Research Article

A Novel Data-Driven Method for Medium-Term Power Consumption Forecasting Based on Transformer-LightGBM

Guang Yang ⓘ, Songhuai Du ⓘ, Qingling Duan ⓘ, and Juan Su ⓘ

College of Information and Electrical Engineering, China Agricultural University, Beijing, China

Correspondence should be addressed to Guang Yang; yangguang2018@cau.edu.cn

With the widespread use of new energy sources and Internet of things, the power market landscape has become complex. In particular, new energy is more stochastic and volatile; it is prone to the problems of inaccurate forecasting on longer time scales, affecting electricity trading. This study proposes a new method for predicting medium-term load series data based on the transformer-lightGBM. The method first preprocesses electricity market data, including missing value processing, outlier processing, overall analysis, and correlation analysis, to extract features with a strong correlation to medium-term electricity consumption forecasts. Then, a transformer neural network is used to learn the complex patterns and dynamic time scales of the load series data to predict the day-ahead market series. Finally, lightGBM is used to combine power characteristics and time characteristics to forecast power consumption. The effectiveness of the proposed method is proved using the ISO-NE dataset. Experimental results indicate that the present method verified more accurate prediction than LSTM-based methods.

1. Introduction

Load forecasts are based on the known demand for electricity and take into account the political, economic, climatic, and other relevant factors to forecast future electricity demand. Internet of things brings more extensive data access to smart grid, making the data involved in prediction richer. From an economic point of view, a load forecast is essentially a forecast of electricity market demand [1]. In recent years, renewable energy has grown considerably under the target incentives and policy support of governments. Owing to the high randomness of renewable energy, there is a trouble of inaccurate forecasting on a longer time scale, which may result in more conservative trading results in medium- and long-term power trading, which is not conducive to the full consumption of renewable energy [2]. Therefore, many countries in the electricity market use medium- and long-term contracts to lock in power generation and create revenue. Then, they introduce electricity spot market construction to promote direct participation of renewable energy in the market competitive transactions [3]. The electricity spot market consists of day-ahead demand, composed primarily of fixed and price-sensitive demand bids, as well as real-time demand, defined as the sum of nondistributed load assets, station-served load assets, and nonmetered load assets.

The accuracy of load forecasting has a large impact on electricity generators and operators. A forecast that is too low may result in reduced revenue from electricity sales; a forecast that is too high may result in poor utilization of new generation capacity or even existing generation capacity. The medium-term load forecast has a forecast period of 1 month to 1 year and is used for the preparation of long-term operational plans for reservoir scheduling, unit maintenance, exchange plans, and fuel plans. The main methods of medium-term load forecasting include dynamic averaging based on extrapolation of time-series trends, exponential smoothing, growth rate methods, grey forecasting, Markov forecasting, and growth curve methods [4]. In medium-term load forecasting, single-factor regression analysis and elasticity coefficient methods are used to take into account unrelated factors, and forecasting methods for multi-correlated factors include multiple regression analysis, clustering forecasting, decision trees, and econometric

methods. Machine learning techniques have developed rapidly; they can establish complex nonlinear relationships through a learning process involving trends in historical data.

The following studies highlight the latest advances in forecasting future energy demand. Guan et al. [5] proposed a method for the kernel function of the Gaussian process model for long-term load probability prediction. Son et al. [6] proposed that LSTM was used for medium- and long-term load forecast. Omaji et al. [7] proposed a model to predict the hourly load for the following month based on historical hourly data and temperature data, and improved entropy mutual information was used in the data pre-processing. Dong et al. [8] proposed a new selective sequence learning method to transform the multiyear long-duration prediction problem into a sequence prediction problem with multiple time steps. Also, a neural network prediction model with singular spectrum analysis was proposed for the problem of the decomposition prediction accuracy of nonsmooth, nonlinear, and medium-term load sequences. Further, Liu and Zhao [9] proposed a medium-term fore-casting method considering economic and meteorological factors. First, the monthly historical electricity consumption is decomposed into long-term trend components, cyclical components, seasonal components, and irregular compo-nents using the seasonal decomposition method. Then, based on electricity, meteorological, and economic data, a support vector machine is used to forecast each component separately and make a comprehensive forecast of the total monthly electricity consumption. Zhang et al. [10] proposed a medium- and long-term forecasting model that takes into account the coordination relationship and lag effect of each influencing factor. The strongly correlated factors affecting the load change are obtained through correlation matrix screening. After obtaining the characteristic decomposition part, the delayed effect test is used to determine the number of lag periods, and the effect of data noise is removed using principal component analysis. He et al. [11] presented a probability density with a continuous conditional quantile function to predict the medium-term electricity load for a given day and introduced the concept of electricity load density. Khatereh et al. [12] proposed a model for predicting solar energy with a HMM to find the energy variation at a specified time over consecutive days. Shang et al. [13] used WNN and generalized regression to predict weather factors, Elman neural networks, and the cuckoo search to optimally predict wind speed. Finally, the variance analysis model was presented to combine the forecast results of weather factors and wind speed data. Shobana Devi et al. [14] presented an integration prediction method to predict wind power to improve the performance of the prediction. A modified LSTM enhanced forgetting gate model was used to optimize the parameters of the LSTM-EFG model using the cuckoo search optimization algorithm for the prediction of subseries data extracted from integrated empirical modal decomposi-tion. Xia et al. [15] improved GRU-RNN structure, and improved training methods were used to improve robust-ness. Yang et al. [16] proposed a multitask prediction framework with BDL to quantify the uncertainty among different groups. Mohsen et al. [17] using customers' con-sumption records at different times, proposed a reliable procedure to check consumption changes. Florian et al. [18] derived speed prediction curves taking into account indi-vidual driving style characteristics and real-time traffic data to obtain EV vehicle consumption predictions. Alonso et al. [19] proposed a multilevel poly-time series clustering method using several representative features to summarize each time series to quantile autocovariance as well as simple and partial autocorrelation. Qian et al. [20] proposed a combination of simulation and transfer learning to improve the prediction accuracy of thermal and cooling loads to address historical data due to changes. Tanveer and Zhang [21] introduced two novel deeply supervised machine learning models, including a fitted stochastic feature ex-pansion Gaussian kernel regression model and a nonpara-metric KNN-based model for demand forecasting for buildings and utility companies. Electricity demand was forecasted in the medium-to-long term by analyzing cus-tomers' electricity consumption patterns to stabilize the supply [6]. Nasir et al. [22] presented a mixture framework of SVM, GRU, and CNN. GreyWolf optimization and EarthWorm optimization were used to optimize the hyperparameters of the SVM and CNN-GRU. Omaji et al. [7] presented a method to forecast monthly hourly loads in advance using hourly load and temperature data. An im-proved entropic mutual information feature selection method was used for data preprocessing, and CRBM was used for load forecasting, while consumer behavior was clustered using adaptive k-means.

The transformer is a new model proposed in 2017 [23]. It is based entirely on the attention mechanism and completely discards the structure of CNN and RNN to solve the long-range dependency problem of RNN and its variants. It has a better memory, remembers information over longer dis-tances, and supports parallelized computing. Its ground-breaking ideas turn the previous equation of sequence modeling with RNNs on its head. So, it has been widely used in various areas of natural language processing [24, 25], but it has been less used in load forecasting tasks.

For the above motivation, we propose a novel data-driven method for medium-term power consumption forecasting based on the transformer-lightGBM method, which aims to further improve the accuracy of electricity market forecasting for long time series. The study's key contributions are as follows:

(i) A data-driven method is adopted to analyze the data of the New England electricity market, including missing value processing, outlier processing, holistic analysis, and correlation analysis, and features strongly correlated with medium-term electricity consumption forecast are extracted for forecasting.

