Journal of Building Engineering 103 (2025) 112055

Contents lists available at ScienceDirect

Journal of Building Engineering

journal homepage: www.elsevier.com/locate/jobe

Full length article

Dynamic energy cost prediction for industrial and commercial buildings using advanced machine learning and multi-pricing models

Qazi Waqas Khan [a,1], Anam Nawaz Khan [b,1], Bibi Misbah [a], Rashid Ahmad [c], Do Hyeun Kim [d,*]

[a] Department of Computer Engineering, Jeju National University, Jeju-si 63243, Jeju Special Self-Governing Province, Republic of Korea
[b] Department of Computer Engineering Bigdata Research Center, Jeju National University, Jeju 63243, Republic of Korea
[c] Department of Computing and IT Sohar University, Sultanate of Oman
[d] Department of Computer Engineering (and Advanced Technology Research Institute), Jeju National University, Jeju 63243, Republic of Korea

ARTICLE INFO

Keywords:
Energy price prediction
Lighting energy management
Exploratory data analysis
Deep learning

ABSTRACT

Accurate electricity energy cost prediction is essential for demand-side management, enabling dynamic pricing strategies and reducing market volatility. This study aims to develop a robust predictive framework for forecasting energy costs in industrial and commercial buildings, addressing challenges such as volatility, non-linearity, and seasonal variations. The framework integrates advanced Machine Learning (ML) models, including Temporal Fusion Transformer (TFT), Convolutional Bidirectional LSTM (Conv-BiLSTM), and Convolutional Bidirectional Gated Recurrent Units (Conv-BiGRU), with imputation methods Variational Autoencoder (VAE), adaptive K-Nearest Neighbor (KNN), and Generative Adversarial Imputation Networks (GAIN) and Hyperparameter Optimization (HPO). These hybrid models are designed to capture complex spatial and temporal dependencies in energy data. Validation using a real-world dataset from South Korean commercial and industrial buildings demonstrates that the Conv-BiLSTM model outperforms other models, achieving the lowest Mean Squared Error (MSE), Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE). Integrating VAE with HPO enhances predictive accuracy, addressing missing data and improving model robustness. Adaptive KNN also proves effective, particularly for industrial datasets, while GAIN shows moderate performance. Notably, the Time of use (ToU) pricing policy gives more accurate predictions than the uniform pricing policy. This study introduces a novel hybrid framework combining Conv-BiLSTM, VAE, and HPO, offering actionable insights for optimizing resource allocation and decision-making. The findings contribute significantly to sustainable energy management and informed decision-making in electricity markets.

1. Introduction

As renewable energy sources such as wind and solar become more prevalent in national electricity markets, new challenges arise in maintaining a balance between power supply and demand, controlling costs, and ensuring affordability. Consequently, there is

* Corresponding author.
E-mail address: kimdh@jejunu.ac.kr (D.H. Kim).

[1] Mr. Qazi Waqas Khan and Dr. Anam Nawaz Khan contributed equally in the paper.

https://doi.org/10.1016/j.jobe.2025.112055
Received 12 September 2024; Received in revised form 13 January 2025; Accepted 6 February 2025
Available online 22 February 2025

Nomenclature	
ML	Machine Learning
TFT	Temporal Fusion Transformer
Conv-BiLSTM	Convolutional Bidirectional LSTM
Conv-BiGRU	Convolutional Bidirectional Gated Recurrent Units
VAE	Variational Autoencoder
KNN	K Nearest Neighbor
EC	Electricity Cost
GAIN	Generative Adversarial Imputation Nets
DSM	Demand Side Management
ToU	Time of Use
CPP	Critical Peak Pricing
RTP	Real-Time Pricing
PRT	Peak Rebate Time
ARMA	Autoregressive Moving Average
GARCH	Autoregressive Conditional Heteroskedasticity
AI	Artificial Intelligence
DL	Deep Learning
CNN	Convolutional Neural Networks
LSTM	Long Short-Term Memory
RNN	Recurrent Neural Networks
DNN	Deep Neural Networks
ANN	Artificial Neural Networks
SDA	Stacked Denoising Auto Encoder
MARS	Multivariate Adaptive Regression Splines
SVM	Support Vector Machine
LASSO	Least Absolute Shrinkage and Selection Operator
GRU	Gated Recurrent Units
GPR	Gaussian Process Regression
RF	Random Forests
MLP	Multi-Layer Perceptrons
MAE	Mean Absolute Error
ANFIS	Adaptive Neuro-Fuzzy Inference System
MLR	Multivariate Linear Regression
GA	Genetic Algorithm
EDA	Exploratory Data Analysis
RMSE	Root Mean Square Error
MAPE	Mean Absolute Percentage Error
GRN	Gated Residual Network
MSE	Mean Squared Error
ELU	Exponential Linear Unit
GLU	Gated Linear Units
HPO	Hyperparameter Optimization

a need for effective demand response strategies that can accurately analyze customer demand patterns [1]. Dynamic Electricity Cost (EC), which periodically adjusts power rates, is already in use in several countries [2]. Prosumers can optimize their energy usage, earn rewards on electricity bills, and contribute to regional and national energy network stability by accurate EC predictions. Predicting electricity costs accurately and adjusting to market-driven movements are of the utmost importance for all market participants in the nationally competitive and ever-changing electricity market. Nevertheless, accurate forecasting is challenging due to its high volatility, nonlinear behavior, and absence of stable patterns [3]. Several factors contribute to this complexity, including electricity demand and supply dynamics, local weather conditions, and market management strategies.

The primary goal of Demand Side Management (DSM) is to encourage adjustments in power use in reaction to changes in prices, as highlighted by the U.S. Department of Energy [4]. Two main approaches can help to accomplish this: changing the activation times of some loads or lowering the power consumption of particular appliances [5]. Under the former, power scheduling systems control appliances are used to reduce consumption during periods of system stress. Task scheduling programs, on the other hand, change flexible loads under dynamic pricing policies, motivating consumers to lower their demand or change their power consumption to off-peak times. Apart from helping consumers save money, this dynamic pricing system greatly benefits power utilities. It helps lower the necessity to run expensive power plants during peak demand hours [6]. Several Asian countries, including Japan, Korea,

Table 1
Review of recent studies on predicting electricity prices (2020–2024).

Ref.	Year	Predictive model	Region	Data used
[25]	2021	LSTM Networks	Nord Pool	Time series of electricity prices
[26]	2021	GRU Networks	PJM, USA	Electricity price time series
[27]	2022	SSA-DELM	Denmark	Electricity prices, demand, renewable sources
[28]	2022	GPR and ANN	Multiple regions	Price, demand data, and market factors
[29]	2022	IDPSO-VMD-XGBoost	Greece	Prices, ambient temperature, humidity levels
[30]	2022	ERC-DNN	Nord Pool, EU	Electricity prices, renewable energy sources
[31]	2022	NARMAX Methods	Ireland	Prices, load data, CO2 emission figures
[32]	2022	ILRCN	Texas, USA	Pricing and load data
[33]	2023	Stacked Autoencoders	PJM, USA	Time series data on electricity prices
[34]	2023	Extreme Learning Machine	Multiple regions	Time series data on electricity prices
[35]	2023	ARIFMA-GARCH	Italy, Belgium	Prices, load, and renewable energy data
[36]	2023	ANN	Russia	Time series of electricity prices
[37]	2023	CNN-BiLSTM-AR	Nord Pool, EU	Time series of electricity prices
[38]	2023	STL-TCN-NBEATS	Spain	Prices, energy consumption, weather conditions
[39]	2023	SSA-NBEATS	Shanxi, China	Time series data on electricity prices
[40]	2023	Logistic Regression-CatBoost	Nord Pool, EU	Time series of electricity prices

and China, were the first to implement dynamic pricing in their power markets. Three prominent programs that are utilized are Time of Use (ToU), Critical Peak Pricing (CPP), and Real-Time Pricing (RTP). According to [7], there are a few additional pricing schemes, including Uniform, Seasonal, and Peak Rebate Time (PRT) approaches.

Energy consumers are incentivized to reduce their energy consumption or plan it during off-peak times through dynamic electricity pricing. There will be less need to activate pricey electricity-generating units because this method lowers overall usage during peak hours. This helps the pressure on power utilities. DSM and dynamic pricing depend greatly on accurate electricity cost estimates [8]. Accurate forecasts help utilities and consumers plan and respond appropriately to price changes. Predictive modeling helps one understand future pricing trends. Thus, dynamic pricing models can be changed in real-time, and DSM becomes more responsive and efficient [9]. These forecasts are vital for utilities in balancing expected demand with energy output, best use of resources, and preparation for integrating renewable energy sources [10]. Consumers can use these projections to deliberately run their systems to maximize savings and help maintain grid stability, particularly those producing electricity.

One of the most challenging aspects of electricity cost forecasting is the inherent volatility, short-term price surges, and shifting seasonality of EC [11]. This complexity has prompted extensive research. The three main types of EC prediction approaches are statistical, computational, and hybrid [12–14]. Traditional statistical techniques such as regression, Autoregressive Moving Average (ARMA), and Generalized Autoregressive Conditional Heteroskedasticity (GARCH) are effective for stable price trends. Still, they frequently struggle with EC's unpredictable and volatile nature.

Conversely, computational models driven by Artificial Intelligence (AI), including ML and Deep Learning (DL), exhibit enhanced capability in handling complex and non-linear patterns. DL methodologies such as Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM), and Recurrent Neural Networks (RNN) are exceptionally well-suited for EC forecasting due to their exceptional proficiency in sequential data processing [15–18]. Research shows that DL models outperform traditional and other ML models regarding accuracy and error rates. For instance, studies have found that Deep Neural Networks (DNN) with recent data have lower error rates than simpler models such as Artificial Neural Networks (ANN) [19]. Similarly, Wang et al. [20] discovered that a Stacked Denoising Auto Encoder (SDA) produced more accurate short-term hourly EP predictions than Multivariate Adaptive Regression Splines (MARS), Support Vector Machine (SVM), and Least Absolute Shrinkage and Selection Operator (LASSO).

Predictions made using a single model tend to be less effective due to the complexity of energy cost dynamics. Recent research has shown that hybrid models integrating various AI approaches are becoming more popular to enhance accuracy [21]. On average, hybrid models that combine different types of models, such as ARFIMA-ANN [22], ARMA-ELM [23], and ARMAX-LSSVM [24], have superior performance. The trade-offs include less transparency, higher computational requirements, and more complexity (see Table 1).

In a 2018 [41] developed a CNN–LSTM network specifically for short-term electricity price predictions, which outperformed competing models like SVM, Random Forests (RF), Multi-Layer Perceptrons (MLP), standalone CNN, and LSTM in terms of Mean Absolute Error (MAE). A paper [42] also used a similar CNN–LSTM model to predict EP in Iran's power market, showing a superior ability to capture the fluctuating and sinusoidal nature of EP time series compared to methods such as SVM, ANN, Adaptive Neuro-Fuzzy Inference System (ANFIS), and Multivariate Linear Regression (MLR). Although other models like ANN, ANN-GA, and ANFIS showed decent performance, the hybrid CNN–LSTM approach better represented the complex patterns of EP than more conventional MLR and SVM techniques. Nevertheless, due to their susceptibility to various influencing factors, these models often fail to create an effective management system between energy producers and consumers [43].

A thorough literature review and recent research suggest that hybrid models are more effective for time series forecasting. However, single-model approaches often fail to address the complexity of electricity energy costs. Hybrid models improve the prediction performance [21], and they can extract more accurate, distinct, and strong features from historical energy data. The existing studies focus on statistical and AI-driven models, but the gaps remain in integrating spatial and temporal data dependencies. In the literature, no study is performed to explore South Korea's industrial and commercial sectors, and there is a need for a method to deal with missing energy data and hyperparameter optimization to enhance the predictive performance of an energy cost.

This study aims to address these gaps by proposing new combinations of models, Convolutional Bidirectional Long Short-Term Memory (Conv-BiLSTM), Convolutional Bidirectional Gated Recurrent Unit (Conv-BiGRU), and Temporal Fusion Transformer (TFT) have been developed. Hybrid techniques, such as convolutional layers with bidirectional sequence modeling (Conv-BiLSTM and Conv-BiGRU), address the complexity of temporal and spatial dependencies in energy data. The Conv-BiLSTM model achieves lower MSE, MAE, RMSE, and MAPE values than the Conv-BiGRU and TFT models. The Variational Autoencoder (VAE) and adaptive K-Nearest Neighbor (KNN) imputation methods effectively address the challenges of missing data, which is essential in energy datasets. The VAE method efficiently handles the complex missing data patterns, improving the model's robustness and accuracy, particularly with Conv-BiLSTM. Bayesian Optimization tunes all learning models' hyperparameters, significantly enhancing their predictive performance. The novelty of a proposed method lies in the unique combination of an imputation method, deep learning models and hyperparameter optimization, which address the challenges of an energy cost prediction for industrial and commercial data.

This paper makes several key contributions to the field of electricity price forecasting in volatile markets:

- A new hybrid forecasting framework that integrates Temporal Fusion Transformer (TFT), Convolutional Bidirectional Long Short-Term Memory (Conv-BiLSTM), and Convolutional Bidirectional Gated Recurrent Unit (Conv-BiGRU) to capture spatial and temporal patterns.
- Implementation of fixed and dynamic pricing strategies for practical, real-world application.
- Exploratory data analysis for lighting energy consumption data to identify crucial patterns for enhancing prediction accuracy and reliability.
- Confirmation of the approach's validity and effectiveness through empirical validation using real data from the South Korean industrial and commercial electricity market, showcasing its applicability in diverse geographic contexts.
- Provision of substantial benefits to electricity market operators by improving energy management, resource allocation, and decision-making processes.

2. Research design and methodology

This section outlines the proposed methodology and details the various steps involved in the proposed predictive modeling framework. Fig. 1 outlines a structured approach for predicting energy cost using DL models, starting with initial energy data processing. In the first phase, data pre-processing, the raw energy data undergoes cleaning to check and fill in missing values using imputation techniques. GAIN, VAE and adaptive KNN methods are used to perform the imputation of missing values. Date–time information is formatted correctly for further analysis. After this, we perform data standardization and train the learning models, such as the Conv-BIGRU, Conv-BILSTM, and TFT models for energy cost prediction. Bayesian optimization is used to tune a learning model. This study performs the experiments with and without the imputation method and Bayesian Optimization to monitor the impact of imputation and parameter optimization. The performance of a learning model is evaluated using the Mean Square Error (MSE), Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) metrics.

