Forecasting Model of Electricity Sales Market Indicators With Distributed New Energy Access

Tao Yao, State Grid Hebei Information and Telecommunication Branch, China
Xiaolong Yang, State Grid Hebei Information and Telecommunication Branch, China
Chenjun Sun, State Grid Hebei Information and Telecommunication Branch, China
Peng Wu, State Grid Energy Research Institute Co., Ltd., China
Shuqian Xue, Beijing Tsingsoft Technology Co., Ltd., China*

ABSTRACT

It is difficult for the existing electricity sales market to adapt to the vast amount of distributed new energy access. This article proposes an electricity sales market index prediction model for high proportion distributed new energy access under the cloud-side cooperation architecture. First, an index prediction system is designed based on the cloud edge collaboration architecture. The edge computing center processes regional data nearby to improve prediction efficiency. Second, on the edge side, a K-means clustering algorithm is used to classify the data. Third, the power data, distributed power output data, load data, weather data, holiday information, and electricity price data are obtained. Finally, the ConvLSTM-Adaboost prediction model is built in the cloud center. The ConvLSTM is used as the base learner, and the Adaboost-integrated algorithm is used for serial training. At the same time, the prediction results of each base learner are weighted and integrated to obtain the final power and load prediction results of the electricity sales market. Experiments show that the prediction results of MAE, PMSE, and MAPE of the proposed model for daily electricity are 52.539MW, 56.859MW, and 2.063%, respectively. Not only is this superior to other models, but it provides a better analysis of influencing factors.

KEYWORDS

Cloud Edge Collaboration, Convlstm-Adaboost Prediction Model, Distributed New Energy, Electricity Forecast, Electricity Sales Market, K-Means Clustering, Load Forecasting

INTRODUCTION

China has actively responded to the carbon peak and carbon neutral carbon emission reduction goals to accelerate the green transformation of energy. Under the dual-carbon background, the traditional power system will usher in comprehensive transformation and upgrades (Singla et al., 2021). Among them, distributed power generation offers advantages like cleanliness, efficiency, and local balancing

DOI: 10.4018/IJITSA.326757 *Corresponding Author

in accordance with the development of China's future power grid (David et al., 2021). Under the combined effect of objective resource conditions and policy promotion, distributed power generation in all regions of China has shown rapid development.

Distributed power can be freely installed at the distribution network terminals. Through the spontaneous self-use of users, its remaining power can be returned to the distribution network to achieve efficient energy consumption and utilization (López et al., 2022). Still, due to the dynamic energy balance, once the power supply is surplus, it will intensify the phenomenon of abandoning distributed new energy. Therefore, when selecting the grid-connection scheme of new energy according to the grid's acceptance capacity of wind and photovoltaic power (WPV) in various regions, accurate power prediction is indispensable (Lu et al., 2022). In the new power grid, the massive access of clean energy power generation, such as WPV, will increase the pressure of power dispatching (Bian et al., 2022). Due to the characteristics of distributed new energy, its power generation form has obvious volatility, intermittence, and unpredictability, creating new challenges to the accurate prediction of power consumption in regional power systems.

Regional electricity consumption is characterized by multiple factors, increased uncertainties, and complex changes due to the improvement of consumer terminals' demand of the new power system and continuous access of new energy sources. The accurate prediction of electricity quantity can provide a reliable basis for the planning and construction of a power grid, optimal dispatching, and optimal load distribution. This poses a challenge to the accurate prediction of regional electricity (Luo et al., 2022). The research results of electricity forecasting are mainly divided into traditional prediction and artificial intelligence prediction (AIP). Among them, AIP methods include tree integration algorithm, support vector machine, and neural network algorithm (Dab et al., 2022; Kalhori et al., 2022; Tan, 2022). Both traditional methods and AIP methods do not fully consider the deep mining and utilization of data under the modern powerful Internet of things (IoT), which limits the highest accuracy of the prediction algorithm.

Based on the above analysis and aiming to accurately predict the electricity consumption, an indicator prediction model under the cloud-side collaborative architecture is proposed. Compared with traditional models:

1. It aims to improve the processing efficiency of the massive distribution of new energy access in the electricity sales market. Thus, the proposed model builds a prediction system based on the cloud-side collaborative architecture. It completes the data classification processing by deploying K-means clustering algorithm on the edge side to improve data quality.
2. The LSTM model lacks the analyzing ability of data space characteristics. Thus, the proposed model uses the ConvLSTM model for data learning and the Adaboost integration strategy to weigh and combine the ConvLSTM base learners. This greatly improves the universality of the model.

The content of the remaining sections of the article is as follows. Section 2 includes related research to summarize existing results. Section 3 is the design of an electricity sales market indicator prediction system based on cloud edge collaboration. It introduces the system architecture and data classification based on K-means clustering algorithm on the edge side. Section 4 is the prediction of electricity sales market indicators based on the ConvLSTM-AdaBoost model in the cloud center. The ConvLSTM-AdaBoost model is constructed to achieve indicator prediction. Section 5 is the experimental results and analysis, which demonstrates the effectiveness of the proposed model. Section 6 presents the conclusion and outlook.

RELATED RESEARCH

According to different classification standards and rules, the prediction methods can be divided into traditional and artificial intelligence (Meng, 2022). Artificial intelligence algorithms are mainly represented by neural network algorithms (Zhu et al., 2022). For the distribution network containing distributed energy, its load forecasting is mainly aimed at the short-term load forecasting, such as daily load. For example, Pereira et al. (2022) integrated the Monte Carlo (MC) and Copula model to study the characteristics of new energy generation. The prediction model is optimized based on the genetic algorithm, which effectively improves the prediction accuracy. However, the Copula model (optimized by the genetic algorithm) is more traditional and, thus, difficult to adapt to complex electricity sales market predictions.

