural Networks, vol. 61, pp. 85–117, Jan. 2015, doi: 10.1016/j.neunet.2014.09.003.

[29] LeCun; Boser, B. E.; Denker, J. S.; Henderson, D.; Howard, R.; Hubbard, W. E.; Jackel, L. D. (1989). "Handwritten digit recognition with a back-propagation network". In Touretsky, David S. (ed.). Advances in Neural Information Processing Systems 2 (NIPS 1989).

[30] J. L. Elman, "Finding Structure in Time," Cognitive Science, vol. 14, no. 2, pp. 179–211, 1990, doi: 10.1207/s15516709cog1402_1.

[31] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," Neural Comput., vol. 9, no. 8, pp. 1735–1780, Nov. 1997, doi: 10.1162/neco.1997.9.8.1735.

[32] A. Graves and J. Schmidhuber, "Framewise phoneme classification with bidirectional LSTM and other neural network architectures," Neural Networks, vol. 18, no. 5, pp. 602–610, Jul. 2005, doi: 10.1016/j.neunet.2005.06.042.

[33] K. Cho, B. van Merriënboer, D. Bahdanau, and Y. Bengio, "On the Properties of Neural Machine Translation: Encoder–Decoder Approaches," in Proceedings of SSST-8, Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation, D. Wu, M. Carpuat, X. Carreras, and E. M. Vecchi, Eds., Doha, Qatar: Association for Computational Linguistics, Oct. 2014, pp. 103–111. doi: 10.3115/v1/W14-4012.

[34] D. Bahdanau, K. Cho, and Y. Bengio, "Neural Machine Translation by Jointly Learning to Align and Translate," May 19, 2016, arXiv: arXiv:1409.0473. doi: 10.48550/arXiv.1409.0473.

[35] C. Fan, S. Nie, L. Xiao, L. Yi, and G. Li, "Short-term industrial load forecasting based on error correction and hybrid ensemble learning," Energy and Buildings, vol. 313, p. 114261, Jun. 2024, doi: 10.1016/J.ENBUILD.2024.114261.

[36] R. Das (2025), "Artificial Intelligence Based Fault Detection and Predictive Analysis for Smart Power Systems", Western Carolina University

[37] A. Harris, R. Das and T. Kandil, "A Comparative Survey of AI-Driven Energy Demand Forecasting Across Diverse Sectors," *2025 57th North American Power Symposium (NAPS)*, Storrs, CT, USA, 2025, pp. 1-6, doi: 10.1109/NAPS66256.2025.11272390.

[38] Q. Ge et al., "Industrial Power Load Forecasting Method Based on Reinforcement Learning and PSO-LSSVM," IEEE Transactions on Cybernetics, vol. 52, no. 2, 2022, doi: 10.1109/TCYB.2020.2983871.

[39] Y. Wang et al., "Short-term load forecasting for industrial customers based on TCN-LightGBM," IEEE Transactions on Power Systems, vol. 36, no. 3, 2021, doi: 10.1109/TPWRS.2020.3028133.

[40] F. L. C. da Silva, K. da Costa, P. C. Rodrigues, R. Salas, and J. L. López-Gonzales, "Statistical and Artificial Neural Networks Models for Electricity Consumption Forecasting in the Brazilian Industrial Sector," Energies, vol. 15, no. 2, 2022, doi: 10.3390/en15020588.

[41] Sravani Parvathareddy, Abid Yahya, Lilian Amuhaya, Ravi Samikannu, Raymond Sogna Suglo, "A hybrid machine learning and optimization framework for energy forecasting and management", Results in Engineering, Volume 26, 2025, 105425, ISSN 2590-1230, https://doi.org/10.1016/j.rineng.2025.105425.

[42] H. Jiang, A. Wu, B. Wang, P. Xu, and G. Yao, "Industrial ultra-short-term load forecasting with data completion," IEEE Access, vol. 8, 2020, doi: 10.1109/ACCESS.2020.3017655.

[43] M. Tan, S. Yuan, S. Li, Y. Su, H. Li, and F. H. He, "Ultra-Short-Term Industrial Power Demand Forecasting Using LSTM Based Hybrid Ensemble Learning," IEEE Transactions on Power Systems, vol. 35, no. 4, 2020, doi: 10.1109/TPWRS.2019.2963109.

[44] N. N. V. Nhat, D. N. Huu, and T. N. T. Hoai, "Evaluating the EEMD-LSTM model for short-term forecasting of industrial power load: A case study in Vietnam," International Journal of Renewable Energy Development, vol. 12, no. 5, 2023, doi: 10.14710/ijred.2023.55078.

[45] C. Fan, S. Nie, L. Xiao, L. Yi, and G. Li, "Short-term industrial load forecasting based on error correction and hybrid ensemble learning," Energy and Buildings, vol. 313, p. 114261, Jun. 2024, doi: 10.1016/J.ENBUILD.2024.114261.

[46] Z. Zhu et al., "A day-ahead industrial load forecasting model using load change rate features and combining FA-ELM and the AdaBoost algorithm," Energy Reports, vol. 9, 2023, doi: 10.1016/j.egyr.2022.12.044.

[47] T. Walser and A. Sauer, "Typical load profile-supported convolutional neural network for short-term load forecasting in the industrial sector," Energy and AI, vol. 5, 2021, doi: 10.1016/j.egyai.2021.100104.