2.1. Data preprocessing

Data imputation is crucial for handling missing or incomplete data, ensuring datasets are complete and reliable for analysis and modeling. Adaptive k-nearest Neighbors (KNN), Variational Autoencoders (VAE), and Generative Adversarial Imputation Nets (GAIN) are some of the advanced techniques for data imputation that are detailed here. This study used this method because of its ability to handle complex and high-dimensional datasets for accurate and realistic results. The adaptive KNN method adjusts its K parameter to identify the most relevant neighbors for imputing missing to preserve the local data pattern and ensure consistency with nearby observations. The VAE utilizes probabilistic modeling to infer missing values by learning latent data representations. This mechanism makes this method suitable for non-linear relationships and complex missing patterns. The GAIN method generates realistic imputations through an adversarial framework and models the distribution of missing data. These techniques perform better than the more straightforward imputation methods by maintaining the dataset's predictive performance and statistical integrity.

2.1.1. Variational Autoencoders (VAE)

Variational Autoencoders (VAE) handle incomplete or unprocessed energy data by encoding it into a more compact latent space and then reconstructing it to fill in missing values. This process involves compressing key features into a latent distribution. The encoded information is decoded back to its original form to complete the data, effectively bridging gaps. This method ensures data is fully restored while minimizing reconstruction errors.

2.1.2. Generative Adversarial Imputation Nets (GAIN)

Generative Adversarial Imputation Nets (GAIN) fill in missing data using a generative adversarial network technique. This approach comprises a discriminator and a generator. The generator generates approximations using partial data, a mask vector emphasizing missing points, and a noise vector. The role of the discriminator is to distinguish between original and filled-in data, hence guiding the refining of the generator's output. During training, the generator and discriminator engage in a minimax game, where the generator aims to generate realistic data imputations while the discriminator attempts to identify them.

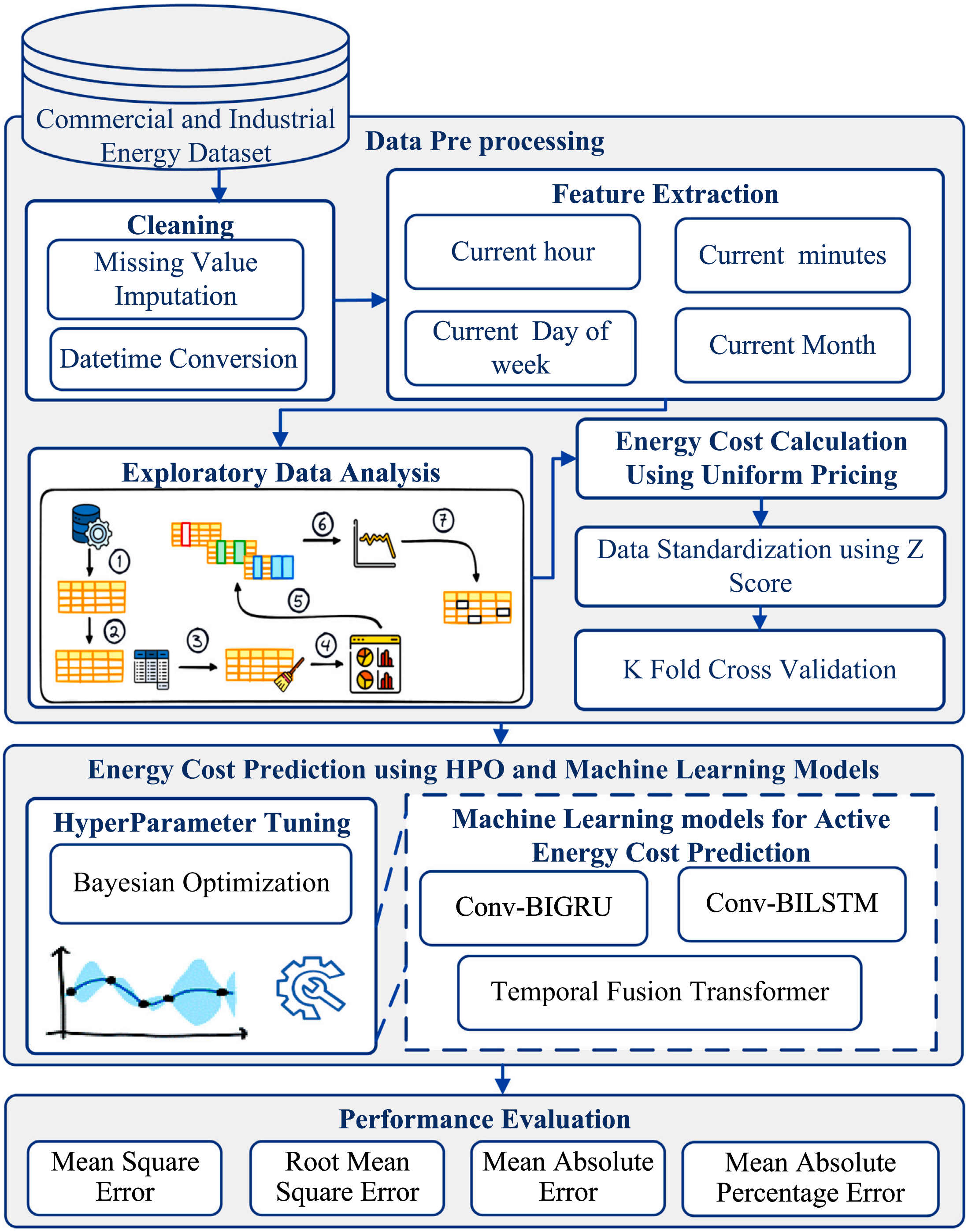


Fig. 1. Overview of the main architecture for energy cost prediction framework.

2.1.3. Adaptive KNN

Adaptive k-nearest Neighbors (KNN) identify the data points closely resembling the missing value, enabling an accurate estimation. The first phase in the technique involves placing the k-nearest neighbors for each point with missing values by computing the distance between data points. Next, the missing values are approximated using a statistical metric derived from the nearest neighboring data points, such as the median or mean. Adaptive KNN is flexible; this strategy enables it to be easily modified to match the specific needs of any dataset. The analysis and prediction of the application are enhanced by completing the dataset with these imputation approaches.

2.2. Feature engineering

In the next step, feature engineering extracts essential data points such as the present hour, minutes, day of the week, and month. Temporal analysis, which predicts energy cost patterns, depends on these characteristics. After feature extraction, exploratory and statistical data analysis is performed to find trends and patterns. This is crucial for understanding energy use patterns. Apart from Exploratory Data Analysis (EDA), the data is also standardized using Z-Score, and K-Fold cross-validation is applied to increase the robustness of the model.

2.3. Energy cost calculation

Further, the energy cost calculation considers the ToU and UP energy price policy. These price policies effectively simulate realistic and dynamic market scenarios in the energy cost framework. The ToU has a variability in energy costs based on the time of use, which reflects the real-time condition where energy prices fluctuate with demand. On the other hand, UP charges consumers based on energy use regardless of time. Utilizing these price policies in the energy cost prediction framework makes it adaptable to various market scenarios. It allows the stakeholders to evaluate energy consumption strategies under static and variable cost conditions. Integrating energy price policies and the advanced DL method enhances the framework's ability to simulate real-world energy cost scenarios, making the framework more reliable. It also helps to simulate realistic energy cost scenarios under dynamic pricing conditions [44]. The details of the uniform price policy are described subsequently.

2.3.1. Time-of-use (ToU) pricing

The Time-of-Use (ToU) pricing mechanism delineates rate variations based on elevated and reduced energy demand periods, categorizing time slots into off-peak, peak, and mid-peak periods. During peak hours, costs are substantially higher due to the increased reliance on peaking power plants, which are typically less efficient and more costly. These peaking plants often utilize diesel, gasoline, or natural gas, contrasting with more environmentally sustainable options like hydroelectric or nuclear power plants that serve base or mid-load demands [45]. The ToU framework addresses the immediate financial impacts of heightened energy usage and strategically plans to enhance transmission infrastructure and construct additional power plants to meet future demand surges. Elevated pricing during peak periods primarily incentivizes consumers to shift their energy usage to more cost-effective off-peak times. In addition to dynamic rate adjustments based on weather and seasonal fluctuations, ToU pricing establishes rates well in advance to aid consumer planning. As a relatively straightforward variable pricing model, this strategy alleviates prices effectively.

2.3.2. Uniform pricing policy (UP)

Uniform pricing refers to pricing structures where electricity costs remain constant regardless of variations in power use. This pricing structure lacks an incentive for users to modify their energy usage by fluctuations in overall demand [46]. As a result, users are protected from experiencing increased electricity energy costs due to unplanned or unnecessary usage. As a result, flat tariffs are commonly employed to prioritize the welfare of stakeholders by offering consistent and reliable invoicing. In this price strategy, the consumers are charged for energy usage regardless of time.

3. Prediction module

The primary objective is to build DL models that predict energy costs accurately. Sequential time-series data is processed using Conv-BiLSTM, TFT, and Conv-BiGRU models. Hyperparameter tuning is implemented through Bayesian Optimization to optimize model efficacy. Accuracy and reliability are evaluated using key performance metrics, including Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE). These procedures guarantee that the models generate energy price estimates that are both precise and consistent, which are indispensable for efficient energy management. This study uses the Conv-BiLSTM, TFT, and Conv-BiGRU models because they capture temporal dependencies and trends in lighting data. These advanced models effectively capture the short and long-term temporal patterns and interactions between features such as historical energy usage and dynamic pricing policies. The TFT model utilizes the attention mechanism to dynamically focus on the most relevant features, and the convolutional layers in the Conv-BiLSTM and Conv-BiGRU extract local temporal patterns effectively. At the same time, the bidirectional LSTM and GRU utilize the past and future context. The existing studies suggest these hybrid models outperform the simpler standalone models in terms of performance and robustness.

3.1. Temporal Fusion Transformer (TFT)

The TFT model is recognized as a powerful tool for analyzing temporal dynamics [47]. The TFT combines high-performance capabilities with advanced techniques for forecasting across multiple time horizons. This method integrates recurrent neural networks with interpretable self-attention layers to improve and streamline temporal processing. The network employs a selective attenuation strategy that filters out irrelevant data across all levels, enhancing the accuracy of feature extraction. Additionally, the adjustable convolutional network depth of the TFT makes it particularly effective for analyzing large-scale datasets, such as those related to energy consumption [48]. The gate functions in the TFT architecture reduce unnecessary components, thereby improving the model's ability to select relevant variables for each time segment. The algorithm efficiently captures short-term and long-term input correlations using encoders for static feature integration via context vector encoding. An interpretable multi-head attention

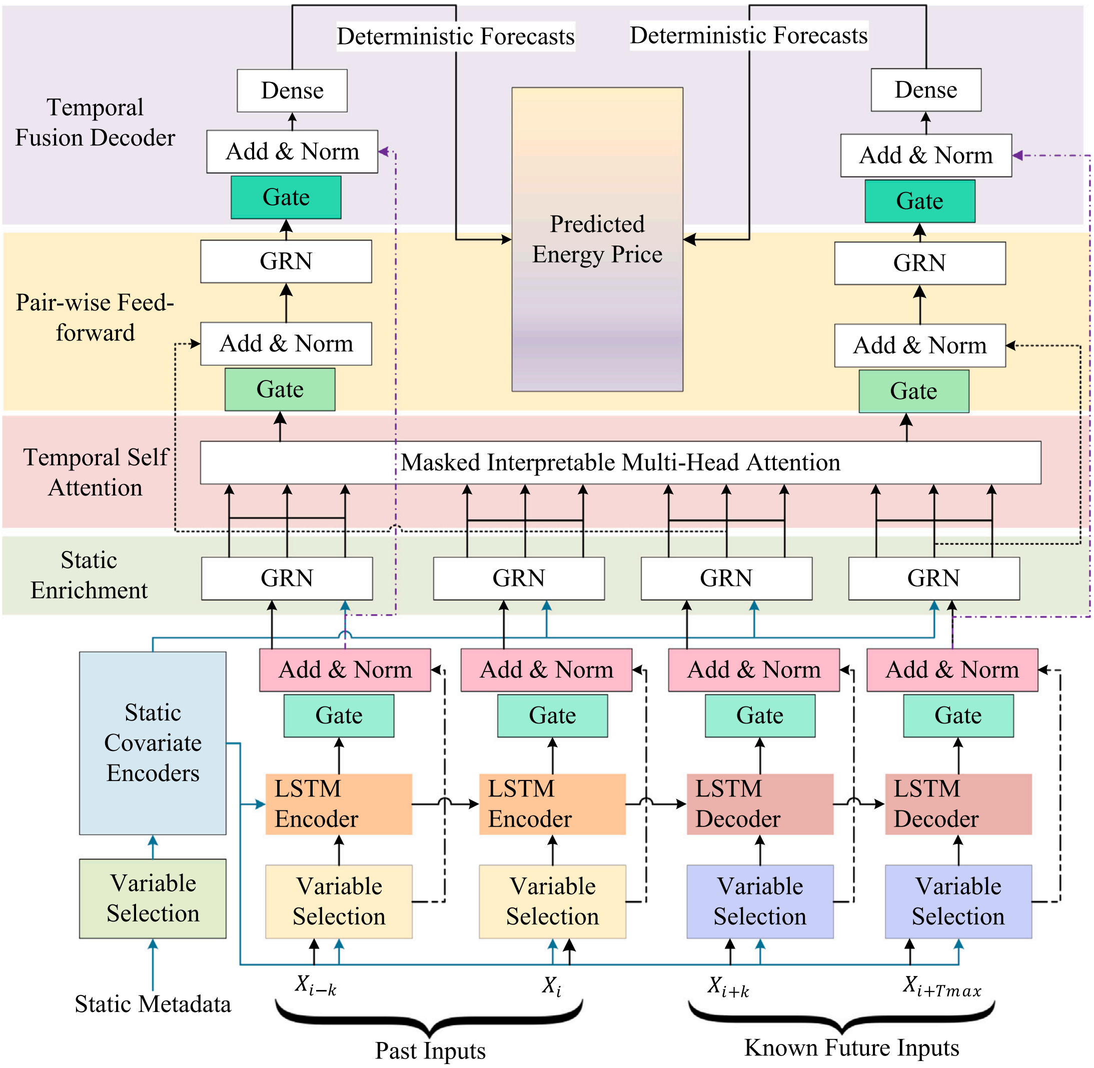


Fig. 2. Temporal Fusion Transformer.

mechanism manages long-term dependencies, while the sequence-to-sequence layer is responsible for localized processing [49]. Unlike traditional quantile forecasting methods, the TFT applies a deterministic forecasting approach using the MSE Loss function. The structure of the model is illustrated in Fig. 2.