Wang et al. (2022) posed a challenge to high-precision load forecasting due to the increasing permeability and uncertainty of DEN. Therefore, the wind, solar, and power load forecasting methods are comprehensively reviewed, and relevant correction methods are introduced. Only existing prediction methods are outlined. Evangelopoulos and Georgilakis (2022) proposed a probabilistic load forecasting method. First, the hierarchical trend method is used to predict the peak load in each zone of the area under study. The spatial error of the entire service area is calculated to build the prediction interval. The algorithm is improved by combining with the MC simulation of probabilistic power flow. This method can effectively achieve load forecasting, but its effectiveness needs to be improved for power grid forecasting with high proportion of distributed new energy access. Therefore, traditional prediction methods lack accurate feature extraction and significant analytical performance models as support, making it difficult to adapt to complex and ever-changing new power systems. In turn, they are unable to reliably predict indicators.

There has been continuous development of intelligent algorithms like deep learning. The advantages of algorithms and deep neural network in the field of load forecasting have also been promoted (Goli et al., 2021). Among them, Zhang et al. (2022) proposed a daily load forecasting model that improves the forecasting accuracy by combining long short-term memory (LSTM) and wavelet transform. This method considers both prediction accuracy and prediction efficiency. Banga et al. (2021) used stack integration and IoT to design an accurate power consumption prediction model to improve the efficiency and accuracy of household load forecasting through data processing in the two stages of prediction model and evaluation model. Hong et al. (2022) proposed a model based on a hybrid convolutional neural network (CNN) to predict the daily peak load one week in advance. It optimizes the network topology and super parameters of the hybrid CNN through the genetic algorithm, effectively improving the accuracy and computational efficiency of the algorithm. Wang et al. (2021) proposed a load forecasting technology based on sequence-to-sequence and gated recurrent unit (GRU). Based on the encoding and decoding framework compatible with input and output data sequences with variable lengths (and combined with BA mechanism), it effectively solves the previous information loss of GRU. Niu et al. (2022) proposed a CNN-BiGRU model for load forecasting through the attention mechanism. This model effectively realizes the overall optimization of the algorithm.

The above methods only considered existing data characteristics, including load size, time characteristics, and other information in the prediction process. There is, however, limited in-depth research on the deep mining of big data under the IoT architecture. Therefore, a prediction model of electricity sales market indicators for high proportion distributed new energy access under the cloud-side collaborative architecture is proposed. It effectively improves the accuracy of load forecasting and provides a theoretical basis for distributed energy access.

DESIGN OF ELECTRICITY SALES MARKET INDEX PREDICTION SYSTEM BASED ON CLOUD-EDGE COLLABORATION

Proposed System Architecture

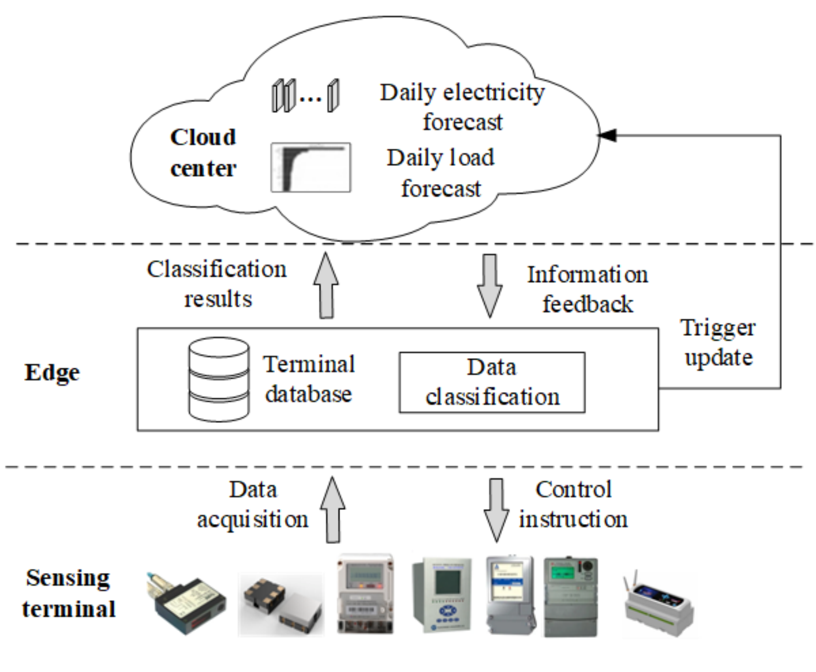
The power grid has gradually integrated the cloud edge collaboration (CEC) concept of the IoT, carried out the application research, and built a cloud-edge data processing mode. Due to the widespread access to distributed new energy, traditional centralized IoT architectures are no longer fully applicable. However, cloud edge collaborative architectures marginalize data processing and are closer to data sources, enabling faster execution of system tasks. Based on this, the proposed model uses the cloud edge collaboration architecture to design the electricity sales market index prediction system. The architecture, as shown in Figure 1, is composed of sensing terminal, edge side, and cloud center.

The cloud center is responsible for the in-depth analysis of data and construction of ConvLSTM-AdaBoost model to predict indicators of the electricity sales market. As an edge node, the intelligent fusion terminal in the electricity sales market performs online monitoring, storage, and classification of the system operation status (Peng et al., 2022). The data of the electricity sales market is diversified, dynamic, and differentiated. It includes distributed new energy generation data, load data, equipment status data, and meteorological data. In addition, the data changes in real time.

DATA CLASSIFICATION BASED ON K-MEANS CLUSTERING ALGORITHM AT THE EDGE

There is an array of data sources and a large amount of data in the electricity sales market. To improve the reliability of prediction, clustering algorithms are deployed in the edge computing server to complete the data preprocessing. The K-means algorithm divides the sample set X into K clusters (Qin et al., 2022). K-means clustering algorithm is then used on the edge side to classify the electricity

Figure 1. Architecture of electricity sales market index prediction system based on cloud-side collaboration



sales market data to obtain electricity data, distributed power output data, load data, weather data, holiday information, and electricity price data.