(ii) In this study, the transformer is used to capture the features of the actual changes and fluctuating trends of the load volume. For the feature in power con-sumption forecasting, the transformer-lightGBM method combines a multiheaded self-attentive mechanism with a temporal modeling capability.
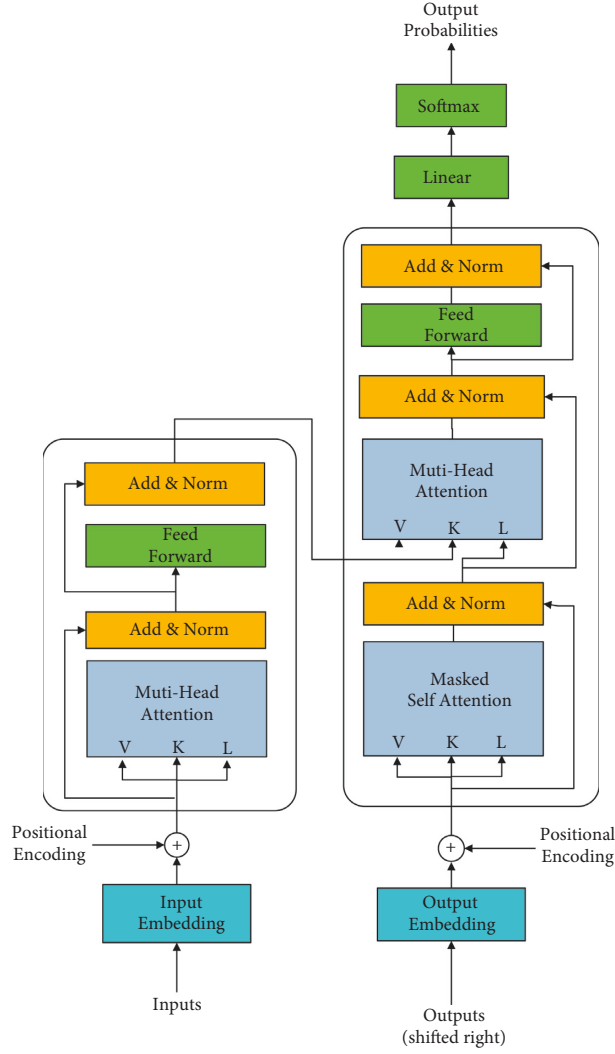
FIGURE 1: Transformer architecture.

The transformer model can extract long-term temporal relationships owing to its special multi-headed attention structure, and lightGBM is an integrated learning framework that integrates feature sequences with efficient adaptive boosting capabilities.

(iii) In this study, we derive monthly and quarterly power consumption from hourly demand forecasts. Our proposed method allows for multitimescale forecasting by paying attention to the selected feature and the results of the feature forecasts, which allows our method to improve the accuracy of power consumption in the context of renewable energy.

2. Summary Method

2.1. PCC. The PCC is a characteristic quantity of a random variable used to measure the linear relationship between two continuous random normal variables, as shown in the following equations:

$$
\begin{aligned}
\rho_{XY} &= \frac{\mathrm{cov}(X, Y)}{\sigma_X \sigma_Y} \\
&= \frac{E\{[X - E(X)][Y - E(Y)]\}}{\sigma_X \sigma_Y} \\
&= \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \overline{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \overline{Y})^2}},
\end{aligned}
\tag{1}
$$

$$
\begin{aligned}
\mathrm{cov}(X, Y) &= E\{[X - E(X)][Y - E(Y)]\} \\
&= E(XY) - E(X)E(Y),
\end{aligned}
\tag{2}
$$

$$
\mathrm{cov}(X, Y) = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{n - 1}.
\tag{3}
$$

In the above formula, X and Y are stochastic variables, ρ_{XY} is PCC, E is the expected value of the random variable, cov is the covariance, *cov (x, y)* is the sample covariance, and σ_X and σ_Y are the standard deviations. The PCC is a value between −1 and 1. When the linear relationship between two

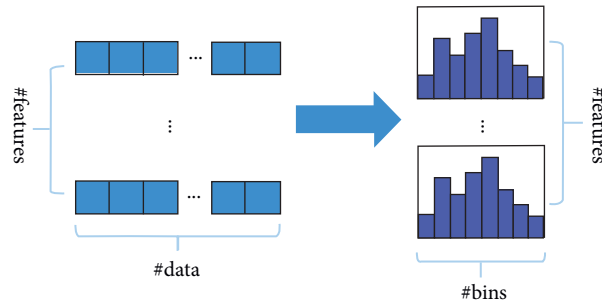FIGURE 2: Histogram optimization algorithm.

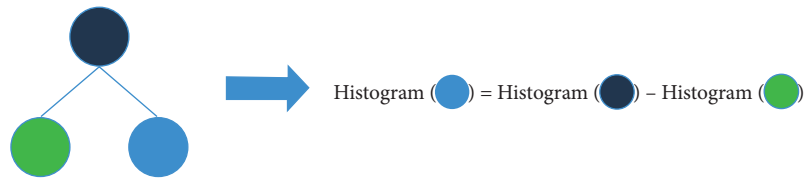Histogram (🔵) = Histogram (⚫) – Histogram (🟢)

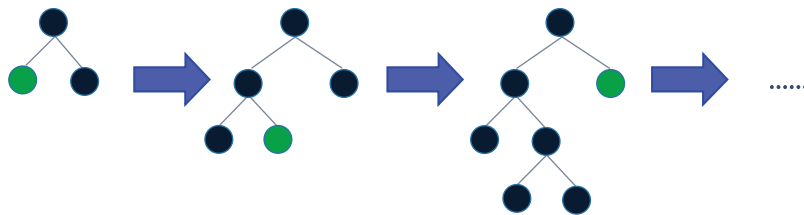FIGURE 3: Construct a histogram of a leaf.
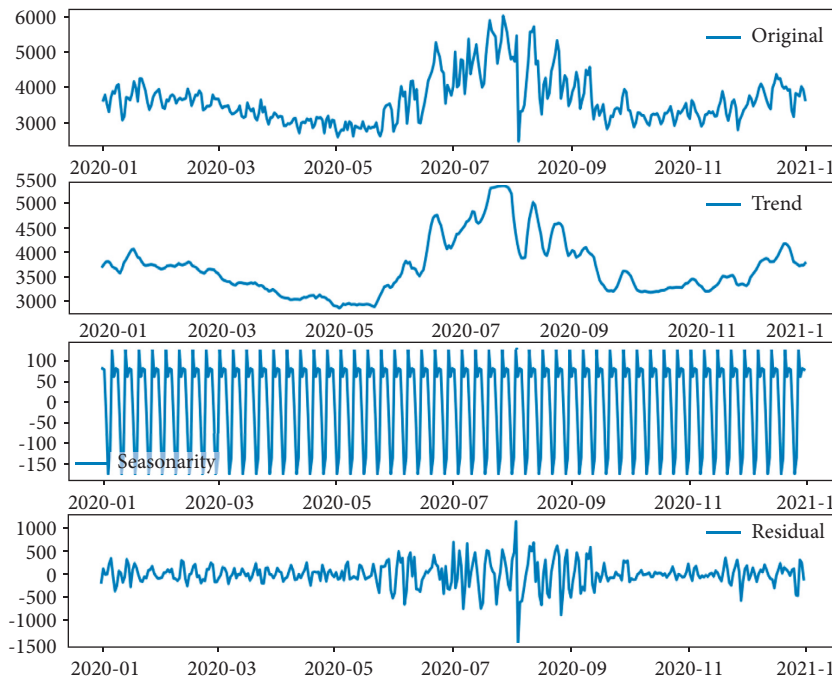

FIGURE 4: Leaf-wise tree growth.

FIGURE 5: The seasonal decomposition of the real-time demand series at 19:00 in 2020.
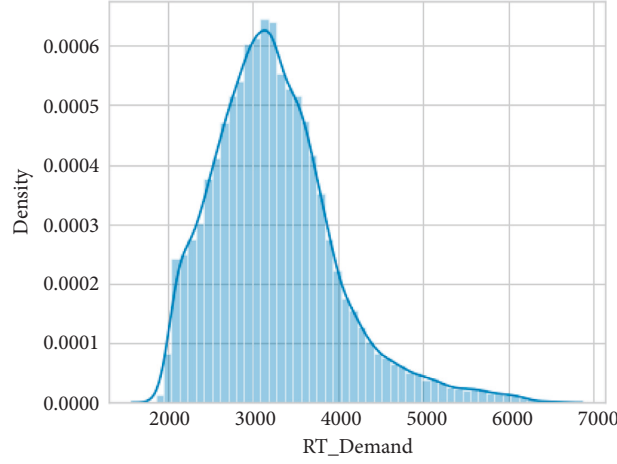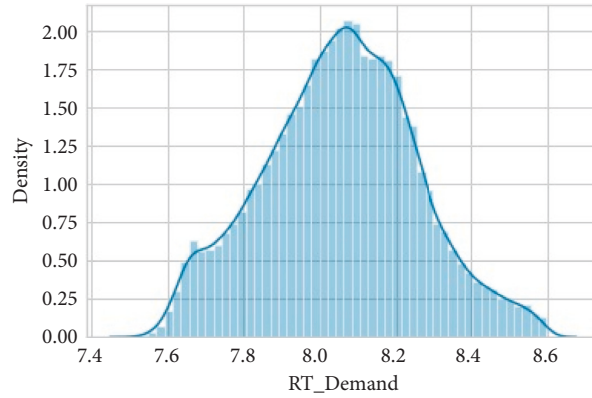
FIGURE 6: Real-time demand density distribution.