Lim et al. [50] have thoroughly examined TFT, providing a comprehensive analysis of its functionality, intricate mathematical calculations for each component, and the complete model flow of the TFT. The mathematical expressions presented here are derived from the model. This paper has outlined the key points to give a concise overview of the functioning of the TFT.

3.1.1. Gated Residual Network (GRN)

The Gated Residual Network (GRN) enhances the model's adaptability by adeptly handling the inputs x and vector y. This is articulated through a series of operations defined as follows:

The core functionality of the GRN is encapsulated in the equation:

$$\mathrm{GRN}_\theta(x, y) = \mathrm{LayerNorm}(x + \mathrm{GLU}_\theta(\zeta_1)) \tag{1}$$

where ζ_1 and ζ_2 are intermediate variables described in subsequent equations. The calculation of ζ_1 is outlined as:

$$\zeta_1 = W_{1,\theta}\zeta_2 + b_{1,\theta} \tag{2}$$

The equation shows how ζ_2 is transformed using weighted summing and bias corrections.

The variable ζ_2 is generated through the activation function ELU, as shown in the equation:

$$\zeta_2 = \mathrm{ELU}(W_{2,\theta}x + W_{3,\theta}y + b_{2,\theta}) \tag{3}$$

At this point in the process, the model improves its ability to identify complex data patterns by integrating non-linear components. Using the Exponential Linear Unit (ELU) expedites convergence and mitigates the effects of vanishing gradients. Layer Normalization (Layer-Norm) stabilizes and expedites learning by normalizing outputs before activation. The network implements shared weights by indexing. Gated Linear Units (GLU) are essential in the GRN because they selectively omit extraneous network components, enabling the model to concentrate on the most significant features.

The GLU has the following formulation:

$$\mathrm{GLU}_{\theta}(z) = \sigma(W_{4,\theta} z + b_{4,\theta}) \otimes (W_{5,\theta} z + b_{5,\theta}) \tag{4}$$

The symbol σ denotes the sigmoid function, while the Hadamard product, which allows element-wise multiplication, is represented by $\otimes$. Solving this equation enables the internal states of the model to be dynamically adjusted according to the importance of the incoming data. Computational efficiency and model accuracy can be improved by effectively suppressing the non-linear effects of the GRN and reducing the GLU outputs to near-zero levels. It may be omitted if the GRN layer does not significantly contribute to the forecast.

3.2. Variable selection networks

The symbol σ denotes the sigmoid function; the Hadamard product, allowing element-wise multiplication, is marked $\otimes$. The performance of models is significantly enhanced by networks that focus the learning process on critical variables at each stage. These networks improve model performance by distinguishing between essential and irrelevant input variables. Linear transformations are applied to describe continuous variables, while entity embeddings, which provide detailed feature representations, are used to encode categorical variables. This approach is further defined using the following mathematical equations;

$$\upsilon_{Xt} = \mathrm{SoftmaxGRN}_{\upsilon X}(\Xi_t, c_s) \quad (7) \tag{5}$$

The equation allows the model to apply context-sensitive weights to variables through the use of c_s produced from a static covariate encoder and υ_{Xt} representing the vector of variable selection weights.

The feature vector is sent through the GRN in the following way for every variable j at time t:

$$\xi_{\sim}(j)_t = \mathrm{GRN}_{\xi_{\sim}}(j)(\xi(j)_t) \quad (8) \tag{6}$$

The transformed feature vector $\xi_{\sim}(j)_t$ is then combined across all variables to create a complete feature vector $\xi_{\sim t}$:

$$\xi_{\sim t} = \sum_{j=1}^{m_x} \upsilon_{jXt} \xi_{\sim}(j)_t \quad (9) \tag{7}$$

The formula assigns a weight to the jth element of the selection vector υ_{Xt} and is represented as υ_{jXt}. This aggregation technique demonstrates the model's ability to dynamically prioritize the most important variables throughout each forecasting interval, resulting in improved accuracy of predictions.

3.3. Temporal fusion decoder

The temporal fusion decoder comprises multiple specialized layers that each process time-based input uniquely.

3.3.1. Sequence-to-sequence layer

The initialization of the cell and hidden state is performed utilizing context vectors c_c and c_h through encoders and decoders in the LSTM architecture of this layer. Gated skip connections enhance the mathematical representation of this interaction.

$$\phi_{\sim}(t,n) = \mathrm{LayerNorm}(\xi_{\sim t} + n + \mathrm{GLU}\phi_{\sim}(t,n)) \quad (14) \tag{8}$$

where n spans a predefined range from $[-k, t_{max}]$.

3.3.2. Static enrichment layer

This layer enhances temporal features by incorporating static information, utilizing:

$$\theta(t,n) = \mathrm{GRN}_{\theta}(\phi_{\sim}(t,n), c_e) \quad (15) \tag{9}$$

The weights of GRN_{θ} are uniformly applied across layers, with c_e serving as the enhancing context vector.

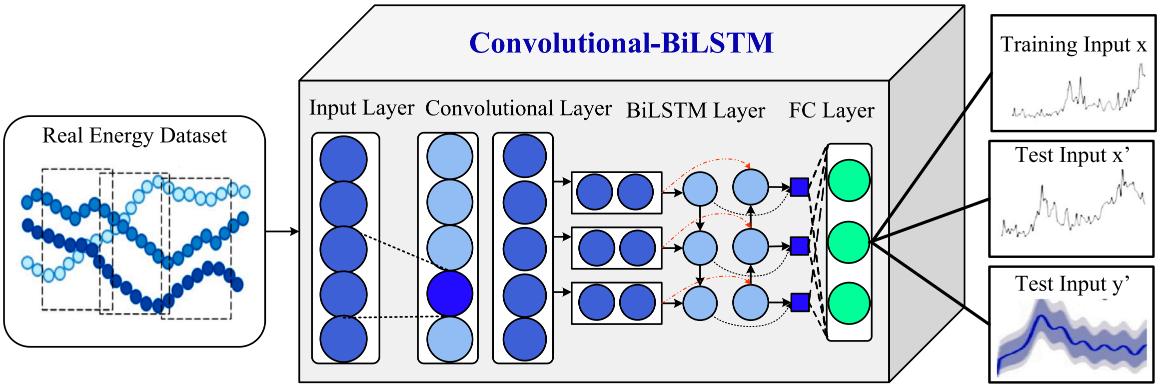


Fig. 3. Conv-BiLSTM architecture.

3.3.3. Temporal self-attention layer

Incorporating multi-head attention, this layer processes:

$$B(t) = \mathrm{IMH}(\Theta(t), \Theta(t), \Theta(t)) \quad (16) \tag{10}$$

This is where the enhanced temporal properties necessary for calculating B(t) are represented by $\Theta(t)$. Data leaking can be prevented by using decoder masking, which limits attention to earlier segments.

3.3.4. Position-wise feed-forward layer

As a means of consistency maintenance and efficiency enhancement, this layer employs Gated Residual Networks:

$$\psi(t, n) = \mathrm{GRN}_{\psi}(\delta(t, n)) \quad (18) \tag{11}$$

The layer that connects sequences is directly linked to by a gated residual connection:

$$\psi_{\sim}(t, n) = \mathrm{LayerNorm}(\phi_{\sim}(t, n) + \mathrm{GLU}_{\psi_{\sim}}(\psi(t, n))) \quad (19) \tag{12}$$

These layers strengthen the Temporal Fusion Transformer by improving the model's accuracy in handling and forecasting temporal sequences.

3.3.5. Convolutional Neural Networks with Bidirectional Long Short-Term Memory (Conv-BiLSTM)

Integrating CNN with Bidirectional BiLSTM creates a powerful Conv-BiLSTM model, enabling accurate predictions of electricity energy costs. This model excels at analyzing energy costs' characteristics and extracting complex information from sensor data, effectively capturing expected and unexpected patterns in energy prices (see Figs. 3 and 4).

The Conv-BiLSTM layer utilizes convolution operations to build spatial connections within the data by processing multiple inputs in its convolution layer. The mathematical expression is:

$$y_{ij} = f\left(b_j + \sum_{m=1}^{M} w_{mj} \cdot x_{i+m-1,j}\right) \tag{13}$$

Here, f represents the ReLU activation function, enhancing non-linearity; b_j is the bias; w_{mj} are the kernel weights; and x denotes the input features.

Following convolution, the pooling layer reduces the dimensionality and complexity of the data, thus aiding in managing computational costs and mitigating over-fitting:

$$p_{ij} = \max_{r \in R} y_{i \times T + r, j} \tag{14}$$

T denotes the stride, and R is the pooling window size, optimizing data pooling across spatial dimensions.

3.3.6. BiLSTM layer: Temporal pattern analysis

The bidirectional layer is a critical element of dependable time series forecasting. It enhances the model's understanding of temporal dynamics by simultaneously analyzing data sequences forward and backwards. The BiLSTM employs numerous gates, including the input gate, to regulate the passage of data throughout the network.

$$i_t = \sigma(W_{pi} p_t + W_{hi} h_{t-1} + W_{ci} c_{t-1} + b_i) \tag{15}$$

Forget Gate:

$$f_t = \sigma(W_{pf} p_t + W_{hf} h_{t-1} + W_{cf} c_{t-1} + b_f) \tag{16}$$

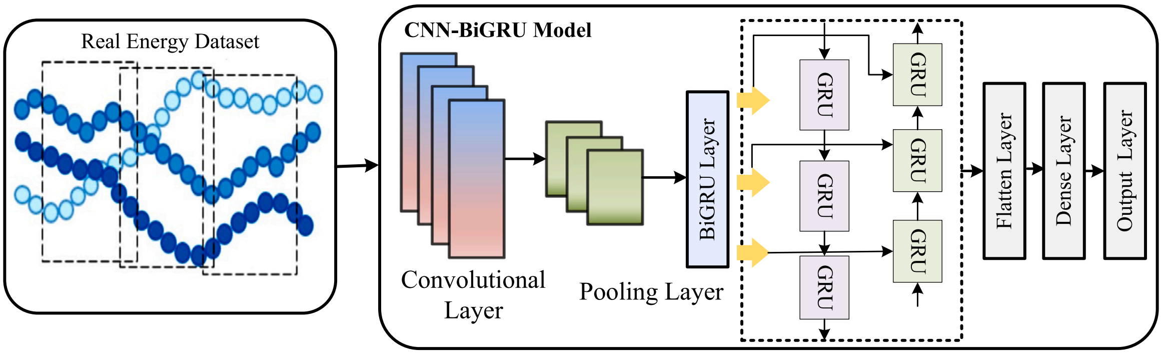


Fig. 4. Convolution Bidirectional Gate Recurrent Unit Architecture.

Table 2
System components and specifications.

S.no	Components	Detail
1	Hardware	Desktop PC Core(TM) i9 @ 2.50 GHz 2.50 GHz
3	Data file storage	MySQL Workbench
4	Programming Language	Python
5	Python required Libraries	Pandas, Scikit, Tensorflow, Seaborn, Matplotlib, pymysql, numpy, keras (for models, layers, optimizers, etc.), sklearn (model_selection, preprocessing, metrics), scipy (optimize), Adam
6	IDE	Spyder and VS Code

Output Gate:

$$o_t = \sigma(W_{po}p_t + W_{ho}h_{t-1} + W_{co}c_{t-1} + b_o) \quad (17)$$

The network depends on these gates to control the flow of incoming, outgoing, and stored data, ensuring precise predictions. The final stage in accurately predicting future energy expenses is consolidating all processed data into a dense feature vector by the fully linked layer.

3.4. *Convolution Bidirectional Gate Recurrent Unit Architecture (Conv-BiGRU)*

The Conv-BiGRU model combines the spatial feature extraction capabilities of CNN with GRU's temporal dependency modeling strength. This makes it highly effective for predicting energy costs. GRU utilize two main gates to identify patterns in sequential data over time: The update gate promotes the acquisition of inter dependencies by balancing old and new information.

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t] + b_z)$$

The reset gate helps the model decide how much past information to forget.

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t] + b_r)$$

Current memory content Combines new input with memory content.

$$\tilde{h}_t = \tanh(W \cdot [r_t * h_{t-1}, x_t] + b)$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t$$

This combination of CNN for robust spatial analysis and GRU for efficient temporal processing makes the CNN-GRU model a powerful tool for predicting dynamic patterns like energy cost, effectively capturing both the immediate and broader temporal trends.

4. Experimental set up, dataset and exploratory data analysis

The system setup outlined in Table 2 consists of a high-performance desktop equipped with an Intel Core i9 processor operating at 2.50 GHz, which is well-suited for training complex ML algorithms and handling large datasets. Python and its robust libraries, including NumPy, Pandas, scikit-learn, TensorFlow integrated with Keras, and many more, are used for programming activities, while MySQL Workbench is used for data administration. Visualization tasks are managed using Matplotlib and Seaborn. MySQL manages database connectivity, and coding primarily uses the Spyder and Visual Studio Code integrated development environments.

	A	B	C	D	E	F	G
1	GID	CHANNEL	ACTIVE_P	hour	minute	DAY_OF_\	MONTH
2	64257	1	146	0	0	3	9
3	64257	1	146	0	1	3	9
4	64257	1	147	0	2	3	9
5	64257	1	147	0	3	3	9
6	64257	1	147	0	4	3	9
7	64257	1	146	0	5	3	9
8	64257	1	147	0	6	3	9
9	64257	1	147	0	7	3	9
10	64257	1	146	0	8	3	9

Fig. 5. Dataset snapshot.

Table 3
Data attributes of energy smart lighting energy dataset.

Attribute	Data type	Description
GID	Integer	A unique identifier for the power grid.
CHANNEL	Integer	Channel number, representing different meters and sensors.
hour	Integer	Hour of the day when the energy usage was recorded (0-23).
minute	Integer	Minute of the hour when the energy usage was recorded (0-59).
Day_of_week	Integer	Day of the week when the data was recorded (1 = Monday, 7 = Sunday).
Month	Integer	Month of the year when the data was recorded (1-12).
Energy Cost	Float	Cost of energy consumed during the specific time, measured using varied price policies.

