Explainable Artificial Intelligence for Load Forecasting Based on Adaptive Baseline Daily Load Curve Optimization

Tian Ni, *Student Member*, *IEEE*, Xixiu Wu, *Member*, *IEEE*, Hui Hou, *Senior Member*, *IEEE*, *Senior Member*, *CSEE*, Jingqi Xu, *Student Member*, *IEEE*, Zhao Yang Dong, *Fellow*, *IEEE*, and Chao Luo

***Abstract*—Accurate daily load curve forecasting is critical to tackle the uncertainty and secure the operation of power systems. With the excellent performance achieved by Artificial Intelligence techniques, it is challenging to explain the models' increasing prediction complexity. This paper proposes an explainable daily load curve forecasting method based on adaptive baseline curve optimization. First, the characteristic of the daily load is reflected by three indicators including the daily peak load, the daily load factor, and the daily minimum load factor. For each indicator, we use Random Forest to select the input features and Gate Recurrent Unit to forecast the indicator. Next, based on the predictions of the load characteristic indicators, the baseline curves are calculated by Adaptive Cosine Distance Nearest Neighbor from historical load curves. Finally, the prediction results of three indicators and the adaptive baseline curve are optimized to obtain the final predictive values. A real-world case from Hubei, China is adopted to evaluate the performance of the proposed method. Experimental results indicate that the proposed method can effectively balance the accuracy and explainability of daily load curve forecasting.**

***Index Terms*—Explainable Artificial Intelligence, load forecasting, daily load curve, adaptive baseline curve, load optimization.**

I. Introduction

THE daily load curve forecasting analyzes and models historical load to predict the 24-hour load for the next day, which is the basis for the economic and secure operation of power systems, especially with the spot market development these days [1], [2]. As a result, load forecasting is required to have higher forecasting performance than traditional power systems. More importantly, the reasons for the decisions made by models receive more attention. It means that daily load curve forecasting needs to consider not only accuracy but also explainability.

Artificial Intelligence (AI) has achieved excellent prediction accuracy in load forecasting and has been rapidly developed and widely used [3], [4]. Models have been compared in [5], [6], and [7], and AI models usually achieve better predictions than others, especially the models based on Recurrent Neural Networks (RNN) such as Long Short-Term Memory (LSTM) and Gate Recurrent Unit (GRU). Considering that a simple model is usually limited, AI models become complex for greater accuracy which has been shown that they may get better expressiveness and performance than conventional shallow models [8]. The complexity of the model is usually manifested in multi-parameter, multi-model, and multi-stage. Multi-parameter refers to setting a large number of parameters for the model. For example, the complexity of neural networks is that they have more complex structures and more neurons [9]. Multi-model is in terms of the stacking of different models [10]. A Sequence-to-Sequence model is proposed to focus on multistep load forecasting. The proposed model utilizes the Temporal Convolution Network as the encoder and LSTM as the decoder [11] Multi-stage means the forecasting process is divided into different stages, with separate forecasting models for each stage. Sharma *et al.* [12] establish a two-stage prediction task using Backpropagation Neural Networks and Radial Basis Function Neural Networks to predict respectively, which takes advantage of different neural networks. However, the complex model is not always beneficial because the training and optimization become difficult. Meanwhile, the more complex models become, the more susceptible to gradient vanishing or explosion they are. In addition, AI models especially deep neural networks are still considered to be unexplainable black-boxes in most cases due to the large number of model parameters, complex mechanism, opaque model, and imprecise significance of parameters. Considering that power systems require security and economy, the application of AI without explainability is not easy.

Recently, research on explainable AI (XAI) has made it possible to address the problems. There are three approaches to understanding the results of AI models according to different modeling stages [13]. The first is the pre-model stage

This work was supported by National Natural Science Foundation of China (52177110), Shenzhen Science and Technology Program (JCYJ20210324131409026), China Power Engineering Consulting Group CO., LTD. Project Funding(DG2-X02-2022).

T. Ni, X. Wu (corresponding author, e-mail: wuxixiu@whut.edu.cn) and H. Hou are with the School of Automation, Wuhan University of Technology, Wuhan 430070, China.

J. Xu is with the Department of Electrical and Computer Engineering, University of Southern California, Los Angeles 90089, USA.

Z. Y. Dong is with the Department of Electrical Engineering, City University of Hong Kong, Hong Kong.

C. Luo is with the Central Southern China Electric Power Design Institute of China Power Engineering Consulting Group, Wuhan 430071, China.

Digital Object Identifier

explanation, which mainly consists of data mining and presentation such as dimensionality reduction [14] or clustering algorithms [15]. Essentially, AI models mine features and discover patterns from data. Pre-model stage explanation is where the researchers actively find the characteristics of the data distribution before training the model. To extract features availably, Ji *et al.* [16] propose Cosine-distance-based K-Medoids clustering algorithms, then they categorize the clustering results into three typical scenarios including weekdays, weekends or holidays, and hot weather days. According to the data characteristics of features, a suitable model can be built. The users are divided into multiple clusters based on load pattern similarity, so the federated learning model can be trained [17]. It can be seen that the purpose of the pre-model stage explanation is to reveal the patterns of the input data. Therefore, this type of model explanation is usually used for feature selection, feature dimensionality reduction, and model selection.

The second one is the post-model stage explanation. At this stage, the main task is to identify the relationship between features and predictions. SHapley Additive exPlanations (SHAP) [18], Local Interpretable Model-agnostic Explanations (LIME) [19], and Standardized Regression Coefficient (SRC) [20] fall into this category. Most of these methods focus on explaining individual predictions or the prediction results of a particular instance, rather than a general overview of the overall model behavior. Based on the same idea, Pelekis *et al.* [21] correlate features with the prediction error of each deep learning model through Multi-Layer Perceptron, and this post-model analysis effectively tests the performance of each model and identifies the factors that most affect the accuracy. The post-model stage explanation is to reveal the relative importance of features. Therefore, this approach is used for feature selection, model improvement, and system monitoring. Neither pre-model nor post-model stage explanation explains AI models in terms of modeling mechanisms, which is not enough to understand the decisions made by the models. Sometimes it is necessary to understand and control the model decision-making process from the mechanism so that researchers can make adjustments and modifications effectively.