FIGURE 7: Real-time demand logarithmic density distribution.

variables is enhanced, the PCC tends toward 1 or −1, When one variable increases and the other also increases, the PCC is greater than 0. If one variable increases and the other decreases, the PCC is less than 0. If the PCC is equal to 0, there is no linear correlation.

2.2. Transformer Model. The transformer model improves the attention mechanism and discards gating models such as LSTM, without the RNNs or CNNs. The transformer uses a stacked self-attention and an overall structure of point-to-point fully connected layers in the encoder and decoder. The structure of transformer is shown in Figure 1.

The proposed attention mechanism is derived from the human visual processing mechanism. After visual input, not all the information is processed, but the attention is focused on a specific part. The transformer is a Seq2Seq model, with an encoder and a decoder, not an RNN. It is based on the mechanisms of attention and self-attention.

Attention involves giving more weight to key information. Here, we use the column vector X to represent the input data and the sequence $[x_1, x_2, \ldots, x_N]$ to represent the relevant input vector. The query vector q is given by the attention mechanism to calculate its correlation with the

input vector. $T \in [1, N]$ represents the position of the selected input sequence. The definition is shown in the following equation:

$$
\begin{aligned}
a_i &= p(t = 1|X, q) \\
&= \text{soft max}(s(x_i, q)) \\
&= \frac{\exp(s(x_i, q))}{\sum_{j-1}^{N} \exp(s(x_i, q))},
\end{aligned}
\tag{4}
$$

where s is the function to calculate the score, and α_i is the distribution to the attention mechanism. In this study, the scaled dot-product attention commonly used in the self-attention model is used to represent the score function, as shown in the following equation:

$$
s(x_i, q) = \frac{x_i^T q}{\sqrt{d}},
\tag{5}
$$

where d is the dimension of x. The self-attention introduces Q (query vectors to match others), K (key vectors to be matched), V (information vectors to be extracted), and the scaled dot-product as the function to generate dynamic weights, which is also used to process sequences containing different lengths.
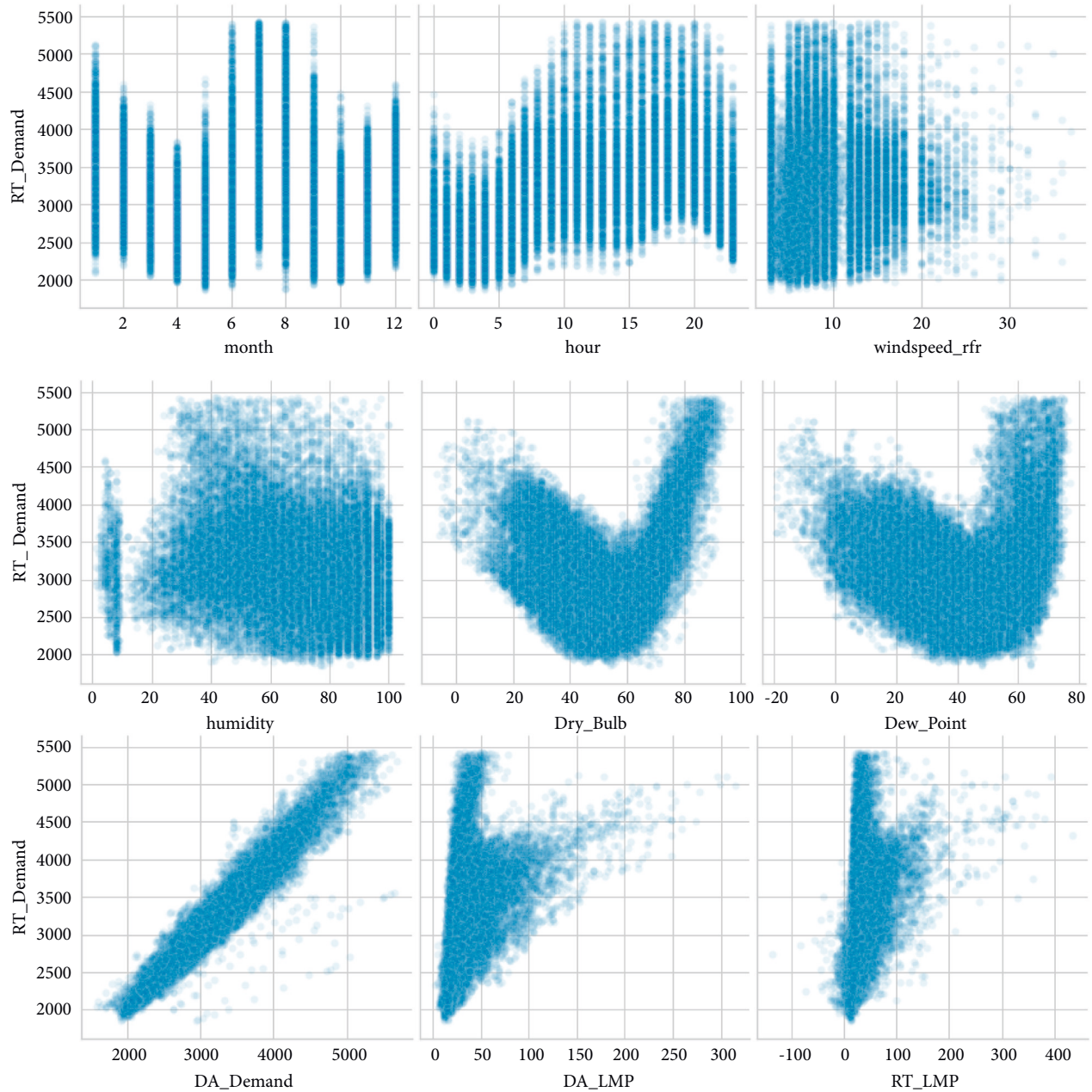
FIGURE 8: The overall analysis.

The transformer model adopts multihead self-attention and positional encoding methods, so it can process information from different locations of the expression subspace synchronously at different locations. Multihead self-attention combined with some self-attention layers provides joint computation. Moreover, as the model does not contain convolutional and recursive units, to learn from the sequences, it must make use of the information about the sequence positions. The positional code layer appears only after the embedding on the encoder side and the decoder side and before the first block. In the absence of this, the transformer model does not work. Position encoding is a unique component of the transformer framework that complements the fact that the attention mechanism itself cannot capture positional information.

In summary, multihead self-attention can be expressed as shown in the following equation:
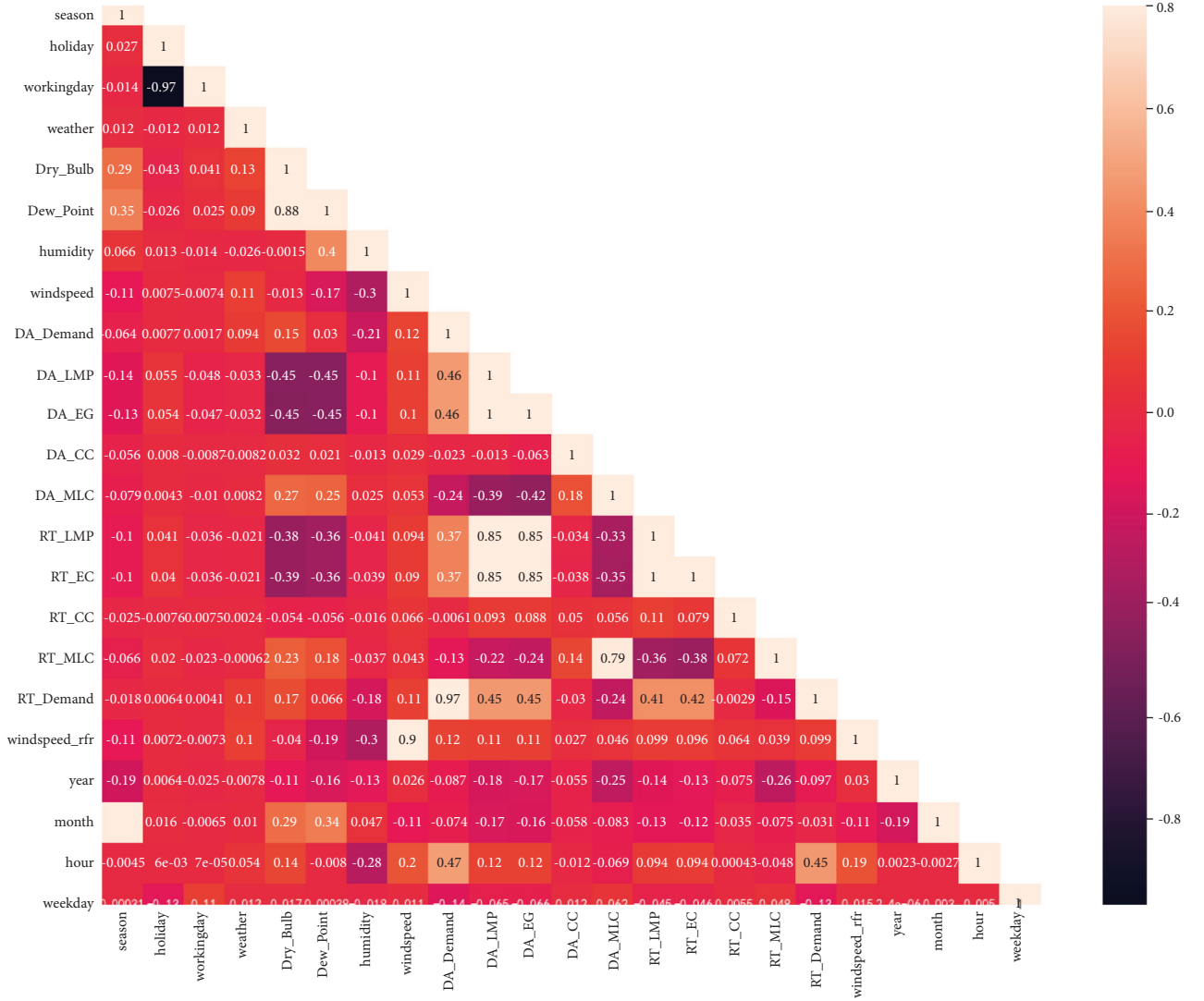
Figure 9: Data correlation analysis.