4.1. Dataset

The energy cost data facilitate an in-depth analysis of usage patterns influenced by temporal factors such as time of day, day of the week, and seasonal changes. This dataset comprises several key attributes that effectively capture and record detailed energy consumption metrics over specific periods, as shown in Table 3 and Fig. 5. Another dataset component is the (GID), a numerical label representing grid ID, linking each GID to specific residential settings. The dataset also features a CHANNEL attribute, which is critical in systems with multiple channels for monitoring light energy consumption in different areas within the premises of industrial and commercial buildings. This field indicates the specific channel or meter from which the data is gathered, facilitating detailed segmentation and analysis of energy usage.

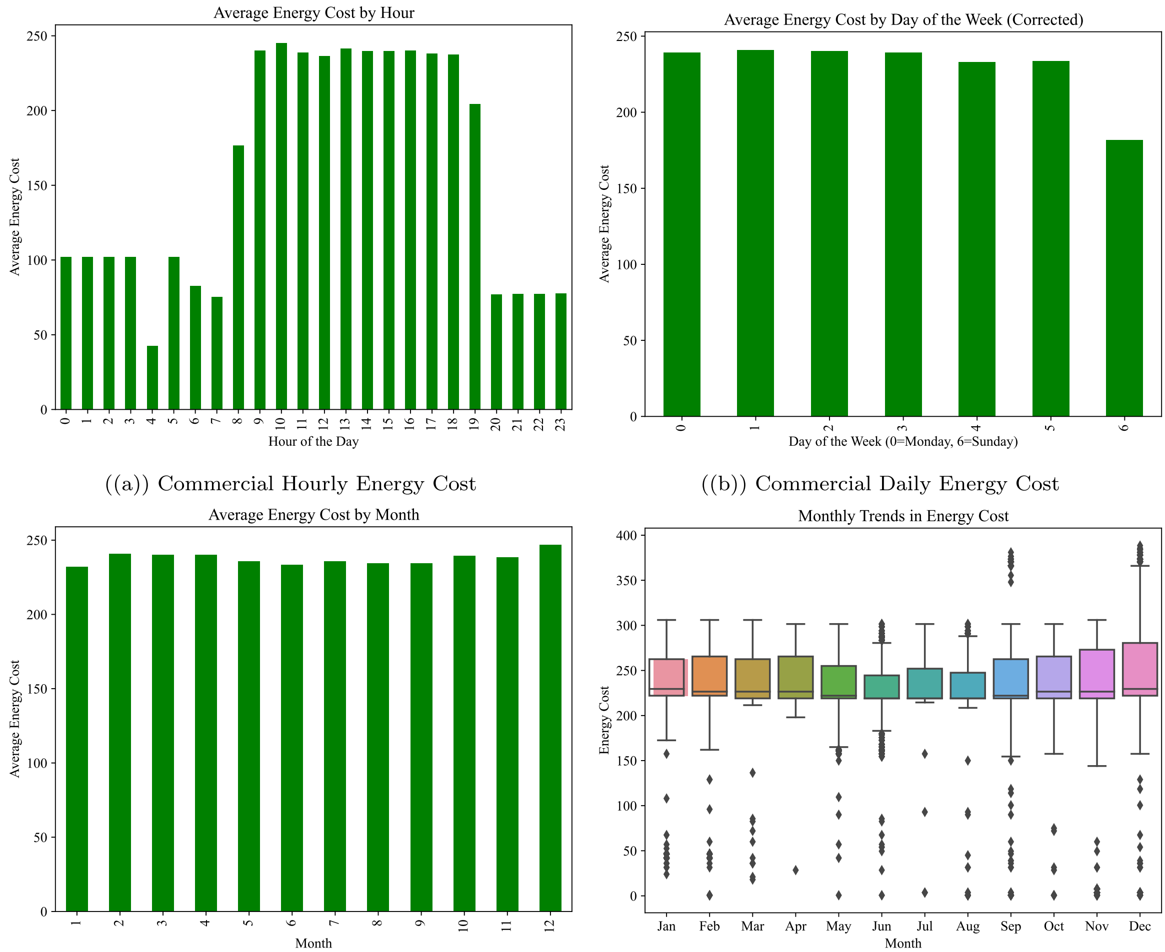
The attributes hour and minute provide a granular temporal resolution, essential for examining daily energy consumption trends. The hour attribute documents the time of day on a 24-hour scale. At the same time, the minute records the minute within that hour, offering the high-resolution data needed for accurate energy monitoring and analysis. The dataset includes day-of-week and monthly fields to help understand changes in energy use, such as the demand for lights in the winter. These variables provide a broader context for energy usage patterns, which help pinpoint variations in consumption on a weekly and seasonal basis. Because it measures the cost of energy utilized at each recorded period, typically represented in units of currency per Kilowatt-hour (Kwh), the energy cost characteristic is essential for economic analysis and billing operations. Integrating these data variables results in a comprehensive and valuable dataset that can be employed to analyze cost fluctuations throughout the day, identify trends across different days of the week, and monitor seasonal or monthly differences in energy usage. It is essential to have comprehensive and precise data to make informed decisions regarding energy management, optimize energy consumption, implement cost-effective strategies, and enhance overall energy efficiency. This dataset is utilized for operational decision-making, strategic planning and policy-making in energy management.

4.2. Exploratory data analysis

This section comprehensively reviews energy cost trends in industrial and commercial buildings. By examining energy use on an hourly, daily, and monthly scale, we can detect patterns and estimate peak times for energy consumption. This enables us to effectively regulate energy use while lowering operational expenses in both sectors.

4.2.1. Data analysis of commercial energy cost

Energy costs vary throughout the day, with certain hours being more expensive than others. This can be observed in Fig. 6(a), which shows the average cost for each hour of the day. Fig. 6(b) demonstrates that the average energy cost for commercial operations



((a)) Commercial Hourly Energy Cost

((b)) Commercial Daily Energy Cost

((c)) Commercial Monthly Energy Cost

((d)) Commercial Monthly Energy Cost Trends

Fig. 6. Commercial energy cost analysis.

remains relatively stable throughout the week, with no significant changes. According to the data presented in Fig. 6(c), energy prices exhibit consistent stability throughout the year, with minor fluctuations every month. This predictability enables businesses to forecast and allocate their energy budgets effectively. Fig. 6(d) illustrates consistent energy costs with negligible variations, suggesting that business energy expenses are easily anticipated for improved budgetary control. Fig. 7 depicts the average energy expenses over a week, showing that costs reach their highest point during the hours of operation (8 AM to 6 PM) on weekdays. This pattern highlights the influence of business activities on energy expenditures, offering valuable information for enhancing energy efficiency and minimizing costs during periods of high demand. The median energy costs are consistent throughout the year, as Fig. 6(d) illustrates. However, there are notable outliers and fluctuations, particularly in the colder and milder months. This data assists businesses in comprehending the monthly trends and prospective fluctuations in energy costs, facilitating more effective planning and budgeting for energy expenses. Energy prices are at their highest during business hours, on weekdays from 8 AM to 6 PM, as shown in Fig. 7. This pattern shows how commercial operations affect energy usage and prices, which help with peak-hour energy optimization and cost reduction.

4.2.2. Data analysis of industrial energy cost

Fig. 8(a) displays the data analysis of industrial energy cost, which reveals that energy costs are highest during business hours, specifically from 7 AM to 7 PM. This trend points to the necessity for optimizing energy consumption to save expenses during these peak hours, which is consistent with the high energy demand during daytime industrial activity. The average energy prices remain the same from Monday to Saturday, with a slight drop on Sunday, as shown in Fig. 8(b). This trend indicates that industrial processes use about the same amount of energy every day of the week, with Sunday seeing a slight decrease. According to Fig. 8(c), energy

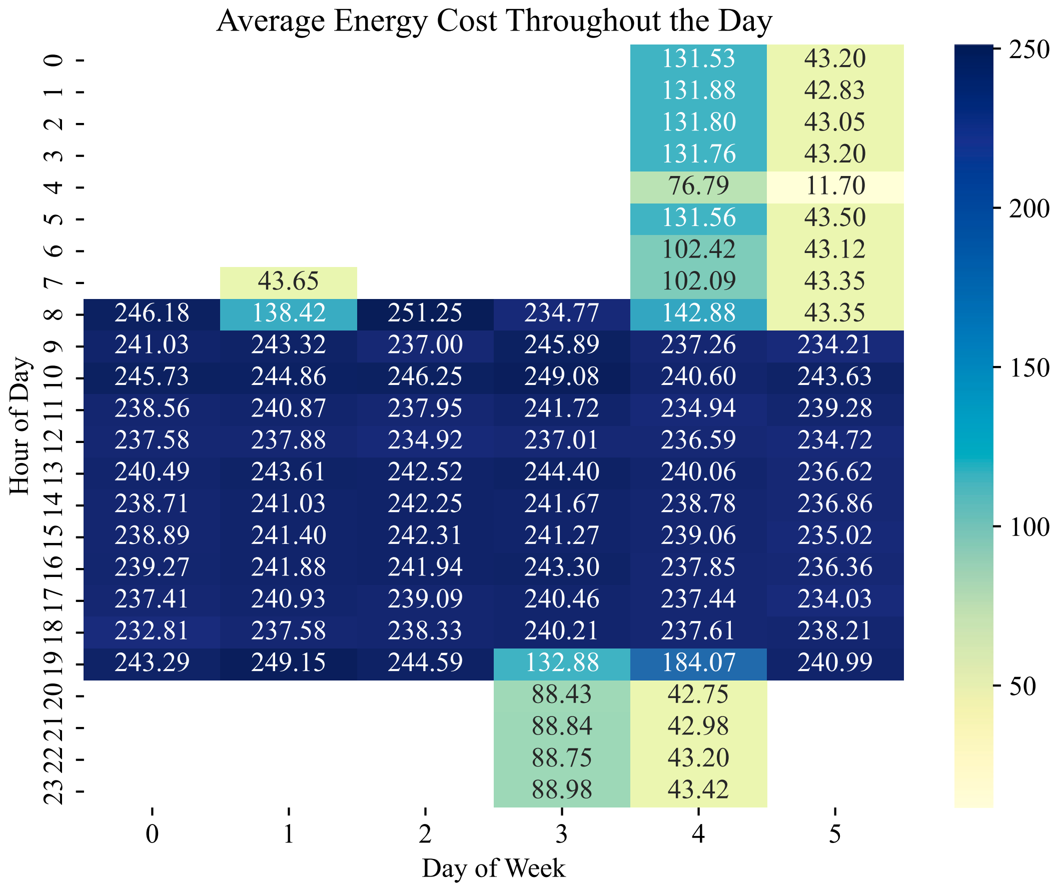


Fig. 7. Commercial average power usage throughout the day.

costs remain consistent year-round, with July showing clear peaks. Energy expenditures remain steady throughout the year, but this pattern indicates that industrial energy usage increases during the summer, perhaps due to increasing cooling demands.

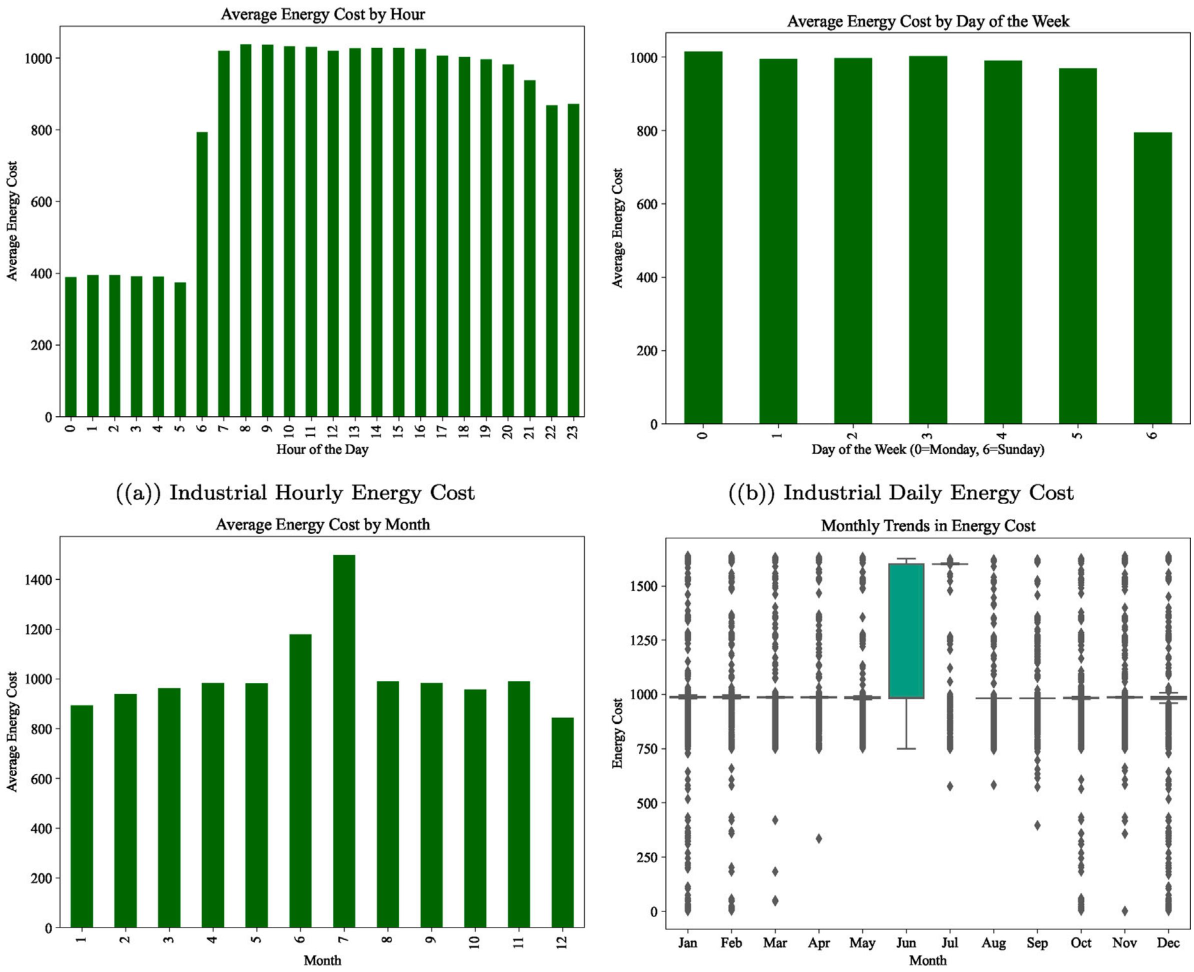
Fig. 8(d) illustrates that, although industrial energy costs remain relatively constant year-round, a notable spike and greater fluctuation characterize June. Better energy cost management may be gleaned from this trend, which shows that energy costs are more volatile in June because of higher demand or changes in operations. Industrial energy expenses reach their highest point during business hours, specifically on weekdays from 7 AM to 6 PM, as shown in Fig. 9. To emphasize the effect of industrial processes on energy consumption, the color gradient shows that costs are higher during these hours.