The last is self-explanatory modeling, which is a way to understand modeled decisions through modeled mechanisms. Linear models, tree models [22], Attention Mechanism (AM) [23], and modeling with physical knowledge are used to make the model prediction mechanism transparent or add physical meaning so that it can be easily understood. Linear models including Linear Regression (LR) [24], Lasso [25], and Ridge [26] calculate the weights for features to build the relationship between the results and the inputs. Therefore, it is easy to understand modeling decisions according to the weights. However, the models are susceptible to outliers, and as nonlinearity increases, the performances of linear models may be doubted. For tree models such as Decision Tree (DT) [27] or Random Forest (RF) [28], each node refers to a judgmental decision, and the conclusion can be obtained through a series of decisions along the branches of the tree. The more important features will be discussed at more important nodes and tree models can transform the data into an explainable form by *if-then*. It makes it easy to understand the contribution of features and model decisions [29]. However, more volatile data needs more nodes, leaves, and layers of the trees, limiting the models' explainability and increasing the risk of overfitting. AM explains what the model learns by assigning different weights to the features [30]. Unlike LR, AM computes the weights in the form of a neural network and thus it has a stronger nonlinear representation than LR. Transformer is an effective utilization of AM and has achieved extensive research in load forecasting [31], [32]. However, the Transformer has complex structures and a large number of weight matrices to be solved, which reduces its explainability. At the same time, AM is often not the whole of the deep learning model, which cannot explain the output of other network layers such as the fully connected layer or RNN layer. The new technology such as Kolmogorov-Arnold Networks (KAN) provides new ideas and tools for XAI [33]. However, the practical application of this model is still doubtful at present. For example, Shukla *et al.* point out that the original KAN based on the B-splines parameterization lacks accuracy and efficiency [34]. Combining physical knowledge is feasible to build models with explainability. These studies have been employed in building load [35], [36], wind or solar power [37], [38], and laboratory quakes [39] forecasting, but fewer have been conducted on load forecasting. Li *et al.* [40] proposed interpretable memristive LSTM for probabilistic residential load forecasting, however, more work on hardware design and a wider range of load types are needed to further develop this method. [41] and [42] build models based on load characterization indicators with clear physical implications. However, both [41] and [42] pay attention to the medium and long-term load. The studies do not conduct an in-depth analysis of short-term forecasting like daily load curve forecasting.

In summary, AI model building usually needs to balance accuracy and explainability. Models with simple structures have better explainability but often have difficulty in improving accuracy. The others with complex structures fit well but are not to be understood. To take into account both model accuracy and explainability for daily load curve forecasting, this paper proposes an explainable daily load curve forecasting method based on adaptive baseline curve optimization. The key contributions of this paper are:

1) We present a daily load curve forecasting framework to balance the accuracy and explainability of the forecasting process. The framework optimizes the load characteristic indicators and the adaptive baseline daily load curve, which has clear physical meaning and explainability.

2) Load characteristic indicator prediction models based on PSO-RF-GRU are proposed. For each load characteristic indicator, the model is trained to improve the accuracy while reducing the model structure. XAI technology was used to understand the proposed model.

3) A novel adaptive baseline curve method suitable for load is introduced. Cosine distance and Softmax are used to calculate the weights of the selected load curves to avoid the drawback of

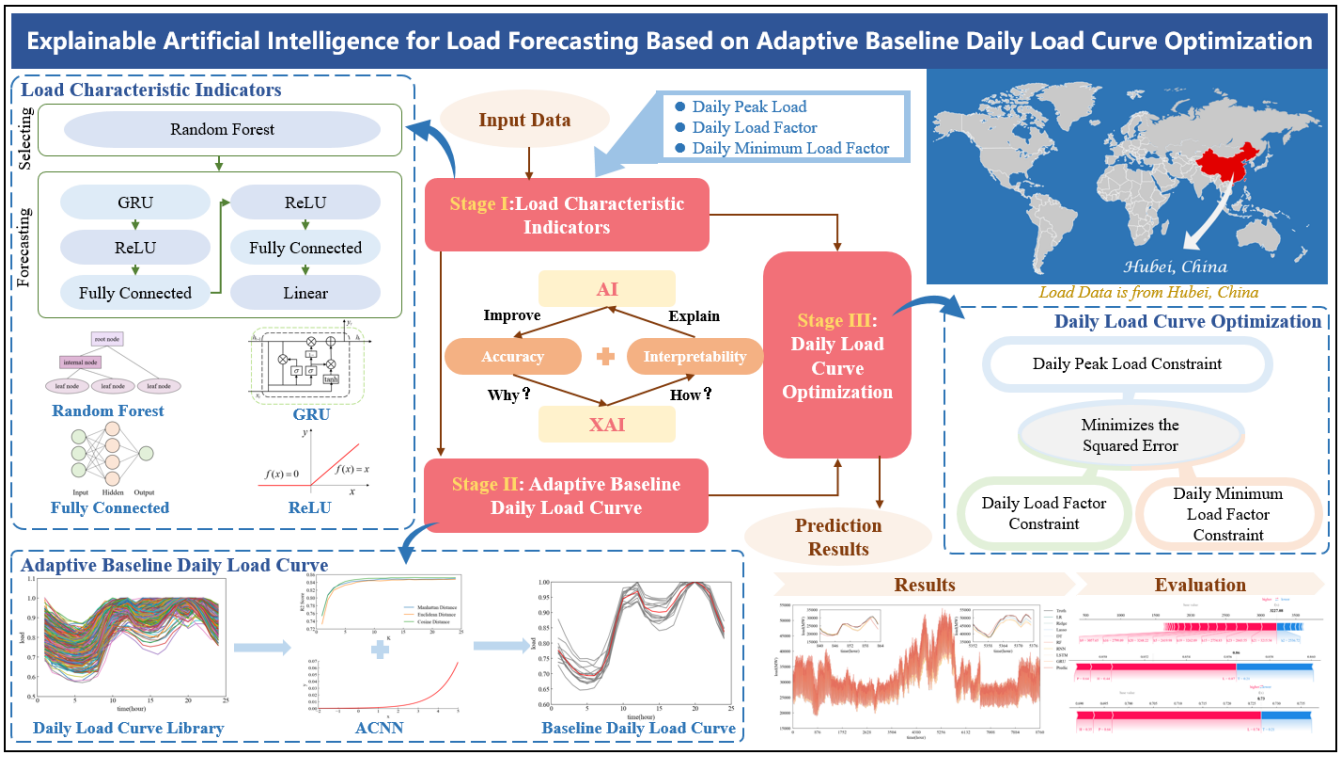


Fig.1. Structure of daily load curve forecasting.

the traditional clustering algorithm which is difficult to reflect the shape change of load.

The rest of this article is organized as follows. In Section II, the structure of an explainable daily load curve forecasting method based on adaptive baseline curve optimization is introduced. Section III describes the proposed method and its architecture methodology. In Section IV, the experimental results are reported and the explainability is discussed. The conclusion is made in Section V.