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \ldots, \text{head}_n)W^0,$$
$$\text{where head}_i = \text{Attention}\left(QW_i^Q, KW_i^K, VW_i^V\right),$$

$$(6)$$

where $W_i^Q \in \mathbb{R}^{d_x \times d_q}$, $W_i^K \in \mathbb{R}^{d_x \times d_k}$, $W_i^V \in \mathbb{R}^{d_x \times d_v}$, and $W_i^o \in \mathbb{R}^{hd_v \times d_x}$.

2.3. LightGBM Model. LightGBM improves the performance of the GBDT, and the optimal segmentation point based on the histogram algorithm and leaf-wise decision tree growth method with depth limitation is adopted. The method eliminates most samples with a small weight in the training process from the perspective of sample reduction, and only the information gain is calculated for the remaining sample data [26–28].

The histogram optimization algorithm needs to convert feature values into bin values in advance before training, that is, make a piecewise function for the value of each feature, divide the values of all samples on this feature into a certain segment (bin), and finally convert feature values from continuous values to discrete values. An intuitive example is shown in Figure 2.

There are many advantages to using histogram algorithms. First, it reduces memory consumption, and it can only store values after feature discretization. Thus, the cost of computing is dramatically reduced. The histogram algorithm only needs to calculate k, which can be considered as a constant and the number of segments. This way, the calculated time complexity measure can be reduced from O (data × feature) to O (k × feature).

The histogram difference only needs to traverse k buckets of the histogram. LightGBM can construct a histogram of a leaf (the parent node is calculated in the previous round) and get a histogram of its brother leaf at a fraction of the cost, doubling its speed, as shown in Figure 3.

LightGBM adds a maximum depth limit to the leaf-wise method to ensure high efficiency while preventing over-fitting, as shown in Figure 4.
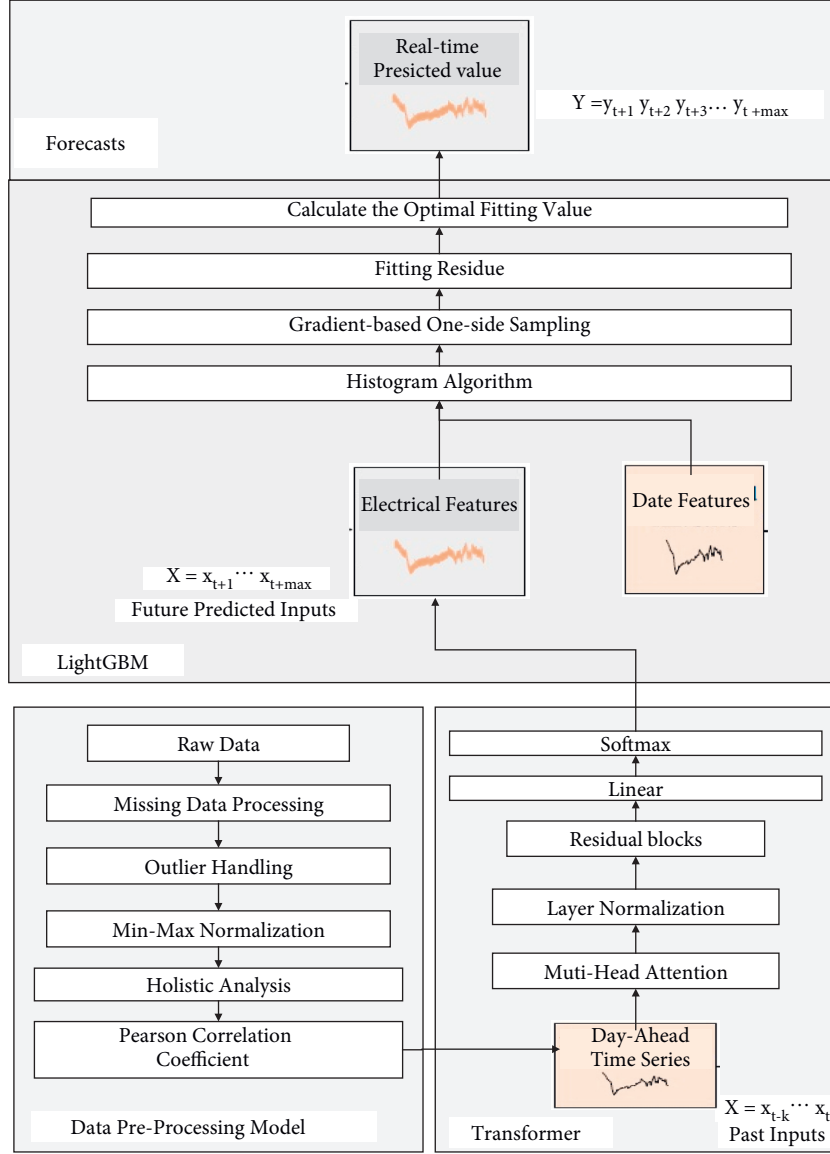
Figure 10: The medium-term power consumption forecasting framework based on the transformer-lightGBM method.

The gain function (G) of the splitting characteristic (j) at point v is defined, as shown in the following equation:

$$
\begin{cases}
G_j(v) = \dfrac{1}{S}\big(G_{j1}(v) + G_{j2}(v)\big), \\[2ex]
G_{j1}(v) = \dfrac{1}{S_l}\left(G_{X_k \in P_l}g_k + \dfrac{1-a}{b}\sum_{X_k \in Q_l}g_k\right)^2, \\[3ex]
G_{j2}(v) = \dfrac{1}{S_h}\left(G_{X_k \in P_h}g_k + \dfrac{1-a}{b}\sum_{X_k \in Q_h}g_k\right)^2,
\end{cases}
\tag{7}
$$