In K-means clustering, suppose the cluster is c_i and the mean vector $\overline{c}_i$ of c_i is:

$$\overline{c}_i = \frac{1}{c_i}\sum\nolimits_{x \in c_i} x_i \tag{1}$$

Then, the minimum square error E is:

$$E = \sum\nolimits_{i=1}^{k}\sum\nolimits_{x \in c_i}\left\| x - \overline{c}_i \right\|_2^2 \tag{2}$$

After several experiments, the prediction effect is best when the K value is 8. Therefore, the clustering process of K-means is:

1. Select eight samples randomly from the sample set as the initial centroid vector: $\left\{s_1, s_2, \cdots, s_8\right\}$.
2. Calculate the Euclidean distance from each point to the eight cluster centers. Then, divide the point into the nearest cluster and cluster center.
3. Recalculate the mean value of each cluster until the mean value does not change. Output the clustering result: $C = \left\{c_1, c_2, \cdots, c_8\right\}$.

K-means algorithm clusters the components with similar morphology. It adds, fuses, and enhances the components with consistent classification. This will reduce the impact of noise on the overall trend and improve the prediction accuracy. In addition, the number of components is reduced by the fusion, which can enhance the training and prediction speed of the model.

FORECAST OF SALES MARKETING INDICATORS BASED ON CONVLSTM-ADABOOST MODEL IN CLOUD CENTER

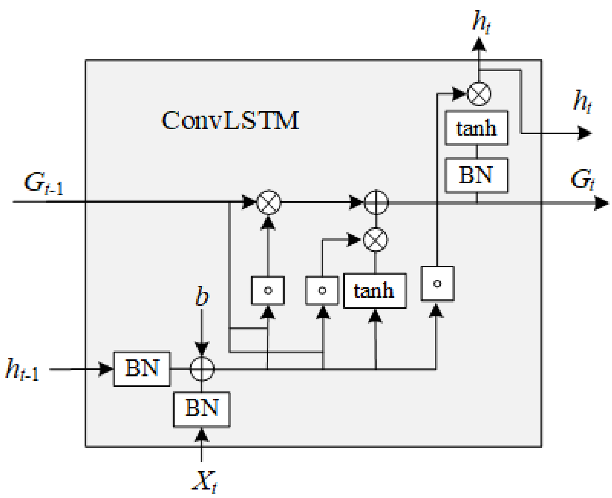
ConvLSTM Model

As a special recurrent neural network, LSTM can make more reasonable use of information in the time dimension (Yang, Guo et al., 2022). However, the LSTM network lacks the ability to process spatial information. Therefore, the CNN-LSTM model is obtained by integrating CNN and the spatial continuity features of information are extracted by CNN convolution (Huang et al., 2022). To simplify the CNN-LSTM model, the proposed model only uses the convolution layer and LSTM for data analysis. The resulting ConvLSTM model structure is shown in Figure 2. ConvLSTM model is widely used in weather and load forecasting by capturing the all-round correlation of data.

ConvLSTM model is composed of memory cell, input, output, and forgetting gate. The difference between the LSTM network and LSTM network is that the connection between input and each gate is replaced by the convolution from the feedforward. The convolution operation is also replaced between states (Jiang, 2022). Regarding the convolution operation, the temporal relationship can be obtained and the spatial features can be extracted (for example, the convolution layer). The mathematical expression of ConvLSTM is as follows:

$$i_t = \delta\left(\omega_{xi} \cdot X_t + \omega_{hi} \cdot H_{t-1} + \omega_{ci} \circ G_{t-1} + b_i\right) \tag{3}$$

Figure 2. Structure of ConvLSTM model



$$f_t = \delta\left(\omega_{xf} \cdot X_t + \omega_{hf} \cdot H_{t-1} + \omega_{gf} \circ G_{t-1} + b_f\right) \tag{4}$$

$$G_t = f_t \circ G_{t-1} + i_t \circ \tanh\left(\omega_{xg} \cdot X_t + \omega_{hg} \cdot H_{t-1} + b_g\right) \tag{5}$$

$$o_t = \sigma\left(\omega_{xo} \cdot X_t + \omega_{ho} \cdot H_{t-1} + \omega_{go} \circ G_{t-1} + b_o\right) \tag{6}$$

$$H_t = o_t \circ \tanh\left(G_t\right) \tag{7}$$

where i_t, f_t, G_t, and H_t represent input gate, forgetting gate, transition state, output gate, and current state, respectively. ω is the weight coefficient matrix corresponding to the convolution kernel. b refers to the paranoia against the door. $\cdot$ represents convolution, is Hadamard product, δ and tanh are the activation functions, and σ is the sigmoid function.

ConvLSTM-AdaBoost Model

AdaBoost algorithm is an integrated learning technology framework that trains different base learners for the same training set. Then, it sets the base learners to form a stronger final learner (Shahare et al., 2023; Yang, Shi et al., 2022). The proposed prediction model uses the ConvLSTM model as a base learner. It is trained serially through the AdaBoost integrated algorithm and adjusts the sample distribution and weight coefficient of each base learner. For the samples whose prediction error is greater than the set threshold in the training process, the researchers increase their weight. Otherwise, the researchers reduce their weight. Finally, the prediction results of each base learner are weighted and integrated to obtain the final prediction results.

The AdaBoost algorithm is used assuming that there are n data $D=\{(x_1,y_1),(x_2,y_2),\cdots,(x_n,y_n)\}$ given ConvLSTM as the basic learner model. At the beginning of the algorithm, all data are given the same weight value ($1/n$) and the number of basic learners is set to M. For the prediction problem, the researchers determine the base learner of the AdaBoost integration algorithm. Then, they initialize the weight of the data set. The training set is adopted to train the first base learner ConvLSTM and to calculate the prediction error rate of the training set on the base learner $\varepsilon 1$. The weight ϖ of training samples is updated according to the performance of prediction error rate. Thus, the weight of sample points with a higher prediction error rate of the base learner $q1$ becomes higher. In the base learners $q2$, these points with higher prediction error rates can receive more attention. The base learners $q2$ are then trained. This operation is repeated operation until all basic learners are trained. Then, the results are integrated to obtain the final strong learner.

The prediction process of ConvLSTM-AdaBoost is shown in Algorithm 1.