II. Structure of Daily Load Curve Forecasting

Fig.1 shows the structure of explainable daily load curve forecasting based on adaptive baseline curve optimization. The first part is the load characteristic indicators, the second part is the adaptive baseline daily load curve, and the last part is the load curve optimization.

In the first part, three indicators with explicit physical significance are adopted to reflect the characteristics of the daily load curve including the daily peak load [43], the daily load factor, and the daily minimum load factor [44]. PSO-RF-GRU models are built to predict load characteristic indicators. Specifically, RF is employed for feature selection while GRU is employed to forecast the load characteristic indicators using the selected features. For each model, we use PSO to tune the parameters.

In the second part, the adaptive baseline daily load curve is calculated by Adaptive Cosine-distance Nearest Neighbor (ACNN) according to the predictions in the first part. A library of daily load curves consists of normalized historical curves. We concatenate feature vectors to compute the Cosine distance between daily load curves and calculate the curve weights based on Softmax, where the weights may help enhance model explainability. The adaptive baseline curve can be obtained after the R2 Score of the curve has been converging.

In the last part, the final prediction is computed by optimizing the load characteristic indicators and the baseline daily load curves. The optimization model is built by minimizing the errors as the objective function with physical significance. The three load characteristic indicator predictions are used as constraints to solve the model. After forecasting, the results are evaluated and analyzed for explainability.

III. Daily Load Curve Forecasting

A. Load Characteristic Indicators

The traditional daily load curve forecasting model outputs all predicted values, which will seem convenient after the model has been built. This approach usually focuses only on the aleatory uncertainty, not the epistemic uncertainty [45]. This means that errors from specific models are unavoidable. Simultaneous output of all predicted values may accumulate errors. In addition, this paper will use RNN to predict. As the forecasting time scale increases, the errors will accumulate at each time step. This is not allowed in the daily load curve forecast, which leads to increasing errors and unreliable forecasts for the next day's backward moments. Therefore, to reduce the prediction step, three characteristic indicators are selected to describe the daily load including the daily peak load $p_d^{\max}$, the daily load factor γ_d, and the daily minimum load factor β_d. The three characteristics can describe the range of load fluctuations [42]. Combined with the shape of the baseline load curve, the change in load can be determined.

Predictive performance is often affected by feature inputs.

For each load characterization indicator, the input features need to be screened separately. Pearson correlation coefficient (PCC) is often used to calculate correlations [46]. However, the effect of features on load may not be linear. PCC, which calculates linear correlations, often does not capture nonlinear features well. To fit the nonlinearity in the data, this paper uses RF to find important features. RF combines multiple DTs. Since each tree is constructed using minimized Mean Square Error (MSE) or Mean Absolute Error (MAE), it makes RF much better for nonlinear fitting. By calculating the frequency or weight of each feature, which is used as a node-splitting criterion when training each tree, random forests can rank the importance of features when mining nonlinear features [47].

However, the features selected by RF are not all available as inputs, or only the most favorable one. As an AI model, the predictive performance of GRU will be closely related to these input features. With the increasing number of features, the model is easy to be overfitting. The model is more likely to learn the noise, especially the features with low relevance. In addition, high dimensional inputs also reduce the speed of model training and take more time. However once the input of features is insufficient, the model does not learn enough load characteristics. This may lead to the lower performance and generalization of the model. Therefore, determining the number of features requires more consideration. So in the study, the three best features I_1, I_2, and I_3 are selected as inputs to the prediction model after building RF.

Three GRU models are built to forecast the three load characteristics indicators, which significantly reduces the number of model forecast outputs. Collecting relevant feature data is challenging, and the correlation between those features and the predicted target may be weak. To address this, RNNs like LSTM and GRU effectively analyze data trends without relying on time labels or explicit features. GRU has fewer number of parameters than LSTM. Therefore, the computation of GRU is faster and less prone to gradient vanishing or explosion. The formulations of all nodes in the GRU unit are given as follows from (1) to (4) [47].

$$z_t = \sigma(\boldsymbol{W}_z \cdot [h_{t-1}, x_t]) \tag{1}$$

$$r_t = \sigma(\boldsymbol{W}_r \cdot [h_{t-1}, x_t]) \tag{2}$$

$$\tilde{h}_t = \tanh(\boldsymbol{W}_h \cdot [r_t * h_{t-1}, x_t]) \tag{3}$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t \tag{4}$$

where z_t is the update door, r_t is the reset door, h_t and $\tilde{h}_t$ are the hidden state and the candidate's hidden state, $\boldsymbol{W}_z$, $\boldsymbol{W}_r$, and $\boldsymbol{W}_h$ are the matrix of parameters to be trained.

Since the performances of RF and GRU are directly affected by the model parameters, Particle Swarm Optimization (PSO) is adopted for model parameter tuning to obtain the optimal model parameters and reduce the error [49].

B. Adaptive Baseline Daily Load Curve

The shape of the daily load curve determines the location of the units carrying the load in the power system, the adequacy of the system's peaking capacity, and the magnitude of the interconnected system's peaking benefits. How to effectively describe the load shape becomes the key to daily load curve forecasting. Daily load curves usually have some similarities. For example, typical daily load curves for the same month in different years have similar shapes, and curves for the same date types have similar shapes. This similarity provides a possibility for the selection of the baseline curve.

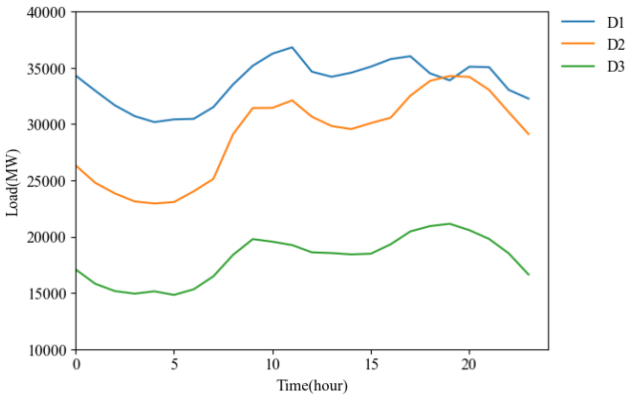


Fig.2. The shape of daily load curve.