where P is the set of extraction, Q is the set of random extraction, a and b are the extraction ratios, S is the value of $P \cup Q$, S_l is the set of data less than v, S_h is the set of data greater than v, and g_k is the opposite direction gradient of X_k.
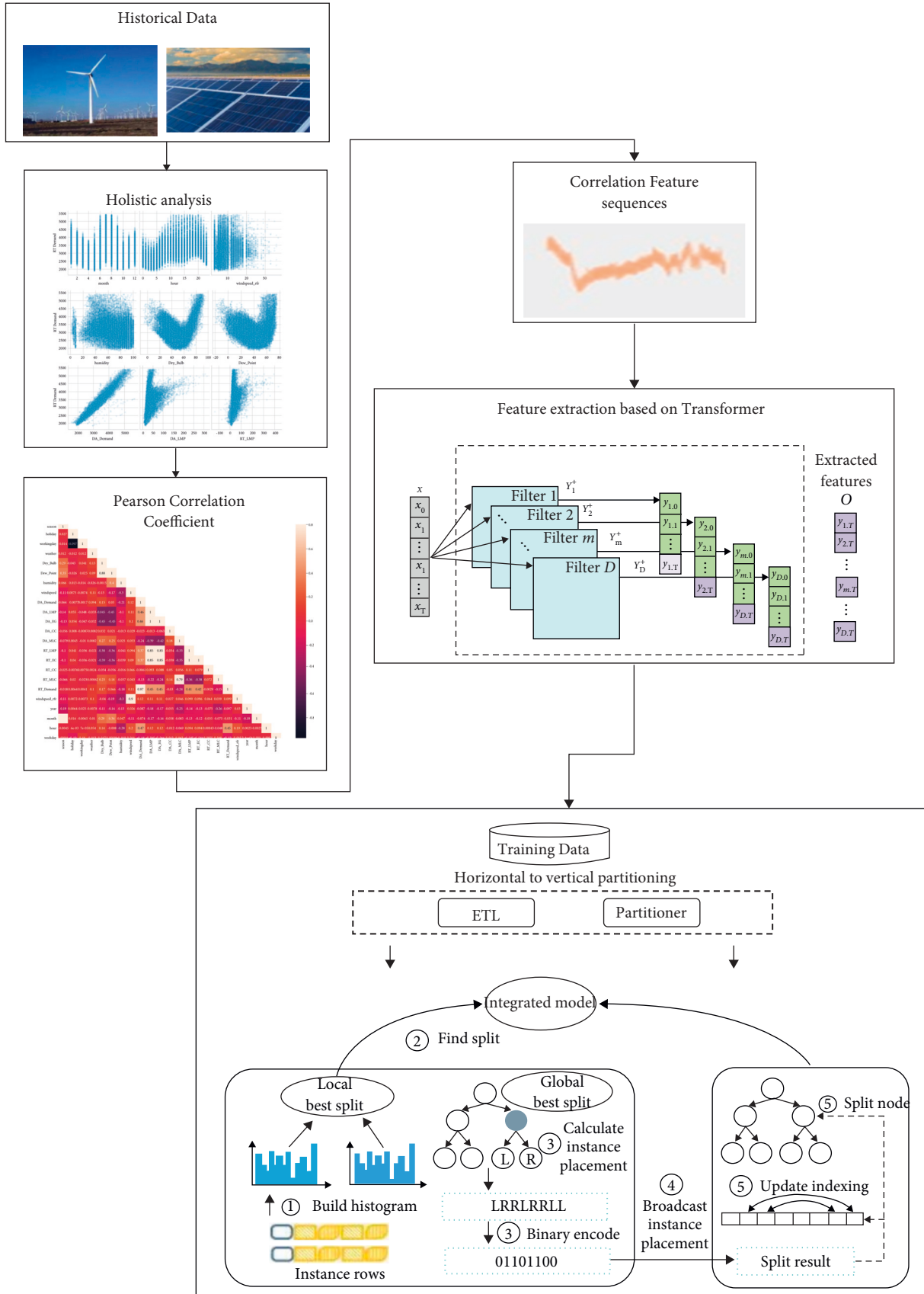
Figure 11: The transformer-lightGBM method processing steps for power markets.

TABLE 1: The transformer-lightGBM parameter settings.

Model		Parameters	Value
Transformer		num_layers	2
		d_model	256
		num_heads	2
		dff	256
		input_vocab_size	8500
		maximum_position_encoding	6
		optimizer	Adam
		learning_rate	0.0001
		loss	mae
		metrics	mse
LightGBM	Estimator	objective	Regression
		boosting_type	gbdt
		n_estimators	81
	GridSearchCV (optimize the parameters method)	metric	rmse
		boosting_type	gbdt
		objective	Regression
		learning_rate	0.3
		num_leaves	50
		max_depth	17
		subsample	0.8
		colsample_bytree	0.8
		max_depth	range (10, 30, 5)
		num_leaves	range (50, 170, 30)
		learning_rate	[0.3, 0.25, 0.2, 0.15, 0.1, 0.05, 0.01]
		feature_fraction	[0.5, 0.6, 0.7, 0.8, 0.9]
		bagging_fraction	[0.6, 0.7, 0.8, 0.9, 1.0]
		subsample	1
		min_samples_split	2
		min_samples_leaf	1
		num_leaves	110
		max_depth	10
		learning_rate	0.1
		feature_fraction	0.8
		bagging_fraction	0.8
		bagging_freq	10
		num_boost_round	531
		early_stopping_rounds	200

TABLE 2: Real-time demand data and related data tables.

Data	Input variable t
t real-time cleared demand	t-744/2208 day-ahead cleared demand t-744/2208 real-time cleared demand t-744/2208 day-ahead cleared LMP t-744/2208 real-time cleared LMP
Transformer preprocessing features	$t+744/2208$ day-ahead cleared demand (transformer predicted value) $t+744/2208$ day-ahead cleared LMP (transformer predicted value)
Date features	Hour Month Holiday

2.4. The Power Consumption Forecasting Framework Based on the Transformer-LightGBM Method

2.4.1. Data Processing. The dataset contains data from the ISO-NE, from a selected Connecticut region [29]. Meteorological data are from [30]. The data are from January 1, 2016, to March 31, 2021, with 46009 data records.

The seasonal decomposition of the real-time demand series is analyzed using the 2020 dataset as an example, and the original, trend, seasonal, and residual series are shown in Figure 5.

The real-time power demand density distribution is shown in Figure 6. The data are still volatile, and we want the fluctuations to be relatively stable; otherwise, it will be easy

to produce overfitting, so we want to process the data to make it relatively stable. Here, we choose logarithmic variation to make the data stable, as shown in Figure 7. After logarithmic transformation, the data distribution is more uniform, and the size difference is also reduced. Using such labels is effective for the training model.

From the data-driven method for medium-term power consumption forecasting, we analyze the ISO-NE hourly dataset as a whole:

(i) *Date Characteristics.* Month and hour.

(ii) *Meteorological Feature.* Humidity, dry bulb temperature, dew point temperature, and wind speed.

(iii) *Power Market Characteristics.* Day-ahead cleared demand, day-ahead LMP (locational marginal price), and real-time LMP.

The overall analysis is shown in Figure 8.

The data correlation analysis is shown in Figure 9.

From the data correlation, it can be seen that the highest correlation coefficient is 0.97 for real-time demand and day-ahead cleared demand, 0.45 for day-ahead LMP, and 0.45 for per hour. So, we analyze the sequence features of day-ahead cleared demand, day-ahead LMP, and per hour as the main feature.

2.4.2. Medium-Term Power Consumption Forecasting Based on Transformer-LightGBM. In this study, we consider the deep relationship between the real-time market demand and the day-ahead market demand and then design a multivariate model to predict the power consumption in the next month and the next quarter based on deep learning. We consider extracting the sequence features of the power spot market by the transformer network. LightGBM gives the final prediction.

The overall research framework is shown in Figure 10, and the processing schematic diagram is shown in Figure 11. The framework consists of three steps: data preprocessing, feature extraction based on the transformer, and load prediction based on lightGBM.

In this study, the transformer-lightGBM parameter settings are listed in Table 1. A cross-validated grid search method is used to optimize the parameters of the estimator for lightGBM.

Processing of the real-time demand data and related data is listed in Table 2.

The method constructed in this study offers two main improvements on the dataset feature extraction and processing:

(i) The prediction of the long time series of real-time loads in the power market is transformed into supervised learning in machine learning; that is, the results of data of length T are predicted for time points $T + 744$ and $T + 2208$ using data of length T. After obtaining data with satisfactory accuracy, the daily and monthly electricity consumption is calculated based on real-time loads.

Table 3: Parameter setting for comparison methods.

Model	Parameters	Value
CNN-LSTM	input	0
	conv1d	48
	max_pooling1d	0
	dropout	0
	dense	17
CNN-BiLSTM	input	0
	conv1d	48
	max_pooling1d	0
	dropout	0
	bilstm	4224
	multiply	0
	dense	33
CNN-ATT-LSTM	input	0
	conv1d	48
	max_pooling1d	0
	dropout	0
	attention_vec	1056
	multiply	0
	dense	33
CNN-ATT-BiLSTM	input	0
	conv1d	48
	max_pooling1d	0
	dropout	0
	bilstm	4224
	attention_vec	1056
	multiply	0
	dense	33

Table 4: Real-time demand forecasting evaluation on the testing set.