Algorithm 1. Pseudo Code for ConvLSTM-AdaBoost Prediction Process

Input:
Training Set $D=\{(x_1,y_1),(x_2,y_2),\cdots,(x_n,y_n)\}$, Base learner Γ_{C-LSTM}, Number of training rounds T, …
Begin
1. Data weight initialization: $W_1=(\varpi_{11},\cdots,\varpi_{1i},\cdots,\varpi_{1n})$, $\varpi_{1i}=1/n, i=1,2,\cdots,n$
2. **For** m = 1, 2, …, M
   do Use training set with weight distribution to train base learners:
3. $q_m=\Gamma_{C-LSTM}(D,W_m)$
   where W_m is the sample weight of the m-th base learner.
4. Calculate the linear prediction error rate ε_m of the training set on $q(x)$:
   $$\begin{cases}\varepsilon_m=\sum_{i=1}^{n}\varpi_{mi}e_{mi}\\ e_{mi}=\dfrac{|y_i-q_m(x_i)|}{\max|y_i-q_m(x_i)|}\end{cases}$$
   where e_{mi} is the relative error on each sample; $q_m(x_i)$ is the predicted value of the i-th sample on the m-th base learner; y_i is the actual value of the i-th sample.
5. Calculate the coefficient β_m of $q_m(x)$:
   $\beta_m=(1-\varepsilon_m)/\varepsilon_m$
6. Update sample weights:
   $$W_{m+1}(x)=\frac{W_m(x)}{Z_m}\beta_m^{1-e_{mi}}$$
   where Z_m is the normalization factor, so that the weight sum of the sample set is 1, $Z_m=\sum_{i=1}^{n}\varpi_{mi}\beta_m^{1-e_{mi}}$.

End
Output: Final predictor: $q(x)=\phi_{m^*}(x)$. Where $\phi_{m^*}(x)$ is the base learner corresponding to the sequence number m^* corresponding to the median of all $\ln(1/\beta_m)$, $m=1,2,\cdots,M$.

Proposed Forecasting Process of Electricity Sales Market Indicators

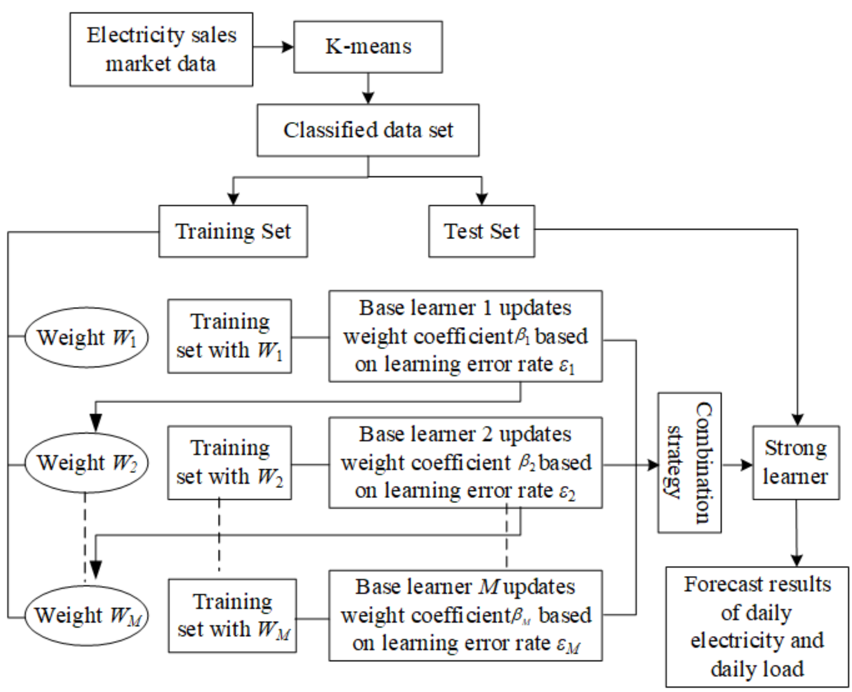
The proposed ConvLSTM-AdaBoost model is used to analyze the electricity sales market data after K-means clustering. This is to obtain the final short-term forecasting results of daily electricity and daily load. For the electricity sales market with a high proportion of DEN access, the prediction process based on ConvLSTM-AdaBoost model is shown in Figure 3.

In the process of forecasting the electricity sales market indicators for the high proportion of DEN access, the K-means is first used to classify n observation data to obtain the electricity data, distributed power output data, meteorological data, electricity price, holidays, and other types of data. Then, the data set is divided into training and test set. The training set is input into the ConvLSTM-AdaBoost model for learning to obtain a strong learning model with the best performance (Hu et al., 2022). Finally, the test set is input into the trained ConvLSTM-AdaBoost model for processing. This obtains the final daily electricity consumption and daily load pre-measurement of the electricity sales market.

ANALYSIS OF EXPERIMENT RESULTS

The testing used daily electricity data and weather data of a city's electricity sales market in western China from January 1 through December 31, 2022. The collection interval was 1h. Throughout the experiment, nine computing nodes were deployed, including one host for task allocation and management of the cloud center and eight edge computing nodes. The host configuration was: 3.5GHz Intel Xeon E5-2697 v2 CPU, 32GB memory. The remaining computing nodes were configured with a 3.2GHz Intel Core i5-6500 CPU and 8GB of memory. All computing nodes were installed with the Linux operating system (Ubuntu 1.04). Meanwhile, a prediction model was built based on the TensorFlow deep learning framework. The ConvLSTM layer number was 2, the convolution core size was 5×5, and M was 12.

Figure 3. Index prediction process based on ConvLSTM-AdaBoost model



Evaluating Indicator

The performance of the proposed model on the test set must be evaluated. At present, there are many evaluation indicators in the prediction, including: mean absolute error (MAE); root mean square error (RMSE); and mean absolute percentage error (MAPE). They are calculated as follows:

$$M_{AE} = \frac{1}{n}\sum_{i=1}^{n}\left|\hat{y}_i - y_i\right| \tag{8}$$

$$R_{MSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left|\hat{y}_i - y_i\right|^2} \tag{9}$$

$$M_{APE} = \left(\frac{1}{n}\sum_{i=1}^{n}\left|\frac{\hat{y}_i - y_i}{y_i}\right|\right) * 100\% \tag{10}$$

where $\hat{y}_i$ is the predicted value.