To obtain a suitable baseline curve, we need to cluster the historical loads efficiently. Common clustering methods like K-means [15] or Fuzzy C-means (FCM) [50] rely on distances like Euclidean or Manhattan to assess similarity. However, in practical applications, these methods can encounter limitations. For instance, in Fig. 2, visually, D2 seems closer to D3 based on similar fluctuations and peak timings. However, the Euclidean distance calculation indicates that D1 is nearer to D2. While this example is extreme, such scenarios can affect the assessment of daily load curve similarities, highlighting the limitations of traditional distance metrics in clustering analysis. In addition, the clustering results are usually a limited number of curves. Although it is very difficult to try to include all the load curve shape types, we still doubt that the load curves to be predicted can be effectively described by the limited number of clustered curves.

ACNN can effectively solve the problems. This method is inspired by the K-Nearest Neighbor (KNN), so it is named Adaptive Cosine-distance Nearest Neighbor. Instead of computing the fixed, finite number of clustered curves, ACNN will form a library of curves with all historical load curves as objects to be selected. The adaptive weights are determined by calculating the Cosine distance so that each baseline curve can be maximally close to the load curve to be predicted. Cosine distance quantifies similarity by calculating the cosine of the angle between vectors, offering advantages over Minkowski distances. First, the Cosine distance is suitable for application to multidimensional positive spaces, and in load forecasting, the load is characterized as a multidimensional positive vector. Second, when dealing with large-scale data in load forecasting, the computational complexity of the Cosine distance is low because only the inner product of vectors is involved in the computation. Finally, load data often suffer from data offsets, which are not affected by the cosine distance since it only considers the direction of the vector. Unlike KNN being used for classification or regression, the curves we obtained are the results since this paper uses ACNN for finding baseline curves.

ACNN can be divided into the following six steps:

1) Assembling a library of historical daily load curves.

Normalize all historical daily load profiles. To reduce the

effect of load size on similarity and to visualize the load shape, all historical daily load curves are normalized, which is expressed as

$$p_h^* = \frac{p_h^{\text{truth}}}{p_d^{\max}} \tag{5}$$

where p_h^* is the load value after normalization, p_h^{truth} is the truth value, h and d are hour and day.

2) Concatenating feature vectors.

For each history day, splicing the three features selected by RF with load characterization indicators. For example, on day d, the vector corresponding to the daily load curve $\boldsymbol{p}_d = [p_{d,0}, p_{d,1}, \cdots, p_{d,23}]$ is $\boldsymbol{v}_d = [I_{d,1}, I_{d,2}, I_{d,3}, \gamma_d, \beta_d]$.

For the future day, the vector is composed of the same elements as above, but the values are predicted. For example, on day d, the vector corresponding to the daily load curve $\hat{\boldsymbol{p}}_d = [\hat{p}_{d,0}, \hat{p}_{d,1}, \cdots, \hat{p}_{d,23}]$ is $\hat{\boldsymbol{v}}_d = [\hat{I}_{d,1}, \hat{I}_{d,2}, \hat{I}_{d,3}, \hat{\gamma}_d, \hat{\beta}_d]$. Since the input features are statistical values related to weather or date, which usually have a high prediction accuracy, the true weather or date values are taken in all subsequent studies. The load characterization indicators are the GRU forecast results.

It is important to note that since the baseline curve consists of the results of load normalization, the vector formed will not contain $p_d^{\max}$.

3) Measuring similarity.

For the forecast day, the Cosine distance between the day's vector and each historical load vector needs to be calculated, which is expressed as

$$D_\theta(d_1, d_2) = \frac{\sum_{i=1}^{v} x_{d_1,i} \cdot x_{d_2,i}}{\sqrt{\sum_{i=1}^{v} x_{d_1,i}^2} \sqrt{\sum_{i=1}^{v} x_{d_2,i}^2}} \tag{6}$$

where $D_\theta(d_1, d_2)$ is the Cosine distance between $\boldsymbol{v}_{d_1}$ and $\boldsymbol{v}_{d_2}$, v is the number of the element belongs to $\boldsymbol{v}_d$, $x_{d_1,i}$ and $x_{d_2,i}$ are the i-th elements of $\boldsymbol{v}_{d_1}$ and $\boldsymbol{v}_{d_2}$.

4) Calculating the weights.

The K daily load curves with the smallest Cosine distance are selected. The weights are calculated by Softmax, which is given as (7). The sum of the weight values through the Softmax function should be equal to 1.

$$W_k = \frac{\exp(D_\theta(d_k, d))}{\sum_{k=1}^{K} \exp(D_\theta(d_k, d))} \tag{7}$$

where W_k is the weight of the k-th curve selected.

5) Obtaining the baseline daily load curve.

The selected K curves are weighted and averaged according to the weights, which are given as (8).

$$p_h^* = \frac{\sum_{k=1}^{K} p_{h,k}^* \cdot W_k}{K} \tag{8}$$

where p_h^* is the weighted average, $p_{h,k}^*$ is the load value at hour h of the k-th curve.

And the resulting load curve $\boldsymbol{p}_h^*$ is normalized to obtain the baseline daily load curve $\boldsymbol{p}_h^{\text{base}}$ by (5).

6) Selecting K.

For the selection of the K, it is similar to the selection of the number of clusters using the contour coefficient. But instead of the Euclidean distance here, the R2 Score is calculated, which is given by (9). This is because the target is not to find the optimal clustering result but to determine a baseline curve, the accuracy of which directly determines the optimization effectiveness. Therefore, we want the baseline curve obtained from the historical load curve library to be as close as possible to the true value. The R2 Score can reflect the degree of fit of the two curves. The optimal number of curves is selected by comparing the fit of different K values on the test set. As a result, the different historical curves can effectively represent new load curves with resilience and adjustability. The different weights ensure that the curves with higher similarity have a more important position. The curve selected after the weighted average also reduces the influence of outliers. Thus, similar to the contour coefficients, a loop is performed for K. The difference is that the loop can be stopped once the R2 Score is judged to have converged.

$$M^{\text{R2}} = 1 - \frac{\sum_{h=1}^{H} (\hat{p}_h - p_h)^2}{\sum_{h=1}^{H} (\bar{p}_h - p_h)^2} \tag{9}$$

where M^{R2} is the R2 Score values, H is the number of load values, $\hat{p}_h$, p_h, and $\bar{p}_h$ are the value to be evaluated, the true value, and the average value.

C. Daily Load Curve Optimization

The prediction results of three indicators and the adaptive baseline curve are optimized to obtain the final predictive values. The objective is to minimize the deviation between the target and baseline daily load curves, which is expressed as

$$f(\hat{p}_d^*) = \min\left(\sum_{h=1}^{H} (\hat{p}_h^* - p_h^{\text{basis}})^2\right) \tag{10}$$

where $f(\hat{p}_d^*)$ is the optimization objective, $\hat{p}_h^*$ is the optimization object.