5. Experimental results and discussion

5.1. Experimental results of commercial energy cost prediction

This section presents the outcomes of our predictive models for commercial energy costs. The analysis highlights the accuracy and effectiveness of various forecasting models in predicting energy costs. The experimental results for commercial energy cost prediction, as shown in Tables 4, 5, and 6, reveal several key insights. Without imputation, the Conv-BiLSTM model achieved the lowest Mean Squared Error (MSE), Mean Absolute Percentage Error (MAPE), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE) compared to TFT and Conv-BiGRU. This suggests that Conv-BiLSTM handles raw data more effectively than the other models. TFT and Conv-BiGRU showed higher error metrics, indicating they might not be as robust as Conv-BiLSTM when dealing with unprocessed data.

Fig. 10 compares the MSE of three models, TFT, Conv-BiGRU, and Conv-BiLSTM, for estimating the energy cost for commercial lighting. The Conv-BiLSTM model has the lowest MSE, indicating improved prediction accuracy over the TFT and Conv-BiGRU models. All models exhibited considerable performance gains after using imputation approaches. When using adaptive KNN imputation, the Conv-BiLSTM model scored the lowest error metrics, demonstrating its successful use of imputed data. Interestingly, the performance gains are less pronounced using GAIN imputation, possibly because the dataset's features do not correspond well with GAIN's fundamental assumptions. While Conv-BiLSTM outperformed all other models, the VAE model showed the most



((a)) Industrial Hourly Energy Cost

((b)) Industrial Daily Energy Cost

((c)) Industrial Monthly Energy Cost

((d)) Industrial Monthly Energy Cost Trends

Fig. 8. Industrial energy cost analysis.

Table 4
Experimental results of commercial energy cost prediction without imputation.

Model	Uniform price				Time of use			
	MSE	MAPE	RMSE	MAE	MSE	MAPE	RMSE	MAE
TFT	0.753861	1.277966	0.868252	0.740075	0.745147	1.19645	0.863219	0.695813
Conv-BiGRU	0.769525	1.29321	0.877226	0.713146	0.73929	1.255895	0.85982	0.699295
Conv-BiLSTM	0.734317	1.198264	0.856923	0.688576	0.722237	1.224296	0.849845	0.684219

significant improvement using advanced imputation approaches, demonstrating its robustness and flexibility to complex imputation strategies.

In commercial settings, Fig. 11 evaluates the impact of three imputation methods, Adaptive KNN, GAIN, and VAE, on the MSE of energy cost forecasts. The Conv-BiLSTM model, with VAE imputation, achieves the lowest MSE, demonstrating superior handling of missing data in commercial datasets compared to other imputation techniques. Performance across all models is further enhanced through Bayesian HPO [51]. Optimizing model parameters proved critical, with Conv-BiLSTM achieving the lowest MSE and MAE when combined with adaptive KNN and HPO. While VAE combined with HPO improved the performance of all models, Conv-BiLSTM consistently performed best overall. These findings underscore the potential of VAE and HPO to enhance the model's predictive performance significantly.

Fig. 12 highlights the benefits of combining imputation methods with HPO. For commercial datasets, the Conv-BiLSTM model with VAE imputation and HPO yields the lowest MSE, significantly improving predictive accuracy compared to models without HPO.

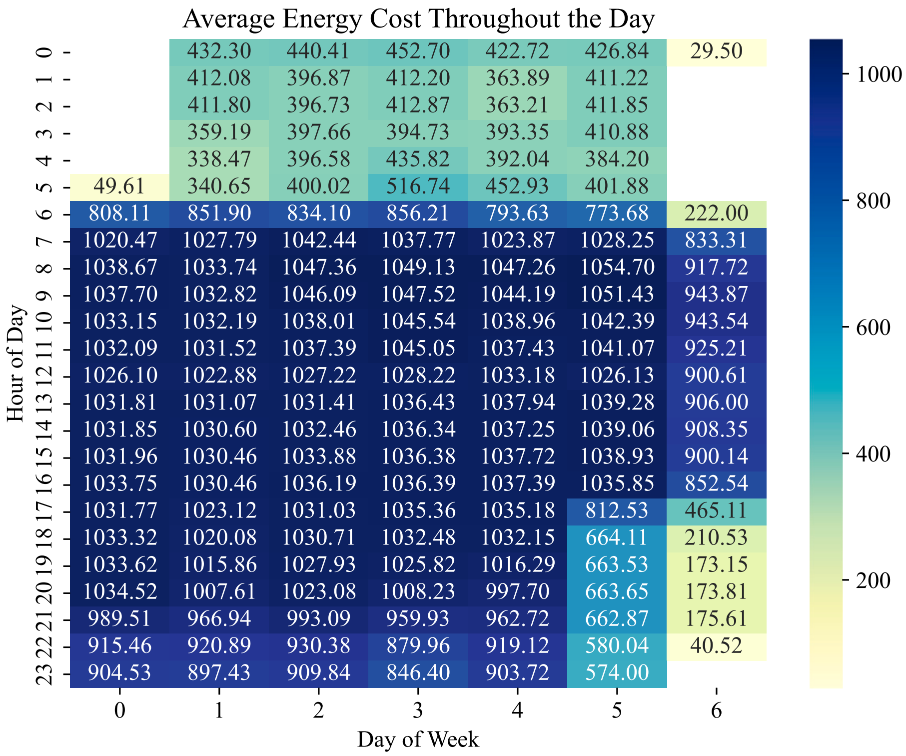


Fig. 9. Industrial average power usage throughout the day.

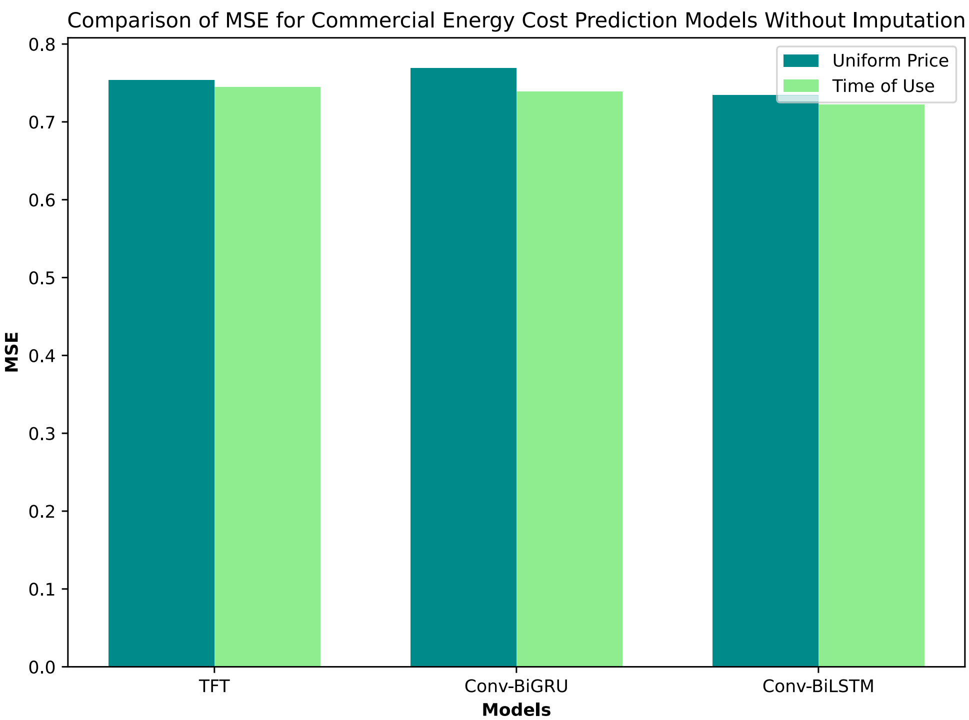


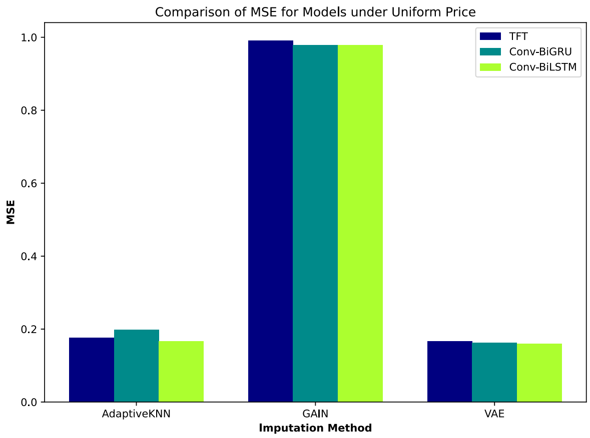
Fig. 10. Models comparison for commercial light data without imputation.

Table 5
Experimental results of commercial energy cost prediction with imputation.

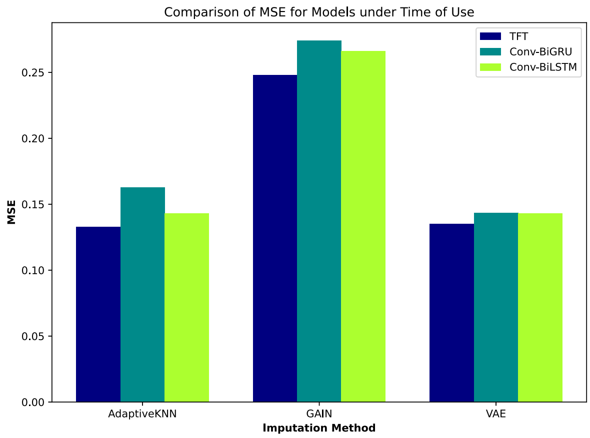
Model	Uniform				Time of use			
	MSE	MAPE	RMSE	MAE	MSE	MAPE	RMSE	MAE
adaptiveKNN								
TFT	0.176058	0.888214	0.419593	0.309227	0.132707	2.005181	0.36429	0.274363
Conv-BiGRU	0.198079	0.988632	0.44506	0.335175	0.162841	2.037851	0.403536	0.306988
Conv-BiLSTM	0.166716	0.91428	0.408309	0.305146	0.142906	1.977378	0.378029	0.285108
GAIN								
TFT	0.990614	0.9856	0.995296	0.646114	0.247915	1.05201	0.49791	0.213275
Conv-BiGRU	0.978348	0.910302	0.99063	0.631795	0.273984	1.35622	0.523435	0.356621
Conv-BiLSTM	0.978948	0.916183	0.989418	0.63247	0.266016	1.260626	0.515768	0.345547
VAE								
TFT	0.16686	0.243142	0.408485	0.219936	0.135197	0.260772	0.367692	0.192868
Conv-BiGRU	0.161948	0.243758	0.402428	0.219627	0.143461	0.274198	0.378762	0.212145
Conv-BiLSTM	0.158966	0.241804	0.398705	0.218111	0.143007	0.294514	0.378162	0.211025

Table 6
Experimental results of commercial energy cost prediction with imputation and HPO.

Model	Uniform				ToU			
	MSE	MAPE	RMSE	MAE	MSE	MAPE	RMSE	MAE
adaptiveKNN								
TFT	0.135269	0.810255	0.367789	0.27847	0.107749	1.936732	0.328251	0.243842
Conv-BiGRU	0.166631	0.818121	0.408205	0.297721	0.145069	2.015339	0.380879	0.287973
Conv-BiLSTM	0.140694	0.760222	0.375092	0.274097	0.119988	2.149662	0.346392	0.258196
GAIN								
TFT	0.97913	0.962217	0.98951	0.642266	0.251415	1.268718	0.501343	0.344027
Conv-BiGRU	0.975581	0.929704	0.987715	0.634443	0.279107	1.226041	0.528306	0.345103
Conv-BiLSTM	0.974399	0.908136	0.987117	0.62721	0.264489	1.183515	0.514285	0.340637
VAE								
TFT	0.16553	0.256709	0.406854	0.224577	0.135876	0.279067	0.366544	0.200537
Conv-BiGRU	0.169022	0.243321	0.401151	0.220262	0.134352	0.274744	0.366541	0.200553
Conv-BiLSTM	0.154126	0.242766	0.392588	0.218508	0.127685	0.252587	0.35733	0.187543

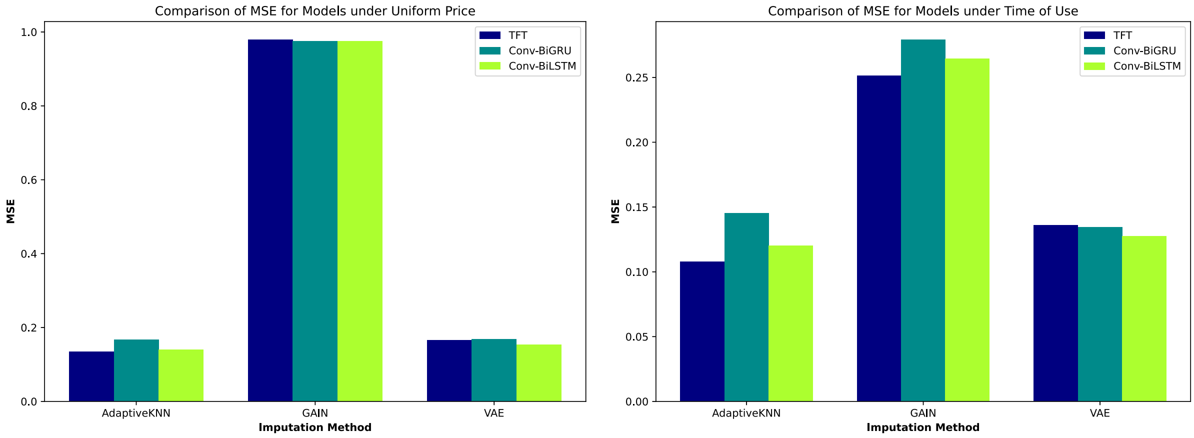


((a)) Uniform



((b)) Time of Use

Fig. 11. Models comparison for commercial light data with imputation.