Impact Analysis of High Proportion DEN Access System

The massive access of DEN will affect the trend of electricity consumption. To study the impact of DEN generation connected to the power grid on the electricity sales market, it is divided into two forecasting scenarios. The first is a consideration of the impact of DEN access. The second is without a consideration of the impact of DEN access. Both prediction scenarios are verified by the proposed model. The former takes the distributed new energy output data into account when constructing the input data of the model. It takes its historical output data as an important input feature. The latter did not consider the distributed new energy output data when constructing the input data of the model; however, it only took the historical data of electricity, weather characteristics, and holiday characteristics as the input of the model. Taking a day in spring as an example, the forecasting results of the proposed model on the electricity consumption trend of the electricity sales market are shown in Figure 4 (with and without considering the impact of DEN access).

From Figure 4, when considering the impact of DEN access factors, the predicted results are closer to the actual values. The consideration of the new energy access is conducive to the proposed model to grasp the trend of electricity change to make a more accurate prediction. However, there are uncertainties in the distributed new energy. If output is not considered, a large deviation will occur in the system power forecast.

Three penetration scenarios (10%, 20%, and 30%) are set to further refine the impact of DEN access to the power grid on the electricity sales market. Under different penetration forecast scenarios, the impact of distributed new energy output on electricity sales market electricity forecast is shown in Table 1.

From Table 1, with different permeabilities, the electricity prediction effect varies. The lower the permeability, the higher the prediction accuracy. When the permeability is 10%, its MAPE is 1.625%. This shows that the higher the penetration rate, the more unstable the system. In addition, the greater the impact of DEN access to the grid on the electricity sales market.

Figure 4. Results of DEN access impact on electricity forecast

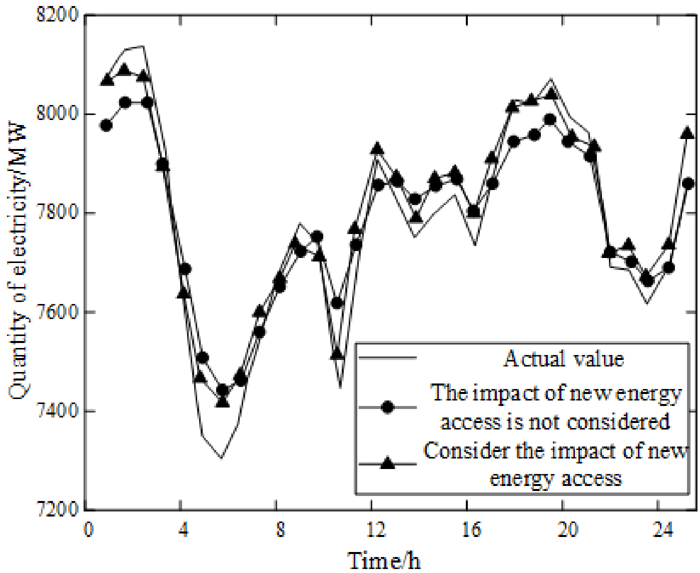


Table 1. Index value of prediction results for proposed model under different permeability scenarios

Scene	Evaluation Index		
	MAE/MW	RMSE/MW	MAPE/%
10% penetration rate	35.691	49.737	1.625
20% penetration rate	39.577	54.106	1.983
30% penetration rate	41.541	57.359	2.112

Comparison of Experimental Results

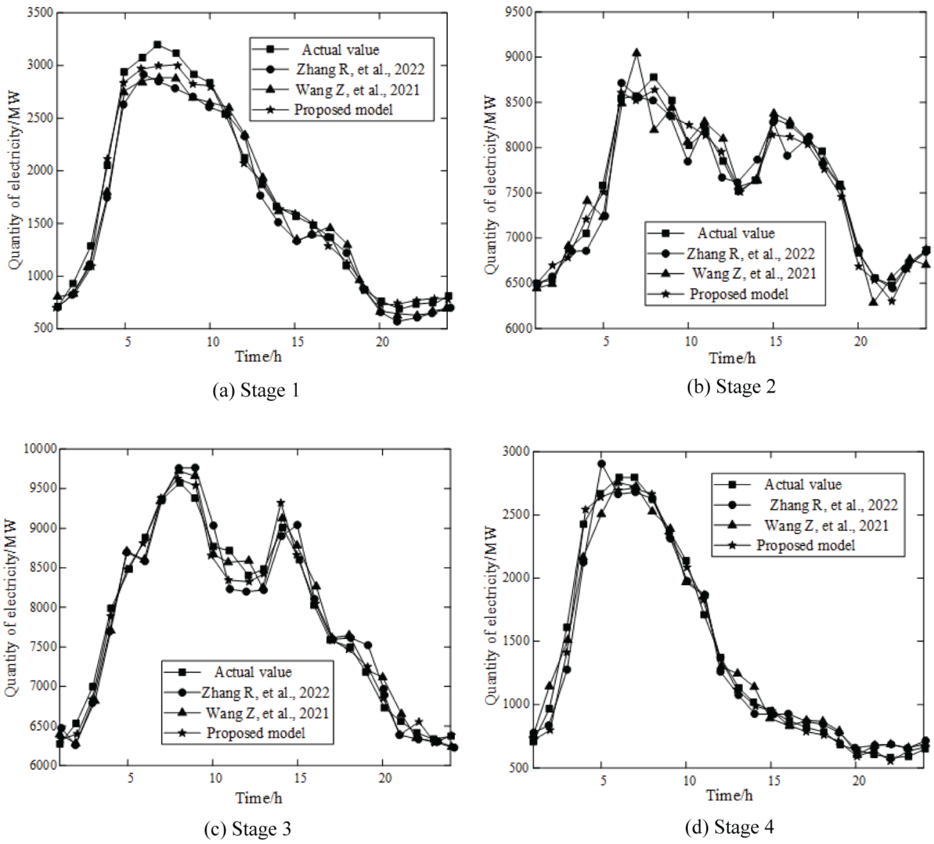
To better demonstrate the prediction performance of the proposed model, the researchers compared it with Zhang et al. (2022) and Wang et al. (2021). Zhang et al. (2022) realized the short-term load forecasting by combining LSTM and wavelet transform, ensuring the forecasting accuracy and efficiency. Wang et al. (2021) proposed an ultra-short-term prediction model based on sequence-to-sequence and GRU. This was combined with the BA mechanism to ensure the reliability of prediction results.