With this objective function, it can be guaranteed that the optimized load at each moment is similar to the trend of the baseline curve. The constraints contain the daily peak load constraints, the daily load factor constraint, and the daily minimum load factor constraints.

1) Daily peak load constraint.

$$\max(\hat{p}_h^*) = 1 \tag{11}$$

This constraint indicates that the maximum value of the obtained load is 1, which after inverse normalization is equal to the predicted daily peak load indicator.

2) Daily load factor constraint.

$$\sum_{h=1}^{H} \hat{p}_h^* = 24\gamma \tag{12}$$

This constraint indicates that the daily load factor is equal to the forecasted daily load factor.

3) Daily minimum load factor constraint.

$$\min(\hat{p}_h^{\text{normal}}) = \beta \tag{13}$$

This constraint indicates that the daily minimum load factor for the day is equal to the predicted daily minimum load factor.

D. Solving and Evaluation

For AI model solving, they are built using the machine learning library *sklearn* or the deep learning framework *TensorFlow.Keras*. The neural network is optimized using Adam.

For optimization model solving, they can be considered quadratic programming models, and be solved using the commercial solver *Gurobi*.

The whole process is implemented using Python 3.9 on a PC with 12th Gen 2.30 GHz Intel(R) Core(TM) i7(12700H) and 64 GB RAM.

The standard performance metrics used to evaluate the prediction are given in (9), (14-17), involving R2, Explained Variance Score (EV), MAE, Root MSE (RMSE), and Mean Absolute Percentage Error (MAPE). R2 represents the model's ability to fit the data and EV measures a model's ability to explain changes in data. MAE, RMSE, and MAPE reflect the prediction accuracy, while RMSE is more sensitive to high error points, and MAPE has no gauge effects. For R2 and EV, a value closer to 1 indicates a better model, whereas for MAE, RMSE, and MAPE, a smaller value indicates a smaller prediction error.

$$M^{\text{EV}} = 1 - \frac{\text{var}(\hat{p}_h - p_h)}{\text{var}(p_h)} \tag{14}$$

$$M^{\text{MAE}} = \frac{1}{H}\sum_{h=1}^{H}\left|(\hat{p}_h - p_h)\right| \tag{15}$$

$$M^{\text{RMSE}} = \sqrt{\frac{1}{H}\sum_{h=1}^{H}(\hat{p}_h - p_h)^2} \tag{16}$$

$$M^{\text{MAPE}} = \frac{1}{H}\sum_{h=1}^{H}\frac{\left|(\hat{p}_h - p_h)\right|}{p_h}\cdot 100\% \tag{17}$$

IV. Case Study

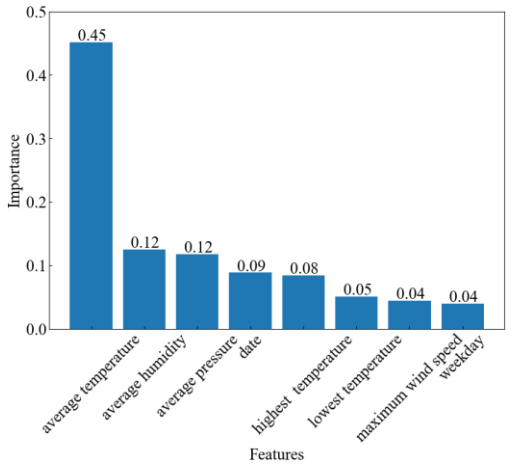
The actual load data from 2014 to 2022 in Hubei Province, China are selected for validation. The datasets from 2014 to 2021 are used as the training set and 2022 as the test set. For forecasting, the inputs are the real load data of the previous day and the real weather of the target day as simulated weather forecast data.

A. Indicators Forecasting Evaluation

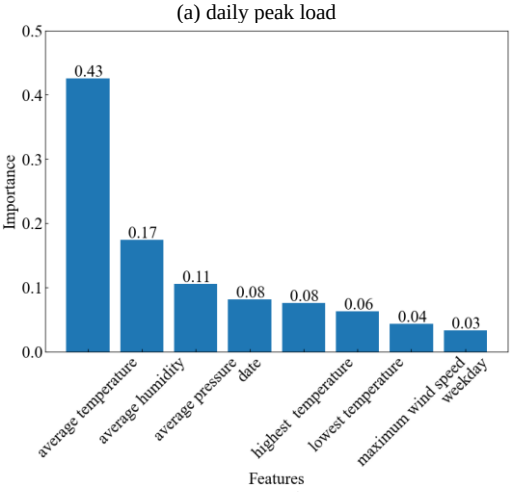
Fig.3 illustrates the results calculated using RF. As a comparison, Fig.4 calculates the correlation using PCC.

As can be seen in Fig.4, there is no strong correlation indicator. Especially for the daily peak load, the correlation is less than 0.2, which cannot meet the correlation requirements. This verifies that PCC only calculates linear correlation, which makes it difficult to achieve better results in areas with strong nonlinear variations such as load forecasting.

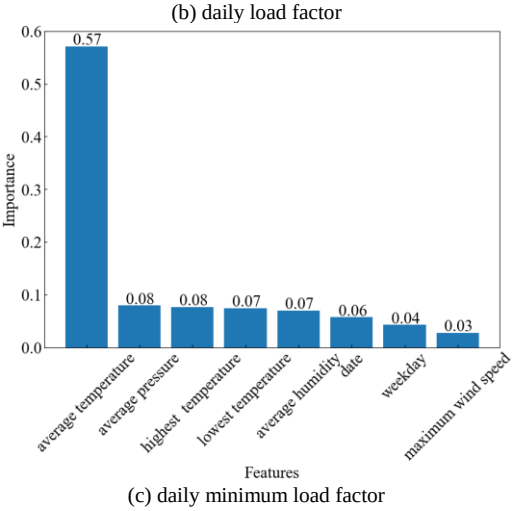
In contrast, the correlation obtained through RF is more significant and intuitive. Because RF is ranking the importance, the first feature is usually with the highest correlation. In addition, the second and third features are also usually higher



(a) daily peak load



(b) daily load factor



(c) daily minimum load factor

Fig.3. RF Correlation Analysis.

than the remaining features. To improve the performance and generalization of the model, the three features are selected, as shown in Table I.

After obtaining the input features, the load characteristic indicators are predicted. Common XAI models are selected for comparison, including linear models (LR, Lasso, Ridge), tree models (DT, RF), and RNN (SimpleRNN, LSTM, GRU).