Time	From March 1, 2021, to March 31, 2021		
Statistical metrics	γmae (kW)	γmape (%)	γrmse (kW)
CNN-LSTM	300.89	11	359.64
CNN-Bi LSTM	312.22	11.36	384.87
CNN-ATT-LSTM	306.96	11.21	374.71
CNN-ATT-BiLSTM	297.89	10.68	380.56
Proposed method	273.59	9.33	330.09

(ii) The transformer-lightGBM method is built to effectively combine the variable timing processing capabilities of the multihead self-attentive mechanism with the ensemble learning model to increase algorithmic accuracy, speed, and generalization capabilities.

3. Experimental Results and Analysis

The experimental environment is as follows: Python 3.7, Tensorflow 2.3 GPU, NVIDIA GeForce 940MX graphics card, Intel i5, 64 bit, and 12 GB RAM. The datasets for experiment 1 and experiment 2 are annual hourly data from ISO-NE, spanning 1916 days from January 2016 to March 2022.

In this experiment, parameter setting for comparison methods is listed in Table 3.
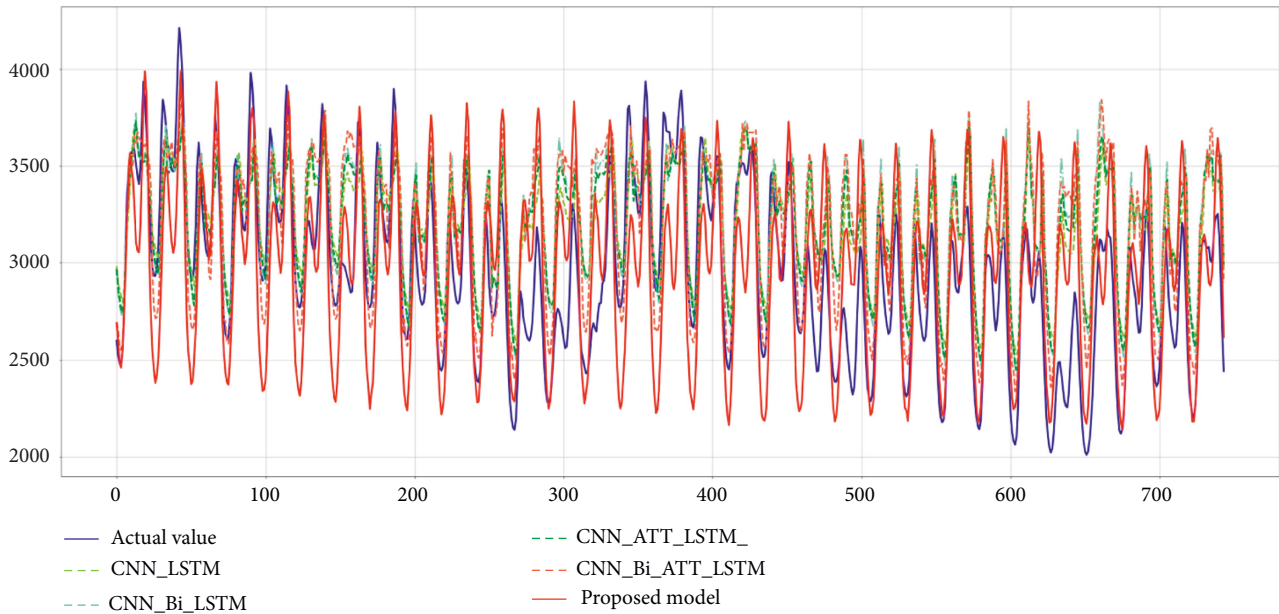
FIGURE 12: Real-time demand forecast from March 1, 2021, to March 31, 2021.
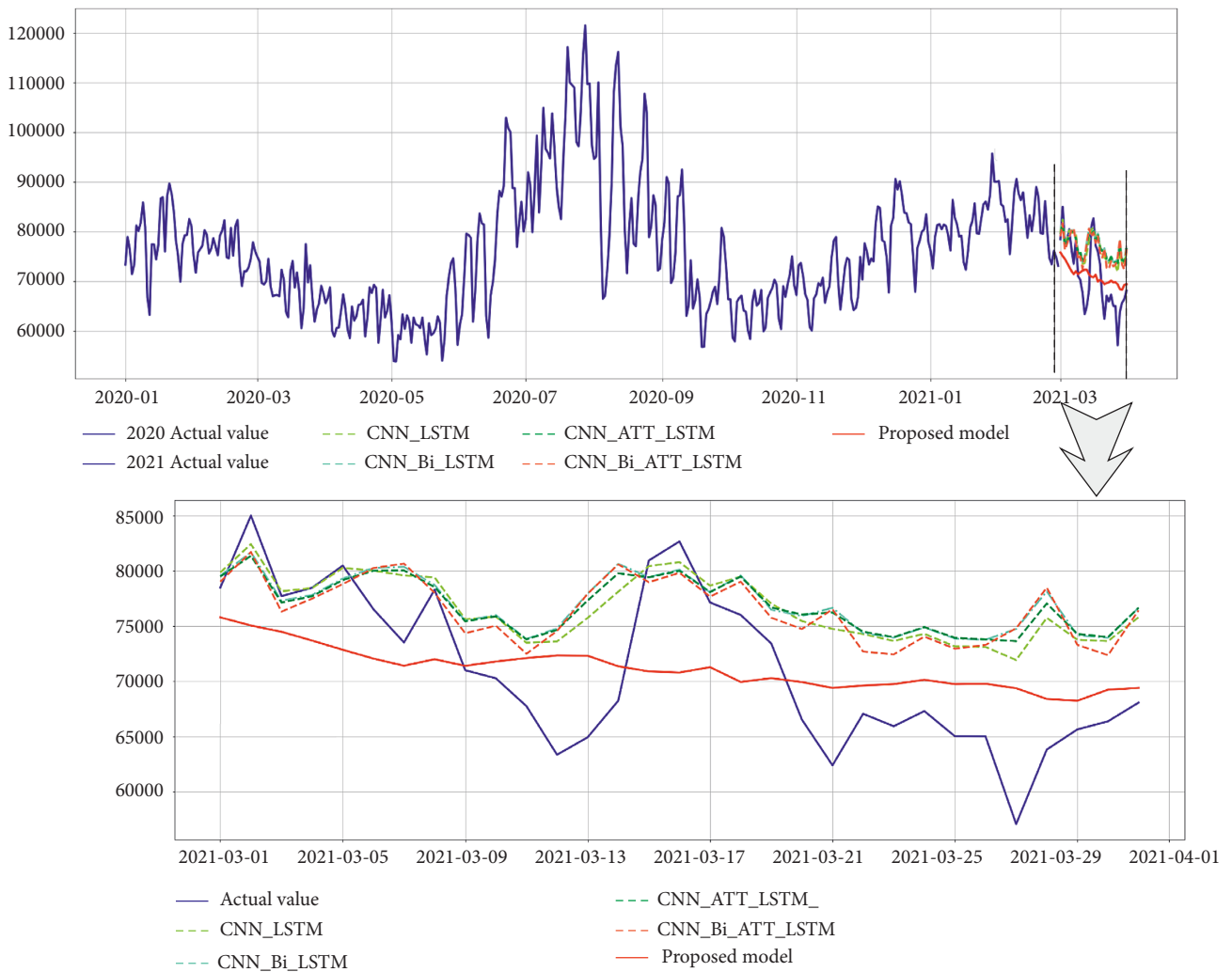

FIGURE 13: Power consumption forecast for the next month from January 1, 2021, to March 31, 2021.

TABLE 5: Real-time demand forecasting evaluation on the testing set.