Analysis of Model Prediction Results in Different Seasons

The study selected the hourly data of daily electricity in 2022. It was then divided into four stages according to the four seasons. The prediction results of typical daily electricity in each stage are shown in Figure 5. Figure 5(a) and (c) are rest days; Figure 5(b) and (c) are working days.

According to Figure 5, the overall predicted value of the proposed model is the closest to the actual value. Only the predicted value at individual moments shows a large deviation. This is because the proposed model combines K-means clustering algorithm and ConvLSTM-AdaBoost model, which can better learn the space-time characteristics of daily electricity and ensure the accuracy of prediction. In addition, it can be found that the proposed model can achieve a higher similarity on

Figure 5. Results of daily load forecasting in different seasons



(a) Stage 1 (b) Stage 2 (c) Stage 3 (d) Stage 4

both working days and rest days, which further verifies its ability to obtain better prediction results. Table 2 shows the RMSE values for prediction results of different models in different seasons to quantitatively explain the prediction reliability of the proposed model.

From Table 2, the RMSE of the proposed model at each stage is smaller than that of the model in Zhang et al. (2022) and Wang et al. (2021). Taking typical day 1 and 3 as examples, the RMSE is 56.284MW and 90.545MW. With the help of K-means and ConvLSTM-AdaBoost models, the accurate prediction of the daily electricity consumption in the electricity sales market in different seasons is achieved, verifying the effectiveness of the model. At the same time, Zhang et al. (2022) combined LSTM and wavelet transform to predict the load. Due to the limited or no consideration of holidays and other factors, the prediction deviation is large. The RMSE value of typical day 3 exceeds 100MW.

Table 2. Predicted RMSE value of different models

Scene	Zhang et al. (2022)	Wang et al. (2021)	Proposed Model
Typical Day 1	66.825	63.153	56.284
Typical Day 2	95.093	92.374	83.091
Typical Day 3	102.739	98.906	90.545
Typical Day 4	62.844	59.137	51.638

Wang et al. (2021) improved the GRU model by integrating the BA mechanism, which can ensure a better prediction effect. However, compared with the proposed model using AdaBoost algorithm to train the ConvLSTM model, its improved GRU model shows a poor prediction ability. Taking the typical day 2 as an example, the predicted RMSE has increased by 9.283 MW.

Impact Analysis of Different Electricity Prices

Three models are used to forecast the selected 12 typical days under the parity scenario and time-of-use electricity price scenario. The prediction effects of each method under different electricity price scenarios are explored. In the time-of-use electricity price scenario, the peak-valley attribute and price difference characteristics are added to the three models to predict the typical day. The average value of each experimental error index is calculated. The results are shown in Table 3.

From Table 3, the prediction effect of the proposed model is best when the electricity price remains unchanged (i.e., under the parity scenario). In addition, the MAPE is only 1.927%. This is because there is no need to consider the electricity price under the parity scenario. The less impact of influencing factors, the higher the prediction accuracy. Under the dynamic time-of-use electricity price scenario, the forecasting effect of the three methods is significantly smaller than that under the parity scenario, especially without considering the electricity price factor. When the time-of-use electricity price exists in the system, the load and distributed new energy generation are adjusted by the electricity price factor. If the electricity price factor is not considered in the prediction model at this time, the prediction accuracy will be significantly reduced and the predicted RMSE value of the three models will be greater than 105MW. If the electricity price factor is considered in the prediction model, the prediction performance of the model is determined by the analysis ability of the influencing factors. The proposed model can comprehensively consider various factors and train them by classification. As a result, the prediction error is lower. Compared with Zhang et al. (2022) and Wang et al. (2021), the MAPE value of the proposed model has decreased by 0.358% and 0.541%, proving that it has a good analysis ability for electricity price.

Prediction Stability Analysis of Different Models

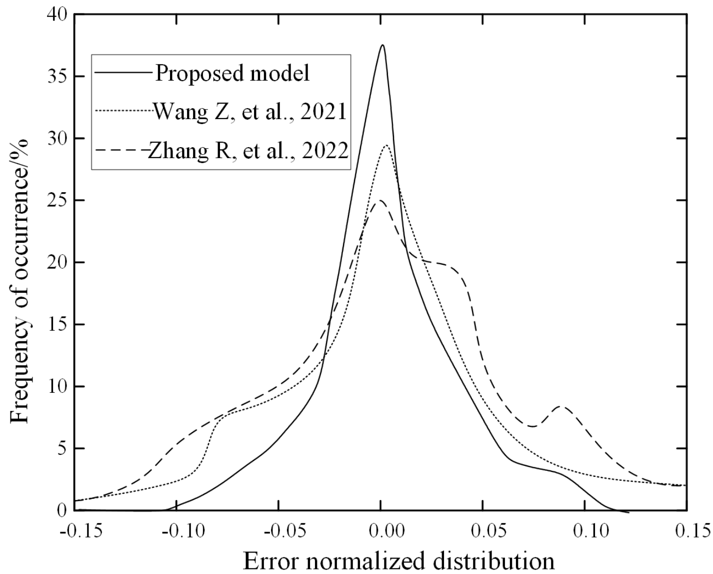
Figure 6 illustrates the statistical error distribution results in 2022, explaining the stability of the prediction results of the ConvLSTM-Adaboost model.

From the perspective of error distribution in Figure 6, the error distribution of the ConvLSTM-Adaboost model is closer to the central zero point. This shows a distribution pattern that is high in the middle and low on both sides. Thus, the ConvLSTM-Adaboost model has a high stability. The curves of Zhang et al. (2022) and Wang et al. (2021) are similar, indicating that their prediction stabilities

Table 3. Predicted RMSE value of different models

	Grid Parity		Time-of-Day Tariff	
	RMSE/MW	MAPE/%	RMSE/MW	MAPE/%
Electricity price is not considered.				
Zhang et al. (2022)	61.482	2.305	118.928	5.413
Wang et al. (2021)	59.735	2.198	114.663	5.206
Proposed model	54.928	1.927	105.715	4.981
Electricity price is considered.				
Zhang et al. (2022)	-	-	64.817	2.598
Wang et al. (2021)	-	-	61.904	2.415
Proposed model	-	-	56.843	2.057

Figure 6. Error distribution curve



are also good. However, the wide curve distribution indicates that the prediction accuracy is not as good as the proposed model. Among them, the error distribution curve of the model in Zhang et al. (2022) is irregular and contains a large maximum error. Thus, the prediction stability is poor.