Fig.5 shows the results of prediction results. GRU can follow the real value changes well. Although there is some deviation in some time steps, the overall difference with the real value is

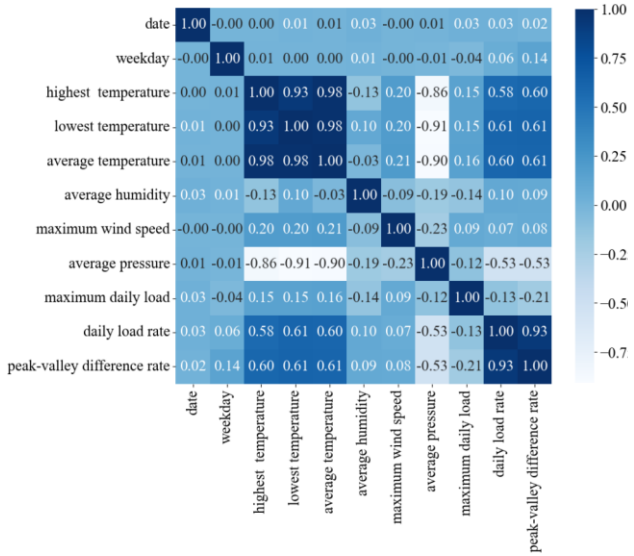


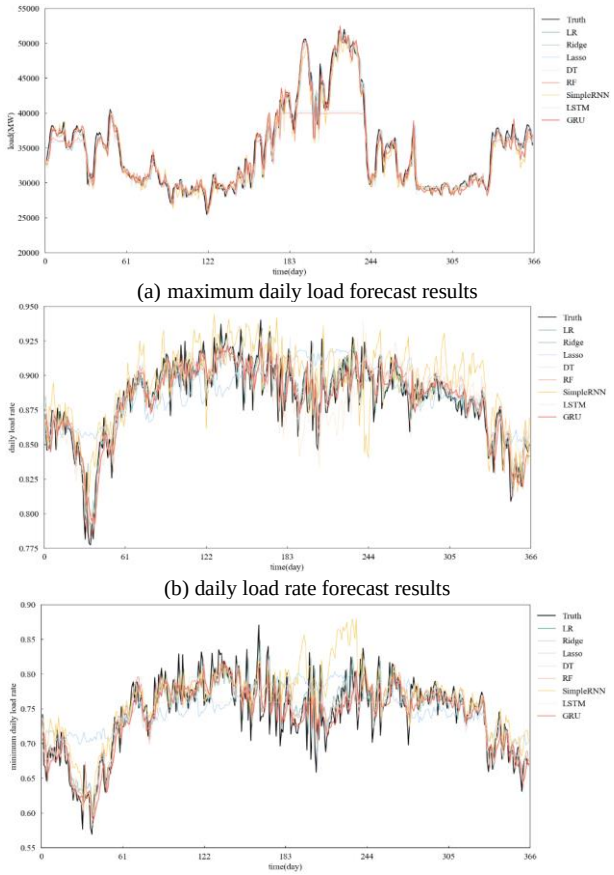
Fig.4. PCC Correlation Analysis.

TABLE I
RESULTS OF INDICATOR SELECTION

	maximum daily load	daily load rate	peak-valley difference rate
feature1	average temperature	average temperature	average temperature
feature2	average humidity	average humidity	average pressure
feature3	average pressure	average pressure	highest temperature

very small. In contrast, the comparative model cannot obtain good prediction results on every load indicator. The linear model cannot effectively follow the time series changes on the more volatile data like the daily load factor or the daily minimum load factor. The tree models have a large defect in peak forecasting. This is because the models are trained by data from previous years. During the development of electricity demand, the load value will gradually increase, but the input characteristics do not fluctuate much from year to year. The *if-then* structure causes the tree model to be unable to reflect the new trend of change, resulting in a decrease in the generalization ability of the tree model.

Tables II, III, and IV show the results of the predictive evaluation of each load characterization indicator. RNN models, especially GRU, show superior performance in the vast majority of cases. R2 and EV of GRU can be kept above 0.75, and the daily peak load can even reach 0.95 and 0.96, which means that the GRU model has a very good fitting performance. In addition, for the daily peak load where the curve changes more gently, the MAE of GRU is reduced by 35.81% compared to the average MAE, and the average RMSE is reduced by 43.76%, which is manifested in the MAPE value, which can be reduced by 1.09% in the GRU model compared to the average MAPE. GRU forecasting effect is still advantageous in the daily load factor and daily minimum load factor, which are more volatile and stochastic. For the daily load factor, comparing the average MAE, GRU reduces by 21.47% and RMSE is 22.85%. In the daily minimum load factor, the average MAE of the comparative models can still be reduced by 19.72% and RMSE by 20.45%. In summary, the GRU



(a) maximum daily load forecast results

(b) daily load rate forecast results

(b) peak-valley difference rate forecast results

Fig.5. Load Characterization Indicators Forecast Results.

model has good forecasting performance and can satisfy the current forecasting needs.

B. Adaptive Baseline Daily Load Curve Evaluation

The most important aspect of the determination of the adaptive baseline daily load curve is the K-value. The optimal K value is found by traversing the K values and calculating R2. To get the best combination, the Manhattan distance, the Euclidean distance, and the Cosine distance used in the proposed model are first compared and the results are obtained as shown in Fig.6.

As shown in Fig.6, after the value of K is greater than 10, the fitting of the three distances does not rise much and can be considered to have converged for all of them. Specifically, when K goes from 1 to 5, R2 improves by 11.98%. When K goes from 5 to 10, R2 improves by 0.88%. And when K goes from 10 to 15 and 15 to 20, the numbers are 0.12% and 0.0073%. The baseline daily load curve selected by cosine distance has a higher coefficient of determination. This verifies that the cosine distance is a more apt description of the shape of the load curve. To balance the degree of fit and prevent excessive computation, the value of K is chosen to be 15.

To further verify the advantages of ACNN, we compared the K-means and FCM and calculated the baseline curve for 2022 by using the clustering center as a typical daily load curve. Comparing the indicators, the results are obtained as shown in Table V. The results show that ACNN minimizes all error values, while the fit is optimal. It means the adaptive baseline daily load curve is closer to the real load curve shape.