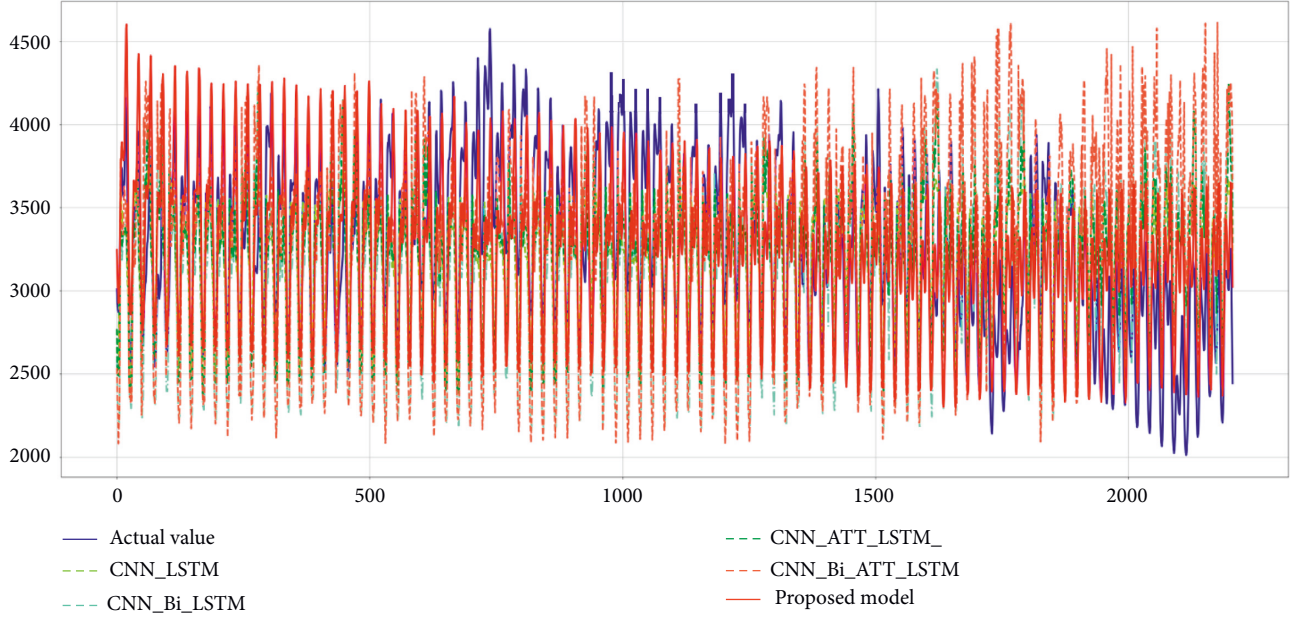| Time	From January 1, 2021, to March 31, 2021		
Statistical metrics	γmae (kW)	γmape (%)	γrmse (kW)
CNN-LSTM	353.86	11.12	432.17
CNN-BiLSTM	443.20	14.04	527.65
CNN-ATT-LSTM	367.80	11.52	444.96
CNN-ATT-BiLSTM	473.44	15.37	595.91
Proposed method	248.99	8.53	300.84

FIGURE 14: Real-time demand forecast from January 1, 2021, to March 31, 2021.

The APE in equation (8), the MAPE in equation (9), the RMSE in equation (10), and the MAE in equation (11) are used:

$$\text{APE} = \frac{\left| P_i^{\text{pre}} - P_i^{\text{real}} \right|}{P_i^{\text{real}}}, \tag{8}$$

$$\text{MAPE} = \frac{1}{N} \sum_{1}^{N} \frac{\left| P_i^{\text{pre}} - P_i^{\text{real}} \right|}{P_i^{\text{real}}}, \tag{9}$$

$$\text{RMSE} = \sqrt{\frac{\sum_{1}^{N} \left(P_i^{\text{pre}} - P_i^{\text{real}} \right)^2}{N}}, \tag{10}$$

$$\text{MAE} = \frac{\sum_{1}^{N} \left| P_i^{\text{pre}} - P_i^{\text{real}} \right|}{N}. \tag{11}$$

Pipre means the predicted data, Pireal means the actual value, and N means the totality.

3.1. Experimental 1: Power Consumption Monthly Forecast.

In the experiment, power consumption is forecasted for 744 points in March 2021. Datasets from January 1, 2016, to January 31, 2021, are used as the training datasets, datasets from February 1, 2021, to February 8, 2021, are used as the validation datasets, and datasets from March 1, 2021, to March 31, 2021, are used as the testing datasets. The prediction results of methods are listed in Table 4. Figure 12 shows the comparison of the real-time demand for each method from March 1, 2021, to March 31, 2021.

It shows the presented approach can provide a better fit to the actual demand in the forecast curve. In especial, the presented method can capture occasional fluctuations. However, the other compared methods show different deviations in the form of sudden changes or peaks. From the experiment 1 results, it can be seen that the presented method improves the prediction accuracy of real-time demand for the coming month.

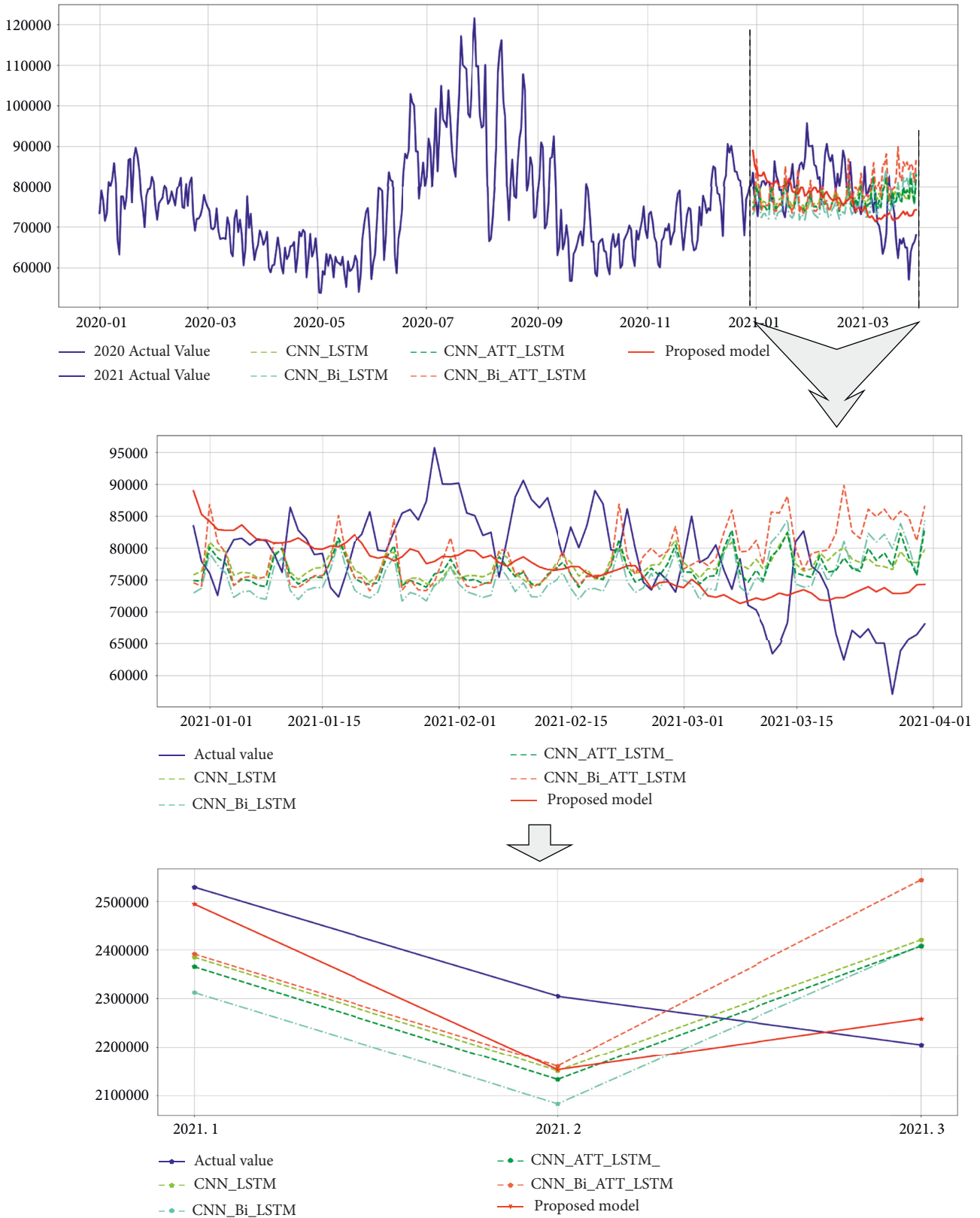Based on the predicted real-time demand at 744 hourly points in the future, we can draw the daily total power

Figure 15: Power consumption forecast for the next quarter from January 1, 2021, to March 31, 2021.

consumption curve for the next months (March 1, 2021, to March 31, 2022), as shown in Figure 13.

3.2. Experimental 2: Power Consumption Quarterly Forecast. In this experiment, real-time electricity demand is forecasted for 2208 points in March 2021. The training set runs from January 1, 2016, to September 31, 2020, the validation set runs from October 1, 2020, to December 31, 2020, and the test set runs from January 1, 2021, to March 31, 2021.