((a)) Uniform   ((b)) Time of Use

Fig. 12. Models comparison for commercial light data with imputation and HPO.

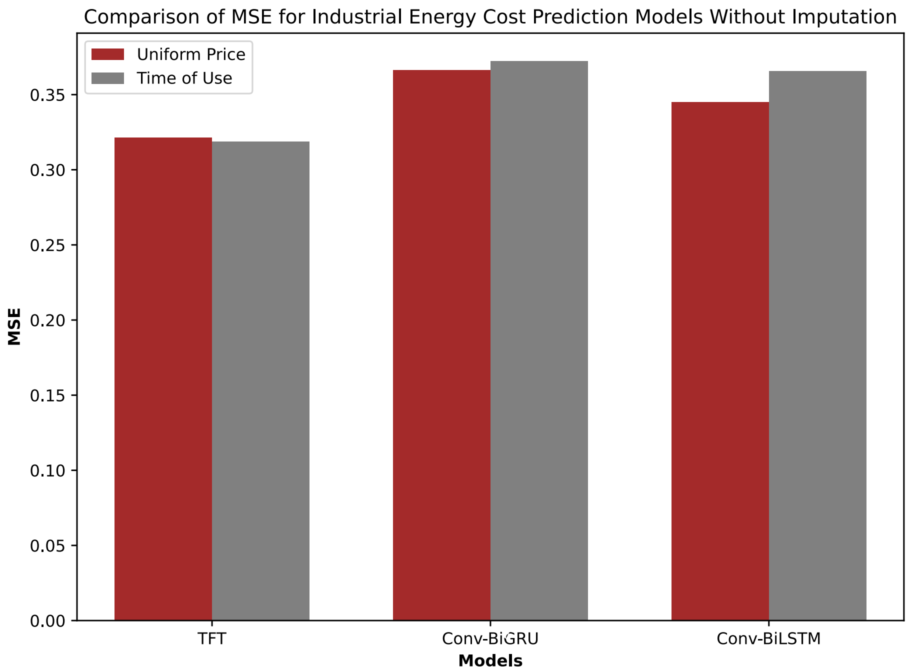


Fig. 13. Models comparison for industrial light data without imputation.

Table 7
Experimental results of industrial energy cost prediction without imputation.

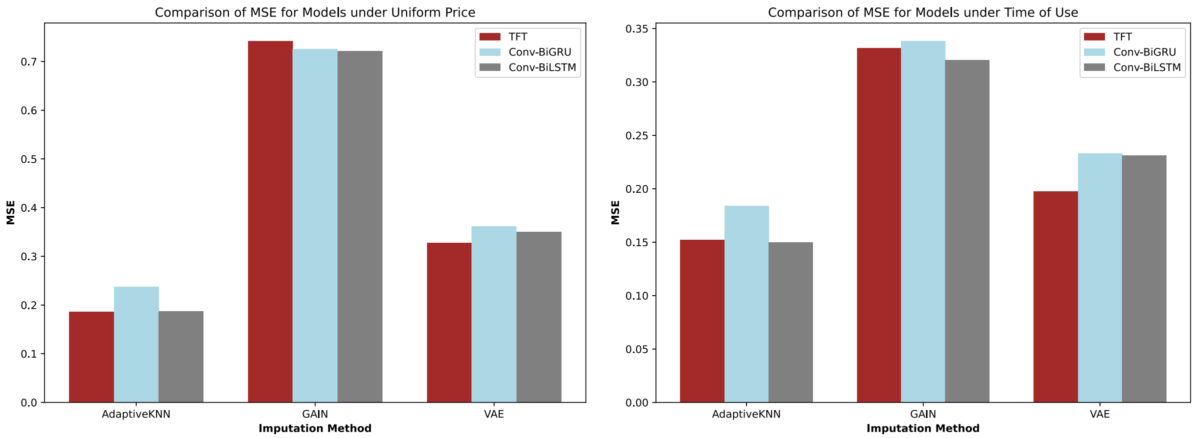
Model	Uniform price				Time of use			
	MSE	MAPE	RMSE	MAE	MSE	MAPE	RMSE	MAE
TFT	0.321269	25.83922	0.568606	0.302467	0.318574	22.24178	0.564423	0.283111
Conv-BiGRU	0.366288	29.11779	0.605218	0.332156	0.372301	29.44156	0.610165	0.344061
Conv-BiLSTM	0.344909	26.6159	0.58729	0.316294	0.365489	27.76749	0.604557	0.329517

5.2. Experimental results of industrial energy cost prediction

This section presents the outcomes of our predictive models for commercial energy costs. The experimental results for industrial energy cost prediction, presented in this section, offer insights into the model's performance in an industrial context.

Table 8
Experimental results of industrial energy cost prediction with imputation.

Model	Uniform				Time of use			
	MSE	MAPE	RMSE	MAE	MSE	MAPE	RMSE	MAE
adaptiveKNN								
TFT	0.185638	0.829974	0.430857	0.274245	0.152188	1.19349	0.390112	0.269254
Conv-BiGRU	0.237533	1.035152	0.487374	0.332797	0.184051	1.4248	0.429012	0.314289
Conv-BiLSTM	0.186837	0.881255	0.432247	0.274533	0.149687	1.417765	0.386894	0.275115
GAIN								
TFT	0.742023	1.675538	0.850896	0.430044	0.331614	0.999451	0.575859	0.341289
Conv-BiGRU	0.725615	1.988786	0.85183	0.459527	0.338259	1.189565	0.5816	0.305878
Conv-BiLSTM	0.721806	1.835564	0.849592	0.448824	0.320327	0.952734	0.565974	0.325707
VAE								
TFT	0.327792	22.16235	0.572531	0.290899	0.197383	0.645836	0.444278	0.25155
Conv-BiGRU	0.361775	31.54827	0.601478	0.339165	0.233134	0.896053	0.482819	0.315589
Conv-BiLSTM	0.350214	28.74828	0.591789	0.325306	0.231233	0.90388	0.480867	0.321036



((a)) Uniform ((b)) Time of Use

Fig. 14. Models comparison for industrial light data with imputation.

Tables 7, 8, and 9, which contain the experimental results for industrial energy cost prediction, demonstrate that Conv-BiLSTM consistently outperforms TFT and Conv-BiGRU, particularly under the Time of Use pricing scheme. TFT exhibits the highest efficacy when operating without imputation (Table 7). Adaptive KNN achieves the most substantial performance improvements with imputation, while Conv-BiLSTM achieves the lowest MSE and MAE (Table 8). VAE also enhances model efficacy; however, GAIN exhibits less significant enhancements. Furthermore, the performance of the model is improved by hyperparameter optimization (HPO), particularly for Conv-BiLSTM and TFT (Table 9). Selecting suitable imputation methods to optimize model parameters and obtain optimal predictive performance is crucial. These findings emphasize the significance of these steps.

Fig. 13 shows the MSE comparison for the same three models applied to industrial datasets. Like the commercial results, the Conv-BiLSTM model outperforms the others, achieving the lowest MSE and highlighting its robustness in industrial applications.

With adaptive KNN imputation, all models showed improved performance similar to the commercial dataset, with Conv-BiLSTM achieving the lowest error metrics, demonstrating its adaptability. GAIN did not improve performance as effectively as other methods, which might be due to the industrial dataset's unique characteristics. VAE provided significant improvements, with Conv-BiLSTM achieving the best results, indicating that VAE is particularly effective for handling missing data in industrial contexts.

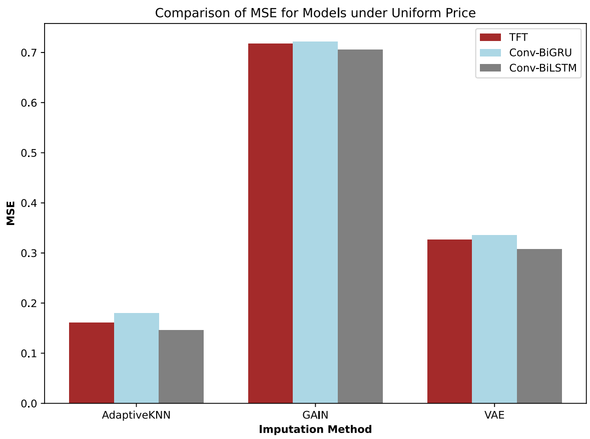
Fig. 14 presents the MSE comparison for industrial datasets using the same imputation methods. Again, the Conv-BiLSTM model with VAE imputation achieves the lowest MSE, confirming VAE's effectiveness in industrial contexts.

HPO further improved the models' performance. With adaptive KNN and HPO, Conv-BiLSTM showed the best results, highlighting the importance of parameter tuning. Even with HPO, GAIN performed better than other methods, suggesting it may not best fit this dataset. The combination of VAE and HPO yielded the best performance, particularly for Conv-BiLSTM, underscoring the efficacy of VAE and HPO in industrial energy cost prediction.

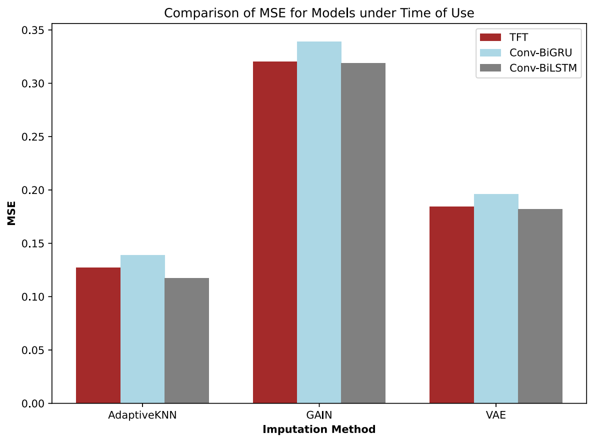
Fig. 15 demonstrates similar findings for industrial datasets. The Conv-BiLSTM model, when enhanced with VAE imputation and HPO, consistently achieves the lowest MSE, underscoring the importance of HPO in refining model performance and robustness.

Table 9
Experimental results of industrial energy cost prediction with imputation and HPO.

Model	Uniform				Time of use			
	MSE	MAPE	RMSE	MAE	MSE	MAPE	RMSE	MAE
adaptiveKNN								
TFT	0.161441	0.841119	0.401797	0.250851	0.127164	0.982212	0.356601	0.231769
Conv-BiGRU	0.180258	0.874118	0.424568	0.278354	0.138741	1.215054	0.372479	0.261526
Conv-BiLSTM	0.145891	0.771836	0.381957	0.234855	0.11735	1.049902	0.342565	0.228378
GAIN								
TFT	0.717705	1.929956	0.847175	0.452545	0.320121	0.85766	0.565793	0.30654
Conv-BiGRU	0.721813	2.102665	0.849596	0.469551	0.339033	1.090184	0.582265	0.351466
Conv-BiLSTM	0.705685	1.911014	0.84005	0.449078	0.318873	0.929392	0.564688	0.323744
VAE								
TFT	0.327182	32.90267	0.571998	0.335524	0.184218	0.544703	0.429207	0.230039
Conv-BiGRU	0.336158	26.82492	0.579791	0.314836	0.19597	0.671499	0.446285	0.259533
Conv-BiLSTM	0.308034	22.81297	0.555008	0.292923	0.182121	0.62143	0.426634	0.248188



((a)) Uniform

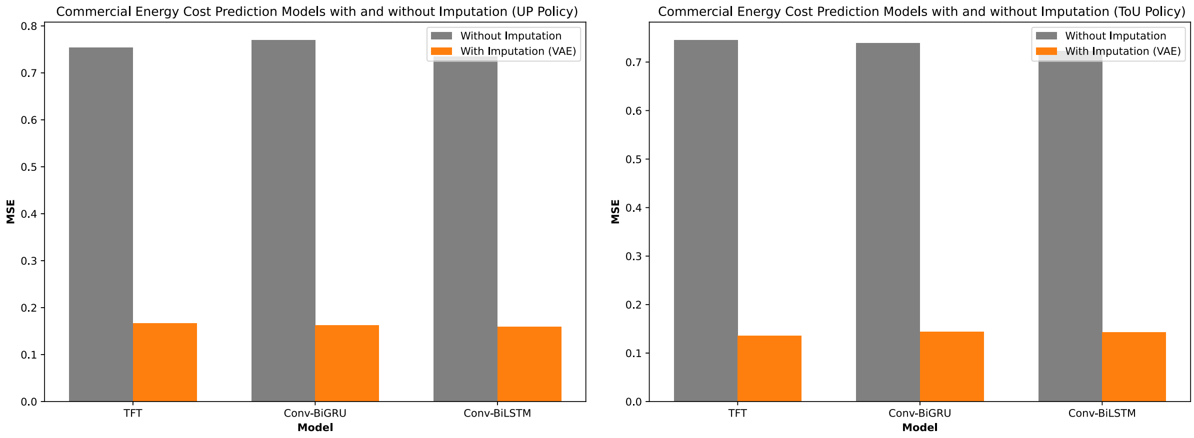


((b)) Time of Use

Fig. 15. Models comparison for industrial light data with imputation and HPO.