TABLE II
RESULTS OF MAXIMUM DAILY LOAD FORECAST EVALUATION

model	R2	EV	MAE(MW)	RMSE(MW)	MAPE(%)
LR	0.93	0.93	1071.46	1559.33	3.01
Ridge	0.93	0.93	1071.46	1559.32	3.01
Lasso	0.93	0.93	1068.30	1557.56	3.00
DT	0.74	0.78	1934.66	3112.01	4.91
RF	0.73	0.77	1884.26	3154.89	4.73
RNN	0.94	0.95	1082.43	1451.84	3.04
LSTM	0.97	0.97	818.72	1113.73	2.35
GRU	0.97	0.97	818.95	1085.31	2.35

TABLE III
RESULTS OF DAILY LOAD RATE FORECAST EVALUATION

model	R2	EV	MAE(10^{-2})	RMSE(10^{-2})	MAPE(%)
LR	0.72	0.72	1.12	1.51	1.28
Ridge	0.73	0.73	1.10	1.49	1.25
Lasso	0.39	0.41	1.67	2.23	1.92
DT	0.74	0.75	1.13	1.45	1.29
RF	0.75	0.75	1.10	1.42	1.25
RNN	0.40	0.57	1.76	2.21	2.01
LSTM	0.62	0.62	1.37	1.77	1.55
GRU	0.77	0.77	1.05	1.36	1.20

TABLE IV
RESULTS OF PEAK-VALLEY DIFFERENCE RATE FORECAST EVALUATION

model	R2	EV	MAE(10^{-2})	RMSE(10^{-2})	MAPE(%)
LR	0.65	0.65	2.26	3.17	3.07
Ridge	0.65	0.65	2.26	3.17	3.07
Lasso	0.30	0.32	3.60	4.50	4.99
DT	0.71	0.72	2.15	2.87	2.92
RF	0.73	0.73	2.09	2.80	2.84
RNN	0.43	0.58	2.99	4.06	4.11
LSTM	0.67	0.68	2.35	3.10	3.12
GRU	0.75	0.76	2.03	2.69	2.74

TABLE V
COMPARISON OF CLUSTERING ALGORITHMS

	R2	EV	MAE	RMSE	MAPE(%)
ACNN	0.85	0.87	2.62	3.53	3.08
K-means	0.78	0.82	3.29	4.33	3.87
FCM	0.77	0.81	3.36	4.44	3.98

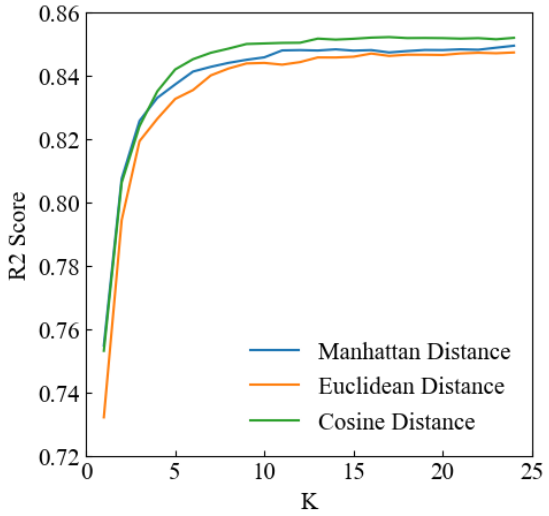


Fig.6. Comparison of different distances.

Subsequent optimizations are based on the curve, so smaller errors mean fewer corrections, thus reducing the errors.

The comprehensive comparison results show that ACNN can effectively follow the real load change when choosing the baseline daily load curve, which provides a good basis for load optimization.

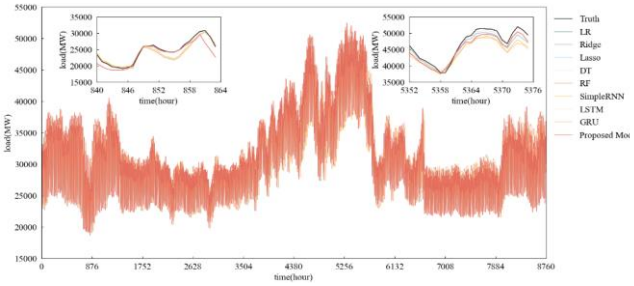
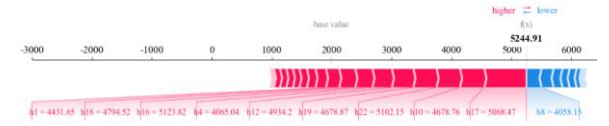
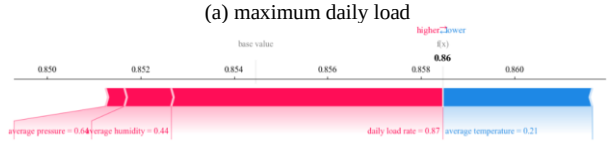


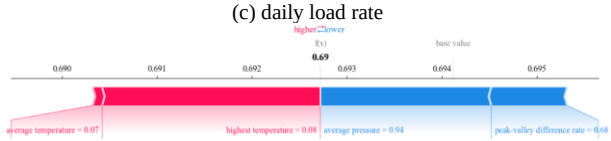
Fig.7. Daily Load Curve Forecast Results.



(a) maximum daily load



(c) daily load rate



(c) peak-valley difference rate

Fig.8. Visualization of explaining GRU.

TABLE VI
COMPARISON OF FORECASTING MODELS

Model	R2	EV	MAE(MW)	RMSE(MW)	MAPE(%)
LR	0.93	0.93	1120.65	1659.92	3.55
Ridge	0.93	0.93	1120.31	1659.66	3.55
Lasso	0.93	0.93	1120.31	1659.66	3.55
DT	0.84	0.88	1645.22	2451.15	4.89
RF	0.86	0.90	1548.73	2292.35	4.59
RNN	0.93	0.93	1133.48	1661.28	3.65
LSTM	0.93	0.93	1107.12	1654.89	3.54
GRU	0.93	0.94	1131.92	1610.64	3.52
Proposed Model	0.95	0.95	980.32	1334.52	3.16

C. Daily Load Curve Forecasting Evaluation

XAI models with explainability were selected for comparative validation again.

The forecast results are depicted in Fig. 7, with the left subplot representing the minimum load of 2022 and the right subplot representing the maximum load. The chosen models effectively track the original load, attributed to single-step predictions based on the actual same-hour load of the previous day. Notably, the proposed model's predictions align closely with the actual values, particularly in peak predictions, successfully capturing intraday load variations.