The prediction results of each model are listed in Table 5. Figure 14 shows the comparison of the demand from January 1 to March 31, 2021. It shows the presented method can provide a better fit to the actual demand power in the forecast curve. From the experiment 2 results, it can be seen that the method proposed in this study improves the prediction accuracy of power consumption for the next quarter.

Based on the predicted real-time power demand at 2208 hourly points in the future, we can draw the daily total power consumption curve for the next three months (January 1, 2021, to March 31, 2022), as shown in Figure 15.

4. Conclusion

In this study, we have presented a novel data-driven approach to forecast medium-term power consumption with transformer-lightGBM, which is summarized as follows:

(1) A data-driven approach was adopted to analyze the data of the ISO-NE, and we derived monthly and quarterly power consumption from hourly demand forecasts. Our proposed method allowed for multi-timescale forecasting by paying the attention to selected features.

(2) A novel method for medium-term power consumption forecasting based on transformer-lightGBM was designed and improved. We used a model architecture consisting of monthly and quarterly forecasts.

(3) The experimental results show that the proposed method verifies more accurate predictions than LSTM-based methods, such as CNN-LSTM, CNN-BiLSTM, CNN-ATT-LSTM, and CNN-ATT-BiLSTM.

Abbreviations

LSTM: Long short-term memory neural network
Transformer: Transformer neural network
LightGBM: Light gradient boosting machine
ISO-NE: US New England electricity market
MSE: Mean squared error
MAE: Mean absolute error
MAPE: Mean absolute percentage error
RSME: Root mean square error
HMM: Hidden Markov model
GRU: Gated recurrent unit
RNN: Recurrent neural network
BDL: Bayesian deep learning
SVM: Support vector machine
CNN: Convolutional neural network
CRBM: Conditional restricted Boltzmann machine
PCC: Pearson correlation coefficient
GBDT: Gradient boosting decision tree
WNN: Wavelet neural network
CNN-LSTM: Convolutional and long short-term memory neural network
CNN-BiLSTM: Convolutional and bidirectional long short-term memory neural network
CNN-ATT-LSTM: Convolutional and LSTM with attention mechanism neural network
CNN-ATT-BiLSTM: Convolutional and bidirectional LSTM with attention mechanism neural network.

Data Availability

The data of the models and algorithms used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] S. Du, B. Wen, and C. Jiang, *Power Market*, pp. 1–9, China Electric Power Press, Beijing China, 2004.

[2] C. Kang and L. Yao, "Key scientific issues and theoretical research framework for power system high proportion renewable energy," *Power System Automation*, vol. 41, no. 9, 2017.

[3] C. Kang, Q. Xia, H. U. Zuohao, and B. Zhang, "New connotation of forecasting problem in power market," *Power System Automation*, vol. 28, p. 18, 2017.

[4] M. Shahidehpour, *Market oriented operation of power system*, pp. 13-14, China Electric Power Press, Beijing, China, 2005.

[5] Y. Guan, D. Li, S. Xue, and Y. Xi, "Feature-fusion-kernel-based Gaussian process model for probabilistic long-term load forecasting," *Neurocomputing*, vol. 426, pp. 174–184, 2021.

[6] N. Son, S. Yan, and J. Na, "Deep neural network and long short-term memory for electric power load forecasting," *Applied Sciences*, vol. 10, no. 18, p. 6489, 2020.

[7] S. Omaji, F. A. Alzahrani, R. J. U. Hussen Khan et al., "Towards modified entropy mutual information feature selection to forecast medium-term load using a deep learning model in smart homes," *Entropy*, vol. 22, no. 1, p. 68, 2020.

[8] M. Dong, J. Shi, and Q. Shi, "Multi-year long-term load forecast for area distribution feeders based on selective sequence learning," *Energy*, vol. 206, Article ID 118209, 2020.

[9] J. Liu and H. Zhao, "Medium-term load forecasting test and seasonal decomposition based onCointegrationGranger causality," *Automation of Electric Power Systems*, vol. 43, no. 1, 2020.

[10] Z. Zhang, Y. Du, H. Cui, and X. Lai, "Intelligent mid-long electricity load forecast method considering associated

factors," *Power System Protection and Control*, vol. 47, no. 2, 2019.

[11] Y. He, C. Wen, Q. Xu, and A. Han, "A method to predict probability density of medium-term power load considering temperature factor," *Power System Technology*, vol. 39, no. 1, 2015.

[12] G. J. Khatereh, D. Gharavian, and H. Mahdiani, "A novel method for day-ahead solar power prediction based on hidden Markov model and cosine similarity," *Soft Computing*, vol. 24, no. 7, pp. 4991–5004, 2020.

[13] Z. Shang, M. Li, Y. Chen, C. Li, Y. Yang, and L. Li, "A Novel Model Based on Multiple Input Factors and Variance Reciprocal: Application on Wind Speed forecastin Soft Computing," *Soft Computing*, vol. 26, Article ID 8877, 2022.

[14] A. Shobana Devi, G. Maragatham, K. Boopathi, and A. G. Rangaraj, "Hourly day-ahead wind power forecasting with the EEMD-CSO-LSTMEFG deep learning technique," *Soft Computing*, vol. 24, Article ID 12411, 2020.

[15] M. Xia, H. Shao, X. Ma, and W. Clarence, "A stacked GRU-RNN-based approach for predicting renewable energy and electricity load for smart grid operation," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 10, 2021.

[16] Y. Yang, W. Li, T. Aaron Gulliver, and S. Li, "Bayesian deep learning-based probabilistic load forecasting in smart grids," *IEEE Transactions on Industrial Informatics*, vol. 16, p. 7, 2020.

[17] M. Kojury-Naftchali, A. Fereidunian, M. Savaghebi, and B. Akhbari, "Change detection in electricity consumption patterns utilizing adaptive information theoretic algorithms," *IEEE Systems Journal*, vol. 15, no. 2, 2021.

[18] M. Florian, B. Rolle, M. Bauer, and O. Sawodny, "Forecasts of electric vehicle energy consumption based on characteristic speed profiles and real-time traffic data," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 2, pp. 1404–1418, 2020.

[19] A. M. Alonso, F. J. Nogales, and C. Ruiz, "Hierarchical clustering for smart meter electricity loads based on quantile autocovariances," *IEEE Transactions on Smart Grid*, vol. 11, no. 5, pp. 4522–4530, 2020.

[20] F. Qian, W. Gao, Y. Yang, and D. Yu, "Potential Analysis of the Transfer Learning Model in Short and Mediumterm Forecasting of Building HVAC Energy Consumption," *Energy*, vol. 193, Article ID 116724, 2019.

[21] A. Tanveer and H. Zhang, "Novel deep supervised ML models with feature selection approach for large-scale utilities and buildings short and medium-term load requirement forecasts," *Energy*, vol. 209, Article ID 118477, 2020.

[22] A. Nasir, M. Irfan, M. Awais et al., "Big data analytics for short and medium-term electricity load forecasting using an AI techniques ensembler," *Energies*, vol. 13, no. 19, p. 5193, 2020.

[23] A. Vaswani, N. Shazeer, and N. Parmar, "Attention Is All You Need," in *Proceedings of the 31st Conference on Neural Information Processing Systems*, LongBeach, CA, USA, December 2017.

[24] S. Li, X. Jin, and X. Yao, "Enhancing the Locality and Breaking the Memory Bottleneck of Transformer on Time Series Forecasting," in *Proceedings of the 33rd Conference on Neural Information Processing Systems (NeurIPS 2019)*, Vancouver, Canada, December 2019.

[25] B. Lima, S. O. Arkb, N. Loeb, and T. Psterb, "Temporal Fusion Transformers for Interpretable Multi-Horizon Time Series Forecasting," *International Journal of Forecasting*, vol. 37, 2020.

[26] G. Ke, "LightGBM: A highly efficient gradient boosting decision tree," in *Proceedings of the. Adv. Neural Inf. Proces. Syst*, pp. 3147–3155, LongBeach, CA, US, December 2017.

[27] R. Punmiya and S. Choe, "Energy theft detection using gradient boosting theft detector with feature engineering-based preprocessing," *IEEE Transactions on Smart Grid*, vol. 10, no. 2, pp. 2326–2329, 2019.

[28] X. Ma, J. Sha, D. Wang, Y. Yu, Q. Yang, and X. Niu, "Study on a prediction of P2P network loan default based on the machine learning lightGBM and XGboost algorithms according to different high dimensional data cleaning," *Electronic Commerce Research and Applications*, vol. 31, pp. 24–39, 2018.

[29] "US New England electricity market," 2021, https://www.isone.com.

[30] "Wunderground," 2021, https://www.wunderground.com.