By combining various scenario factors, the three models are used to forecast daily electricity and daily load, respectively. The resulting index values are shown in Table 4.

From Table 4, the proposed model maintains a good forecasting accuracy for both load forecasting and electricity forecasting. Especially for load forecasting, the MAE, PMSE, and MAPE are only 20.351MW, 25.717MW, and 1.924%, respectively. This is because the ConvLSTM model can learn the space-time characteristics of the load. The prediction effect is guaranteed through the serial training of AdaBoost algorithm. At the same time, the load prediction considers fewer factors. Thus, the overall prediction effect seems to be ideal. Similarly, the MAE, PMSE, and MAPE of the proposed model for the daily electricity forecast results are 52.539MW, 56.859MW, and 2.063%, respectively. This is superior to other models. The LSTM model in Zhang et al. (2022) lacked the analysis of data

Table 4. Comparison of prediction results for different models

Model	MAE/MW	RMSE/MW	MAPE/%
Load forecasting			
Zhang et al. (2022)	25.003	29.998	2.217
Wang et al. (2021)	23.926	28.236	2.155
Proposed model	20.351	25.717	1.924
Electricity forecast			
Zhang et al. (2022)	60.852	64.824	2.601
Wang et al. (2021)	57.054	61.915	2.427
Proposed model	52.539	56.859	2.063

space characteristics; therefore, the prediction error is relatively large and the daily electricity forecast MAPE is 2.601%. Wang et al. (2021) integrated the BA mechanism and improved the prediction performance. Still, this does not consider the comprehensive impact of multiple factors. Thereby, the prediction accuracy needs to be further improved.

CONCLUSION

The high proportion of DEN connected to the power grid has brought new challenges to its stable and reliable operation, especially in the electricity sales market. There have been serious problems regarding ways to maintain the balance of electric energy. Thus, a power sales market index prediction model for high proportion DEN access under the CEC architecture is proposed. In the index prediction system based on the CEC architecture, the data classification is completed by deploying K-means clustering algorithm on the edge side. It is sent into the ConvLSTM-Adaboost prediction model of the cloud center. Through the Adaboost integration algorithm, the ConvLSTM base learner is trained serially. The prediction results are weighted and integrated to output the final power and load prediction results. The experimental conclusions are as follows:

1. The proposed model can access a high proportion of DEN. At the same time, the lower the permeability, the higher the prediction accuracy. When the permeability is 10%, its MAPE is only 1.625%.
2. The prediction reliability is improved by using ConvLSTM-Adaboost. The prediction results of MAE, PMSE, and MAPE for daily electricity are 52.539MW, 56.859MW, and 2.063%, respectively. These results are superior to other models.

In the future, the massive access of electric vehicles will pose a serious challenge to the electricity sales market. Access rules will be affected by factors like user behavior and holidays. Therefore, the next study will focus on the impact of disorderly access of electric vehicles to the electricity sales market on its indicator prediction.

AUTHOR CONTRIBUTIONS

Shuqian Xue is responsible for working on the entire manuscript. Tao Yao: Conceptualization, Methodology, Writing (original draft), Funding Acquisition. Shuqian Xue: Project Administration, Methodology. Xiaolong Yang: Investigation, Writing (review and editing). Chenjun Sun: Validation, Writing (review and editing). Peng Wu: Validation, Writing (review and editing).

FUNDING

This work was supported by science and technology projects of State Grid Corporation of China (5204XA22000D).

CONFLICTS OF INTEREST

The authors declare that publication of this material does not involve conflicts of interest.

REFERENCES

Banga, A., Ahuja, R., & Sharma, S. C. (2021). Stacking machine learning models to forecast hourly and daily electricity consumption of household Internet of things. *Journal of Scientific and Industrial Research*, *80*(10), 894–904.

Bian, H., Guo, Z., Zhou, C., & Peng, S. (2022). Multi-time scale electric vehicle charging load forecasting considering constant current charging and parallel computing. *Energy Reports*, *8*, 722–732. doi:10.1016/j.egyr.2022.08.034

Dab, K., Agbossou, K., Henao, N., Dubes, Y., Kelouwani, S., & Hosseini, S. S. (2022). A compositional kernel based gaussian process approach to day-ahead residential load forecasting. *Energy and Buildings, 254*(1), 111459.1-111459.10.

David, M., Boland, J., Cirocco, L., Lauret, P., & Voyant, C. (2021). Value of deterministic day-ahead forecasts of PV generation in PV + Storage operation for the Australian electricity market. *Solar Energy*, *224*(8), 672–684. doi:10.1016/j.solener.2021.06.011

Evangelopoulos, V. A., & Georgilakis, P. S. (2022). Probabilistic spatial load forecasting for assessing the impact of electric load growth in power distribution networks. *Electric Power Systems Research, 207*(6), 107847.1-107847.10.

Goli, A., Tirkolaee, E. B., & Weber, G. W. (2021). An integration of neural network and shuffled frog-leaping algorithm for CNC machining monitoring. *Foundations of Computing and Decision Sciences*, *46*(1), 27–42. doi:10.2478/fcds-2021-0003

Hong, Y. Y., Chan, Y. H., Cheng, Y. H., Lee, Y.-D., Jiang, J.-L., & Wang, S.-S. (2022). Week-ahead daily peak load forecasting using genetic algorithm-based hybrid convolutional neural network. *IET Generation, Transmission & Distribution*, *16*(12), 2416–2424. doi:10.1049/gtd2.12460

Hu, Y., Li, J., Hong, M., Ren, J., & Man, Y. (2022). Industrial artificial intelligence based energy management system: Integrated framework for electricity load forecasting and fault prediction. *Energy, 244*(4), 123195.1-123195.16.