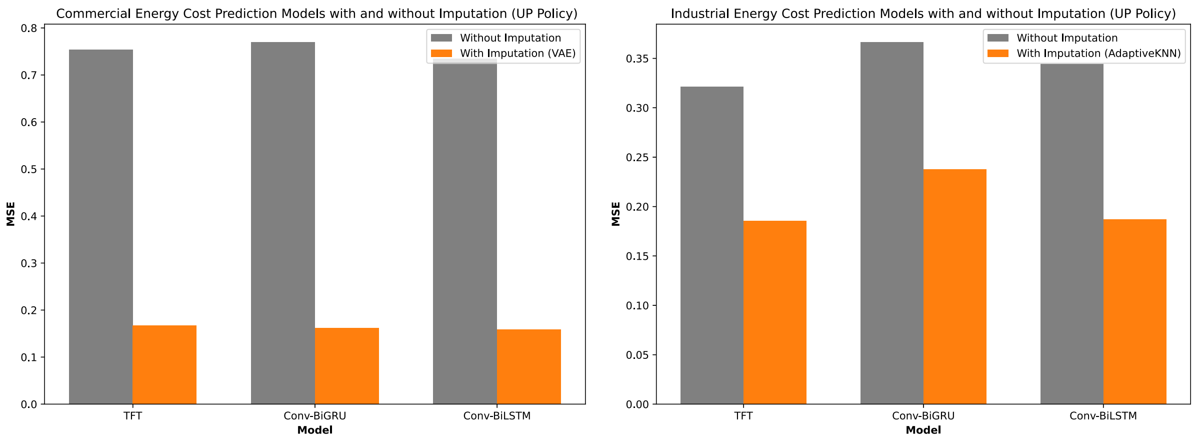
5.3. Impact of imputation methods and HPO for commercial and industrial prediction

This section presents an imputation and the HPO method's impact on commercial and industrial data. Figs. 16 and 17 compare a model with and without imputation for both commercial and industrial data, respectively. The analysis demonstrates that applying VAE imputation significantly reduces the MSE of energy cost prediction models across both Uniform Price and time-of-use pricing schemes. Conv-BiLSTM consistently achieves the lowest MSE, highlighting its effectiveness in handling missing data. These results underscore the critical importance of selecting appropriate imputation methods, as the right choice can significantly impact the accuracy of energy cost prediction models. Next, Imputation techniques considerably lower the MSE for models of energy cost prediction, according to the analysis in Fig. 17. VAE imputation greatly reduces MSE for commercial predictions; Conv-BiLSTM performs best. Adaptive KNN imputation improves model accuracy for industrial predictions; once more, Conv-BiLSTM shows better performance. Fig. 18 compares a model with and without the HPO for commercial data, and 19 compares a model for industrial light data. Under the Uniform Price Policy, HPO considerably lowers the MSE for commercial energy cost prediction models. All models, especially Conv-BiLSTM, show lower MSE after HPO using Adaptive KNN (Fig. 18 (a)). With GAIN (Fig. 18 (b), HPO has little effect. HPO also clearly results in performance improvements for VAE (Fig. 18 (c), stressing its relevance for model accuracy improvements. Fig. 19 shows how HPO affects MSE in industrial energy cost prediction models under the Time of Use Policy. With Adaptive KNN (Fig. 19 (a)), HPO considerably reduces MSE, particularly for Conv-BiLSTM. HPO exhibits a minimal increase in GAIN (Fig. 19 (b)). VAE (Fig. 19(c)) paired with HPO improves performance, with significant decreases in MSE for all models, highlighting the value of HPO in optimizing predictive accuracy.



((a)) Uniform Price Policy ((b)) Time of Use Price Policy

Fig. 16. Impact of imputation for commercial energy light prediction.



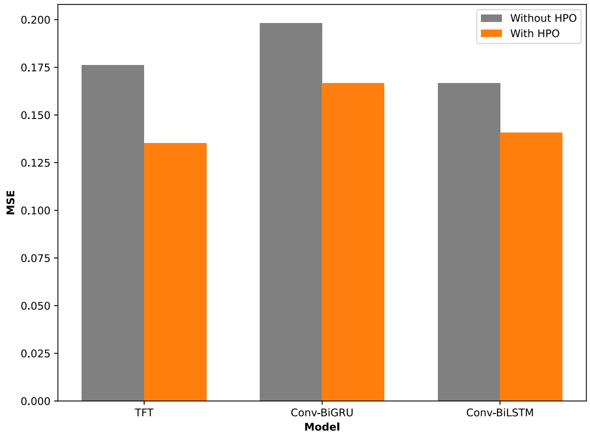
((a)) Uniform Price Policy ((b)) Time of Use Price Policy

Fig. 17. Impact of imputation for industrial energy light prediction.

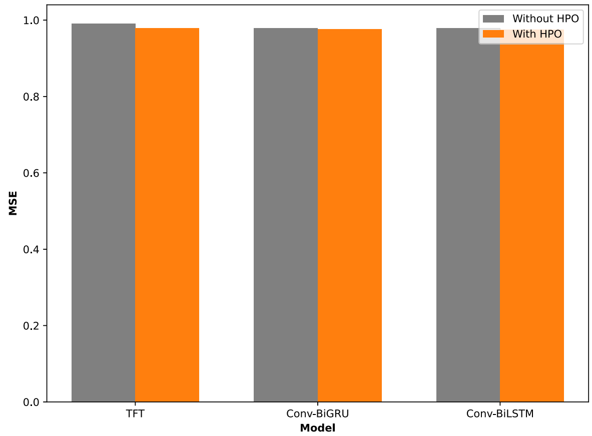
5.4. Discussion section

The results highlight several key considerations for model selection, imputation methods, and HPO. The Conv-BiLSTM model consistently outperformed others in commercial and industrial contexts, particularly when combined with advanced imputation methods and HPO. This indicates its high effectiveness for energy cost prediction across different datasets. Advanced imputation techniques like VAE and adaptive KNN significantly enhance model performance, with VAE showing the best results and demonstrating its suitability for handling complex missing data patterns. While GAIN also improves performance, it is less effective than VAE and adaptive KNN, suggesting it is not the best choice for all datasets. Optimizing model parameters is crucial for maximizing performance, as shown by the substantial improvements observed with HPO, underscoring its importance in refining model accuracy and efficiency. The analysis also indicates that models perform better under the time-of-use price policy than the Uniform Price policy, as evidenced by the lower MSE values. This suggests that energy cost predictions are more precise and reliable when based on time-of-use pricing.

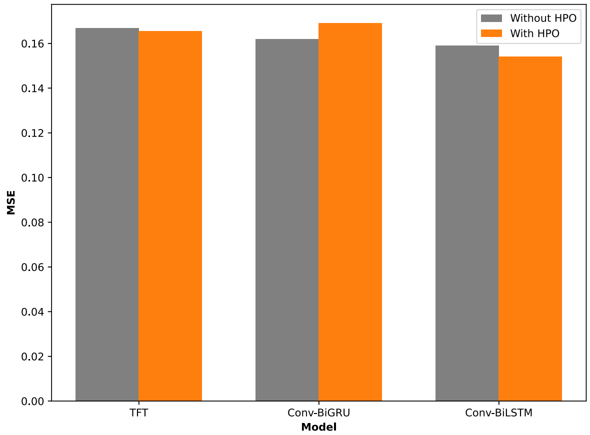
Overall, the Conv-BiLSTM model, particularly when combined with VAE imputation and HPO, offers the most accurate and reliable predictions for energy costs in commercial and industrial settings. These findings emphasize the critical role of data preprocessing, imputation, and parameter tuning in developing accurate and reliable predictive models for energy consumption. This study can inform future research and practical applications in energy management, leading to more efficient and accurate forecasting methods.



((a)) HPO and Non-HPO with Adaptive KNN



((b)) HPO and Non HPO with GAIN



((c)) HPO and Non HPO with VAE

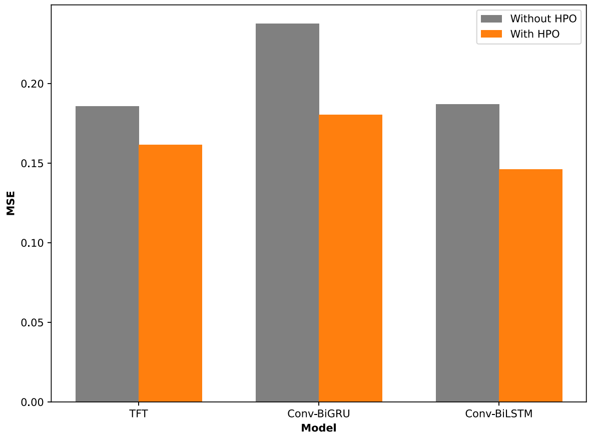
Fig. 18. Impact of HPO for commercial energy light prediction with uniform price policy.

5.5. Ablation study

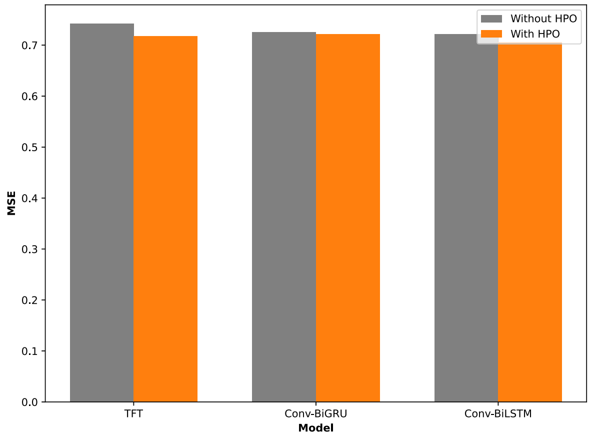
The experimental result shows the significance of imputation techniques for energy cost prediction. Amongst the imputation methods, the VAE and adaptive KNN have low MSE values for commercial and industrial energy cost prediction. The VAE imputation method can handle the complex missing data patterns, which helps the learning models consistently achieve the lowest MSE values. The adaptive KNN method has slightly lower performance than the VAE method for commercial datasets, and it outperformed other methods in industrial settings, demonstrating its adaptability to structured missing patterns. The GAIN method performs less than the VAE and adaptive KNN, potentially due to mismatches between the dataset characteristics and GAIN's underlying assumptions. The experimental findings demonstrate the importance of an appropriate imputation strategy, which influences the model performance.

Amongst the learning, the Conv-BiLSTM model shows a superior predictive capability than the Conv-BiGRU and TFT models for both commercial and industrial datasets. Without imputation, the Conv-BiLSTM achieves lower MSE, MAE, RMSE and MAPE values, outperforming the other models handling the raw data. After applying the imputation method, the performance of a TFT and Conv-BiGRU improved slightly. Still, the Conv-BiLSTM model utilizes the imputed data more effectively than the other models, which indicates that it captures the complex temporal dependencies and interactions effectively. The findings suggest that the architectural advantages of Conv-BiLSTM, such as its convolutional and bidirectional sequence modeling components, make the learning model well-suited for energy cost prediction.

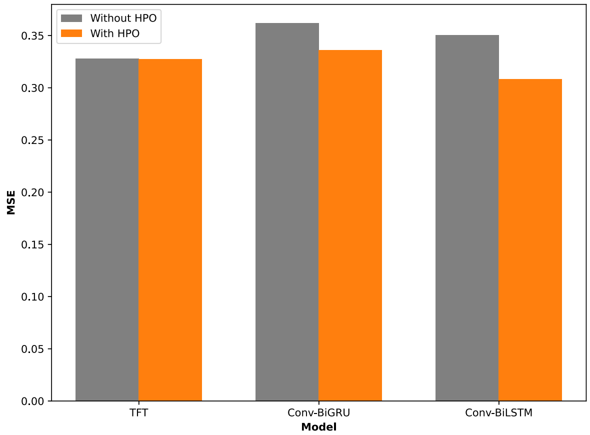
Bayesian optimization enhances the performance of a learning model and reduces the MSE and MAE for all models. Conv-BiLSTM has a lower prediction error with tuned parameters than the other learning model, demonstrating the effectiveness of tuning parameters to align with dataset characteristics. The combination of Conv-BiLSTM, VAE, and HPO results in more accurate energy cost predictions, which shows the importance of parameter tuning and imputation strategies in optimizing predictive performance. This combination of a model is new and not utilized for lightning electricity energy cost predictions.



((a)) HPO and Non-HPO with Adaptive KNN



((b)) HPO and Non HPO with GAIN



((c)) HPO and Non HPO with VAE

Fig. 19. Impact of HPO for industrial energy light prediction with time of use policy.

5.6. *Practical implications of results*

Integrating the Conv-BiLSTM model, VAE-based imputation, and hyperparameter optimization (HPO), the proposed energy cost prediction framework is engineered for seamless deployment within modern Building Management Systems (BMS). This system is tailored to the requirements of smart building ecosystems, utilizing advanced metering infrastructure and IoT-enabled devices for real-time energy monitoring, analytics, and control. The Conv-BiLSTM model forms the framework's core, providing high-fidelity energy cost predictions critical for data-driven energy management. Accurate cost forecasts enable precise budget planning and allocation, mitigating the risk of financial inefficiencies and unforeseen expenses. VAE-based imputation addresses the pervasive issue of missing data in building energy datasets, ensuring that analytics remain robust and consistent despite data incompleteness. HPO further optimizes model hyperparameters, enhancing predictive accuracy and ensuring the framework's adaptability and resilience across heterogeneous operational environments.

The predictive insights generated by the proposed system empower stakeholders to make strategic, data-informed decisions regarding light energy consumption. Detailed analytics derived from the framework identify energy usage patterns, uncover inefficiencies, and highlight peak demand intervals, optimizing energy-intensive operations. This includes aligning such operations with off-peak periods or low-tariff intervals, leading to substantial cost savings. Additionally, by integrating real-time analytics, the framework supports the dynamic adjustment of energy schedules, facilitating demand-response strategies to prevent grid overloads and leverage real-time pricing incentives. Predictive analytics extend beyond cost savings to support broader sustainability and operational goals. By pinpointing sources of energy wastage and proposing corrective measures, the framework promotes carbon footprint reduction and aligns with global sustainability objectives. Integrated into BMS platforms, the system enables predictive maintenance, reducing the likelihood of equipment failures, prolonging the operational lifespan of critical infrastructure, and maintaining optimal system performance. This creates a foundation for intelligent, resilient, and environmentally sustainable building operations. Experimental validation underscores the feasibility of implementing the framework in practical settings, demonstrating its capacity to transform energy management practices. The proposed system presents a cutting-edge solution for optimizing energy efficiency and achieving long-term sustainability within modern smart buildings by delivering accurate energy cost predictions, actionable insights, and robust performance under varying conditions.

6. Conclusion

The experimental results of energy cost prediction models for commercial and industrial buildings have unveiled several important insights. The Conv-BiLSTM model consistently outperforms the TFT and Conv-BiGRU models across both datasets, particularly when combined with HPO and advanced imputation techniques. The Conv-BiLSTM model achieves superior accuracy in commercial energy cost prediction, evidenced by the lowest MSE, both with and without imputation. Integrating VAE with Conv-BiLSTM and HPO significantly enhances performance, as VAE is particularly effective in addressing complex patterns of missing data. While adaptive KNN imputation offers some benefits, it is more effective than GAIN but less effective than VAE. HPO further validates the superior performance of the Conv-BiLSTM model by improving its accuracy. For industrial energy cost predictions, similar patterns emerge. The Conv-BiLSTM model, when optimized with HPO and paired with VAE and adaptive KNN imputation, consistently achieves the lowest error metrics. This study introduces a novel hybrid cost prediction framework that integrates the Conv-BiLSTM, Conv-BiGRU, and TFT models, addressing the complexities of temporal and spatial dependencies in energy data. The Conv-BiLSTM model, particularly when combined with VAE imputation and HPO, stands out as the most effective technique for energy cost prediction in both commercial and industrial environments. Additionally, this study incorporates dynamic pricing data, demonstrating that Time-of-Use (ToU) pricing achieves more accurate predictions than uniform pricing policies. The novelty of this research lies in the unique combination of imputation methods, hybrid deep learning models, and hyperparameter optimization techniques. It addresses critical challenges in energy cost prediction, such as handling missing data, improving model robustness, and capturing complex temporal dependencies. This study underscores the importance of selecting appropriate preprocessing methods and fine-tuning model parameters to achieve optimal predictive performance. These insights can inform future research and practical applications, ultimately contributing to sustainable energy management in commercial and industrial contexts. In the future, we aim to employ a decentralized, federated learning mechanism to perform collaborative training between buildings while preserving local data privacy. This approach ensures the confidentiality of sensitive local data by training the model locally and only sharing model updates, such as weights or gradients, with neighboring participants. By learning insights from multiple buildings collaboratively, this method enhances prediction performance and provides a privacy-preserving solution for energy management and cost prediction in heterogeneous environments.