Huang, Y., Hasan, N., Deng, C., & Bao, Y. (2022). Multivariate empirical mode decomposition based hybrid model for day-ahead peak load forecasting. *Energy, 239*(1), 122245.1-122245.15.

Jiang, W. (2022). Deep learning based short-term load forecasting incorporating calendar and weather information. *Internet Technology Letters, 5*(4), e383.1-e383.6.

Kalhori, M., Emami, I. T., Fallahi, F., & Tabarzadi, M. (2022). A data-driven knowledge-based system with reasoning under uncertain evidence for regional long-term hourly load forecasting. *Applied Energy, 314*(5), 118975.1-118975.15.

López, M., Sans, C., & Valero, S. (2022). Automatic classification of special days for short-term load forecasting. *Electric Power Systems Research, 202*(1), 107533.1-107533.9.

Lu, S., Xu, Q., Jiang, C., Liu, Y., & Kusiak, A. (2022). Probabilistic load forecasting with a non-crossing sparse-group Lasso-quantile regression deep neural network. *Energy, 242*(3), 122955.1-122955.12.

Luo, X., Huang, Y., Zhang, F., & Wu, Q. (2022). Study of the load forecasting of a wet mill based on the CEEMDAN-refined composite multiscale dispersion entropy and LSTM nerve net. *International Journal of Automotive Technology*, *16*(3), 340–348. doi:10.20965/ijat.2022.p0340

Meng, Z. (2022). Bagging based multi-source learning and transfer regression for electricity load forecasting. *IAENG International Journal of Computer Science*, *49*(2), 335–340.

Niu, D., Yu, M., Sun, L., Gao, T., & Wang, K. (2022). Short-term multi-energy load forecasting for integrated energy systems based on CNN-BiGRU optimized by attention mechanism. *Applied Energy, 313*(5), 118801.1-118801.17.

Peng, C., Tao, Y., Chen, Z., Zhang, Y., & Sun, X. (2022). Multi-source transfer learning guided ensemble LSTM for building multi-load forecasting. *Expert Systems with Application, 202*(9), 117194.1-117194.13.

Pereira, L. D. L., Yahyaoui, I., Fiorotti, R., de Menezes, L. S., Fardin, J. F., Rocha, H. R. O., & Tadeo, F. (2022). Optimal allocation of distributed generation and capacitor banks using probabilistic generation models with correlations. *Applied Energy, 307*(2), 118097.1-118097.13.

Qin, J., Zhang, Y., Fan, S., Hu, X., Huang, Y., Lu, Z., & Liu, Y. (2022). Multi-task short-term reactive and active load forecasting method based on attention-LSTM model. *International Journal of Electrical Power & Energy Systems, 135*(2), 107517.1-107517.12.

Shahare, K., Mitra, A., Naware, D., Keshri, R., & Suryawanshi, H. M. (2023). Performance analysis and comparison of various techniques for short-term load forecasting. *Energy Reports*, *9*(7), 799–808. doi:10.1016/j.egyr.2022.11.086

Singla, A., Singh, K., & Yadav, V. K. (2021). Optimization of distributed solar photovoltaic power generation in day-ahead electricity market incorporating irradiance uncertainty. *Journal of Modern Power Systems and Clean Energy*, *9*(3), 1–16. doi:10.35833/MPCE.2019.000164

Tan, M., Hu, C., Chen, J., Wang, L., & Li, Z. (2022). Multi-node load forecasting based on multi-task learning with modal feature extraction. *Engineering Applications of Artificial Intelligence*, *112*(11), 104856–104869. doi:10.1016/j.engappai.2022.104856

Wang, H., Zhang, N., Du, E., Yan, J., Han, S., & Liu, Y. (2022). A comprehensive review for wind, solar, and electrical load forecasting methods. *Global Energy Interconnection*, *5*(1), 9–30. doi:10.1016/j.gloei.2022.04.002

Wang, Z., Li, H., Tang, Z., & Liu, Y. (2021). User-level ultra-short-term load forecasting model based on optimal feature selection and Bahdanau attention mechanism. *Journal of Circuits, Systems and Computers, 30*(15), 2150279.1-2150279.21.

Yang, D., Guo, J., Sun, S., Han, J., & Wang, S. (2022). An interval decomposition-ensemble approach with data-characteristic-driven reconstruction for short-term load forecasting. *Applied Energy, 306*(1), 117992.1-117992.16.

Yang, W., Shi, J., Li, S., Song, Z., Zhang, Z., & Chen, Z. (2022). A combined deep learning load forecasting model of single household resident user considering multi-time scale electricity consumption behavior. *Applied Energy, 307*(2), 118197.1-118197.18.

Zhang, R., Yu, M., & Zhang, C. (2022). A similar day based short term load forecasting method using wavelet transform and LSTM. *IEEJ Transactions on Electrical and Electronic Engineering*, *17*(4), 506–513. doi:10.1002/tee.23536

Zhu, K., Li, Y., Mao, W., Li, F., & Yan, J. (2022). LSTM enhanced by dual-attention-based encoder-decoder for daily peak load forecasting. *Electric Power Systems Research*, *208*(3), 107860–107870. doi:10.1016/j.epsr.2022.107860

Tao Yao (1991.09-) is with State Grid Energy Research Institute Co., Ltd., Her research directions include electrical engineering, smart energy systems, virtual power plants, etc.

Xiaolong Yang (1989.03-) is with State Grid Hebei Information & Telecommunication Branch, Shijiazhuang, Hebei Province. His now research interests include electrical engineering, smart energy systems, virtual power plants, and so on.

Chenjun Sun is with State Grid Hebei Information & Telecommunication Branch. His research interests are cyber physical systems, smart grid, vurital power plant, and so on.

Peng Wu (1978.09-) is with State Grid Energy Research Institute Co., Ltd., his research directions include electrical engineering, smart energy systems, virtual power plants, etc.

Shuqian Xue (1988.10-) is with Beijing Tsingsoft Technology Co., Ltd., her research directions include electrical automation, new energy power prediction, and power grid load prediction.