CRediT authorship contribution statement

Qazi Waqas Khan: Writing – review & editing, Writing – original draft, Visualization, Methodology, Formal analysis, Conceptualization. **Anam Nawaz Khan:** Writing – review & editing, Writing – original draft, Methodology, Conceptualization. **Bibi Misbah:** Writing – original draft, Software, Formal analysis, Data curation. **Rashid Ahmad:** Writing – review & editing, Visualization, Validation, Methodology, Investigation. **Do Hyeun Kim:** Writing – review & editing, Validation, Supervision, Software, Resources, Project administration, Funding acquisition.

Qazi Waqas Khan and Anam Nawaz Khan contributed equally to this paper.

Declaration of competing interest

The authors declare no competing interests.

Acknowledgments

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No. RS-2024-00405278). and This research was supported by ‘‘Regional Innovation Strategy (RIS)" through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(MOE)(2023RIS-009)]. Any correspondence related to this paper should be addressed to Do Hyeun Kim

Data availability

Data will be made available on request.

References

[1] Keke Wang, Min Yu, Dongxiao Niu, Yi Liang, Sha Peng, Xiaomin Xu, Short-term electricity price forecasting based on similarity day screening, two-layer decomposition technique and Bi-LSTM neural network, Appl. Soft Comput. 136 (2023) 110018.

[2] Tingting Zhang, Zhenpeng Tang, Junchuan Wu, Xiaoxu Du, Kaijie Chen, Short term electricity price forecasting using a new hybrid model based on two-layer decomposition technique and ensemble learning, Electr. Power Syst. Res. 205 (2022) 107762.

[3] Anam-Nawaz Khan, Naeem Iqbal, Atif Rizwan, Rashid Ahmad, Do-Hyeun Kim, An ensemble energy consumption forecasting model based on spatial-temporal clustering analysis in residential buildings, Energies 14 (11) (2021) 3020.

[4] QJUDE Qdr, Benefits of demand response in electricity markets and recommendations for achieving them, Tech. Rep., 2006, US Dept. Energy, Washington, DC, USA, 2006, p. 95.

[5] Mohammad Rastegar, Mahmud Fotuhi-Firuzabad, Farrokh Aminifar, Load commitment in a smart home, Appl. Energy 96 (2012) 45–54.

[6] Alexander Boogert, Dominique Dupont, When supply meets demand: the case of hourly spot electricity prices, IEEE Trans. Power Syst. 23 (2) (2008) 389–398.

[7] Qazi Waqas Khan, Rashid Ahmad, Atif Rizwan, Anam Nawaz Khan, KyuTae Lee, Do Hyeun Kim, Optimizing energy efficiency and comfort in smart homes through predictive optimization: A case study with indoor environmental parameter consideration, Energy Rep. 11 (2024) 5619–5637.

[8] Kaile Zhou, Shanlin Yang, Demand side management in China: The context of China's power industry reform, Renew. Sustain. Energy Rev. 47 (2015) 954–965.

[9] Yong Qiang Tan, Yan Xia Shen, Xin Yan Yu, Xin Lu, Day-ahead electricity price forecasting employing a novel hybrid frame of deep learning methods: A case study in NSW, Australia, Electr. Power Syst. Res. 220 (2023) 109300.

[10] Amir Safdarian, Mahmud Fotuhi-Firuzabad, Matti Lehtonen, Benefits of demand response on operation of distribution networks: A case study, IEEE Syst. J. 10 (1) (2014) 189–197.

[11] Goutam Dutta, Krishnendranath Mitra, A literature review on dynamic pricing of electricity, J. Oper. Res. Soc. 68 (2017) 1131–1145.

[12] Touraj Ashrafian, Enhancing school buildings energy efficiency under climate change: A comprehensive analysis of energy, cost, and comfort factors, J. Build. Eng. 80 (2023) 107969.

[13] Lize Wang, Dong Xie, Lifeng Zhou, Zixuan Zhang, Application of the hybrid neural network model for energy consumption prediction of office buildings, J. Build. Eng. 72 (2023) 106503.

[14] Mohd Herwan Sulaiman, Zuriani Mustaffa, Muhammad Salihin Saealal, Mohd Mawardi Saari, Abu Zaharin Ahmad, Utilizing the Kolmogorov-arnold networks for chiller energy consumption prediction in commercial building, J. Build. Eng. 96 (2024) 110475.

[15] Jingtao Liu, Zhihong Zhai, Yuxiang Zhang, Yixian Wang, Yunfei Ding, Comparison of energy consumption prediction models for air conditioning at different time scales for large public buildings, J. Build. Eng. 96 (2024) 110423.

[16] Yiqing Wang, Hengda Chen, Huanxin Chen, Mingshu Ye, Yicheng Ren, Chuang Yang, A hybrid model based on wavelet decomposition and LSTM for short-term energy consumption prediction of chillers, J. Build. Eng. (2024) 111539.

[17] Anam-Nawaz Khan, Naeem Iqbal, Atif Rizwan, Rashid Ahmad, Do-Hyeun Kim, An ensemble energy consumption forecasting model based on spatial-temporal clustering analysis in residential buildings, Energies 14 (11) (2021) 3020.

[18] Xiaojie Zhou, Ning Wang, Jixin Zou, Guiting Liu, Xucheng Zhuang, Guodan Liu, Analysis and prediction of energy consumption in office buildings with variable refrigerant flow systems: A case study, J. Build. Eng. 97 (2024) 110936.

[19] Fazıl Gökgöz, Fahrettin Filiz, Electricity price forecasting: A comparative analysis with shallow-ann and dnn, Int. J. Energy Power Eng. 12 (6) (2018) 421–425.

[20] Long Wang, Zijun Zhang, Jieqiu Chen, Short-term electricity price forecasting with stacked denoising autoencoders, IEEE Trans. Power Syst. 32 (4) (2016) 2673–2681.

[21] Tasmeea Rahman, Mohammad Lutfi Othman, Samsul Bahari Mohd Noor, Wan Fatinhamamah Binti Wan Ahmad, Mohamad Fani Sulaima, Methods and attributes for customer-centric dynamic electricity tariff design: A review, Renew. Sustain. Energy Rev. 192 (2024) 114228.

[22] Foued Saâdaoui, Hana Rabbouch, A wavelet-based hybrid neural network for short-term electricity prices forecasting, Artif. Intell. Rev. 52 (2019) 649–669.

[23] Zhang Yang, Li Ce, Li Lian, Electricity price forecasting by a hybrid model, combining wavelet transform, ARMA and kernel-based extreme learning machine methods, Appl. Energy 190 (2017) 291–305.

[24] Xing Yan, Nurul A. Chowdhury, Mid-term electricity market clearing price forecasting utilizing hybrid support vector machine and auto-regressive moving average with external input, Int. J. Electr. Power Energy Syst. 63 (2014) 64–70.

[25] Wei Li, Denis Mike Becker, Day-ahead electricity price prediction applying hybrid models of LSTM-based deep learning methods and feature selection algorithms under consideration of market coupling, Energy 237 (2021) 121543.

[26] Haolin Yang, Kristen R. Schell, Real-time electricity price forecasting of wind farms with deep neural network transfer learning and hybrid datasets, Appl. Energy 299 (2021) 117242.

[27] Yangrui Zhang, Peng Tao, Xiangming Wu, Chenguang Yang, Guang Han, Hui Zhou, Yinlong Hu, Hourly electricity price prediction for electricity market with high proportion of wind and solar power, Energies 15 (4) (2022) 1345.

[28] Paolo Gabrielli, Moritz Wüthrich, Steffen Blume, Giovanni Sansavini, Data-driven modeling for long-term electricity price forecasting, Energy 244 (2022) 123107.

[29] Athanasios Ioannis Arvanitidis, Dimitrios Bargiotas, Dimitrios Kontogiannis, Athanasios Fevgas, Miltiadis Alamaniotis, Optimized data-driven models for short-term electricity price forecasting based on signal decomposition and clustering techniques, Energies 15 (21) (2022) 7929.

[30] Dimitrios Kontogiannis, Dimitrios Bargiotas, Aspassia Daskalopulu, Athanasios Ioannis Arvanitidis, Lefteri H Tsoukalas, Error compensation enhanced day-ahead electricity price forecasting, Energies 15 (4) (2022) 1466.

[31] Catherine McHugh, Sonya Coleman, Dermot Kerr, Hourly electricity price forecasting with NARMAX, Mach. Learn. Appl. 9 (2022) 100383.

[32] Vasudharini Sridharan, Mingjian Tuo, Xingpeng Li, Wholesale electricity price forecasting using integrated long-term recurrent convolutional network model, Energies 15 (20) (2022) 7606.

[33] Jingyi Shang, Jinfeng Gao, Xin Jiang, Mingguang Liu, Dunnan Liu, Optimal configuration of hybrid energy systems considering power to hydrogen and electricity-price prediction: A two-stage multi-objective bi-level framework, Energy 263 (2023) 126023.

[34] Chenxu Zhang, Yong Fu, Probabilistic electricity price forecast with optimal prediction interval, IEEE Trans. Power Syst. (2023).

[35] Anna Gloria Billé, Angelica Gianfreda, Filippo Del Grosso, Francesco Ravazzolo, Forecasting electricity prices with expert, linear, and nonlinear models, Int. J. Forecast. 39 (2) (2023) 570–586.

[36] Anany Pandey, Manish Pandey, A robust neural network based short time electricity price prediction, in: 2023 3rd International Conference on Intelligent Communication and Computational Techniques, ICCT, IEEE, 2023, pp. 1–8.

[37] Abdallah Abdellatif, Hamza Mubarak, Shameem Ahmad, Saad Mekhilef, Hamdan Abdellatef, Hazlie Mokhlis, Jeevan Kanesan, Electricity price forecasting one day ahead by employing hybrid deep learning model, in: 2023 IEEE IAS Global Conference on Renewable Energy and Hydrogen Technologies (GlobConHT), IEEE, 2023, pp. 1–5.

[38] Biao Zhang, Chao Song, Xuchu Jiang, Ying Li, Electricity price forecast based on the STL-TCN-NBEATS model, Heliyon 9 (1) (2023).

[39] Feihong Xu, Xianliang Teng, Jixiang Lu, Tao Zheng, Yulong Jin, Prediction of day-ahead electricity price based on N-BEATSx model optimized by SSA considering coupling between features, in: The Purple Mountain Forum on Smart Grid Protection and Control, Springer, 2022, pp. 178–194.

[40] Hamza Mubarak, Shameem Ahmad, Al Amin Hossain, Ben Horan, Abdallah Abdellatif, Saad Mekhilef, Mehdi Seyedmahmoudian, Alex Stojcevski, Hazlie Mokhlis, Jeevan Kanesan, et al., Short-term electricity price forecasting using interpretable hybrid machine learning models, in: 2023 IEEE IAS Global Conference on Renewable Energy and Hydrogen Technologies (GlobConHT), IEEE, 2023, pp. 1–6.

[41] Ping-Huan Kuo, Chiou-Jye Huang, An electricity price forecasting model by hybrid structured deep neural networks, Sustainability 10 (4) (2018) 1280.

[42] Mohammadreza Heidarpanah, Farhad Hooshyaripor, Meysam Fazeli, Daily electricity price forecasting using artificial intelligence models in the Iranian electricity market, Energy 263 (2023) 126011.

[43] Gholamreza Memarzadeh, Farshid Keynia, Short-term electricity load and price forecasting by a new optimal LSTM-NN based prediction algorithm, Electr. Power Syst. Res. 192 (2021) 106995.

[44] Moein Qaisari Hasan Abadi, Russell Sadeghi, Ava Hajian, Omid Shahvari, Amirehsan Ghasemi, A blockchain-based dynamic energy pricing model for supply chain resiliency using machine learning, Supply Chain Anal. 6 (2024) 100066.

[45] Giorgia Oggioni, Alexandra Schwartz, Ann-Kathrin Wiertz, Gregor Zöttl, Dynamic pricing and strategic retailers in the energy sector: A multi-leader-follower approach, European J. Oper. Res. 312 (1) (2024) 255–272.

[46] Hannah Covington, Brian Woo-Shem, Chenli Wang, Thomas Roth, Cuong Nguyen, Yuhong Liu, Yi Fang, Hohyun Lee, Method for evaluating fairness of electricity tariffs with regard to income level of residential buildings, Appl. Energy 353 (2024) 122130.

[47] William Gomez, Fu-Kwun Wang, Jia-Hong Chou, Li-ion battery capacity prediction using improved temporal fusion transformer model, Energy 296 (2024) 131114.

[48] Amril Nazir, Abdul Khalique Shaikh, Abdul Salam Shah, Ashraf Khalil, Forecasting energy consumption demand of customers in smart grid using temporal fusion transformer (TFT), Results Eng. 17 (2023) 100888.

[49] Sinan Rasiya Koya, Tirthankar Roy, Temporal fusion transformers for streamflow prediction: Value of combining attention with recurrence, J. Hydrol. (2024) 131301.

[50] Bryan Lim, Sercan Ö Arık, Nicolas Loeff, Tomas Pfister, Temporal fusion transformers for interpretable multi-horizon time series forecasting, Int. J. Forecast. 37 (4) (2021) 1748–1764.

[51] Jia Wu, Xiu-Yun Chen, Hao Zhang, Li-Dong Xiong, Hang Lei, Si-Hao Deng, Hyperparameter optimization for machine learning models based on Bayesian optimization, J. Electron. Sci. Technol. 17 (1) (2019) 26–40.