Table 6 shows the results of the prediction error evaluation. The linear models perform well overall since the inputs contain the true value of the load from the previous day. With a large number of data updates, Ridge and Lasso with regularization compute almost the same linear weights, which do not differ much from the LR. But for one day, the predictive effectiveness of the linear models declines. This is because, after a large amount of data training, the previous day's load values are

weighted very high, failing to reflect the changes in the next day's load. In contrast, the tree models, which are only obtained by training on previous years' data, do not forecast as well as the linear model overall, even though the inputs are the same. This is also because the tree models learn more about the feature inputs. But load has more complex variations, similar to load characterization indicator forecasting. Therefore, on the whole, the machine learning model forecasting ability decreases and is not as good as the linear model which is directly linked to the previous day's load. The RNN model, on the other hand, can effectively follow the load changes, as reflected by its performance in the evaluation metrics. Especially for RMSE, both LSTM and GRU outperform the other compared models. It means that the RNN is more advantageous when there are load fluctuations.

As a comparison, the proposed model can be optimized in all the metrics. Among them, R2 and EV are both 0.95, and all the indicators are improved compared to other models. MAE is reduced by 21.00% on average compared to other models, RMSE is reduced by 27.12% on average compared to other models, and MAPE is reduced by 0.70%. This indicates that the framework proposed in this paper has improved the prediction performance over common XAI models.

D. Explainability Analysis

1) Explainability of load characterization indicator forecasting.

The models or theories involved in the forecasting of load characteristic indicators are mainly related to RF and GRU.

The understanding of RF follows a model self-explanatory. The feature importance level can be understood in Fig.3. As an example when searching for the feature of daily peak load, the average temperature appeared most frequently in all DTs. Therefore, in the ranking of feature importance, the average temperature has a 45% share and is ranked the highest. This means that the average temperature influences the load values and should be given more attention in the study. Thus, tree models can be fitted with nonlinearities while their decision logic can be easily obtained.

GRU has a very complex network of internal connections that prevents a direct explanation of its predictive mechanisms. Fortunately, due to the simple structure of the proposed model, it is easy to understand the model with the help of post-model stage explanation. It has been verified that SHAP is more suitable for deep learning frameworks than LIME or SRC [22]. Force diagrams [51] are used to show the results of the visual model explanation on the highest load day of the summer season as Fig.8. In Fig.8-(a), for example, the values show a high level because the forecast is for the daily maximum load. For each feature, the figure demonstrates that times closer to evening are more inclined to increase load values. It's also closer to peak power demand. On the contrary, 8 a.m. tends to reduce the load. This is because many work hours in Hubei begin at 8:30 or 9:00 a.m., and the electricity use is in a gap before that time. By the same method, one can try to understand other GRUs.

2) Explainability of adaptive baseline daily load curve calculating.

Unlike GRU, ACNN has good explainability with pre-model stage explanation as well as a linear model. ACNN averages the historical load curves by weights and the selection of K is an artificially operable process. Both distance calculation and weight calculation are very intuitive. So it is straightforward to get the shape and weight of each selected curve. Compared to K-means or FCM, which are more complex computational processes, the ACNN method has superior results while being easier to understand.

3) Explainability of daily load curve optimization.

As for load curve optimization, it is a physically meaningful process. The optimization uses power system utility metrics based on one base curve. Unlike the black-box structure of AI models, this process is very intuitive. Each load characteristic indicator or the baseline daily load curve has a direct impact on the directly observable optimization results.

4) Explainability of the overall model.

In summary, the proposed model is highly explainable, as each step is grounded in physical meaning or an explainable method. The model is flexible and can adjust each part based on real-world significance, allowing it to adapt to different scenarios easily. This adaptability makes it well-suited to meet practical needs, ensuring that managers can fully understand the model's predictions and their practical implications.

V. Conclusion

This paper presented an explainable daily load curve forecasting method based on adaptive baseline curve optimization. The proposed method employs RF and GRU to select features and forecast the short-term load considering three characteristic indicators. According to the predictions, ACNN is proposed to calculate the adaptive baseline daily load curve. Therefore, the daily load curve can be obtained by optimizing the three indicators predictions and the baseline daily load curve. The experimental results have indicated that the proposed method can improve the prediction accuracy compared with the XAI models. The model performed optimally on the evaluation metrics such as MAE and RMSE. On the other hand, the proposed method has good explainability. Each of the three steps in the framework has a clear physical meaning, where the AI model can be understood through XAI methods.

In future research, we will further optimize GRU to develop models with more accuracy and explainability. For example, the proposal of KAN provides a new idea of modeling. Since forecast accuracy is affected by load characterization metrics, more relevant metrics will be used to describe load profile changes. But more metrics mean more predictive modeling, perhaps possibly with less predictive accuracy. How to accurately describe the load profile needs more discussion.

Overall, the method in this paper provides ideas for balancing accuracy and explainability. It is significant that build physically meaningful models instead of stacking models or increasing the complexity.

References

[1] L. J. Ge, Y. M. Xian, Z. G. Wang, B. Gao, F. J. Chi, and K. Sun, "Short-term load forecasting of regional distribution network based on generalized regression neural network optimized by grey wolf optimization algorithm," *CSEE Journal of Power and Energy Systems*, vol. 7, no.5, pp. 1093-1101, Sep. 2021.

[2] Z. R. Tian, W. C. Liu, W. Q. Jiang, and C. Y. Wu, "CNNs-Transformer based day-ahead probabilistic load forecasting for weekends with limited data availability," *Energy*, vol. 293, pp. 130666, Feb. 2024.
[3] K. Qu, G. Q. Si, Z. H. Shan, Q. Y. Wang, X. L. Liu, and C. Yang, "Forwardformer: Efficient Transformer With Multi-Scale Forward Self-Attention for Day-Ahead Load Forecasting," *IEEE Transactions on Power Systems*, vol. 39, no. 1, pp. 1421-1433, Jan. 2024.
[4] F. J. Han, X. H. Wang, J. Qiao, M. J. Shi, and T. J. Pu, "Review on Artificial Intelligence based load forecasting research for the new-type power system," *Proceedings of the CSEE*, vol. 43, no.22, pp. 8569-92, Nov. 2023.
[5] Y. X. Chen, C. J. Lin, Y. L. Zhang, J. F. Liu, and D. R. Yu, "Day-ahead load forecast based on Conv2D-GRU_SC aimed to adapt to steep changes in load," *Energy*, vol. 302, pp. 131814, Sep. 2024.
[6] W. J. Xiao, L. Mo, Z. X. Xu, C. Liu, and Y. C. Zhang, "A hybrid electric load forecasting model based on decomposition considering fisher information," *Applied Energy*, vol. 364, pp. 123149, Jun. 2024.
[7] M. Ghafoori, M. Abdallah and S. Kim, "Electricity peak shaving for commercial buildings using machine learning and vehicle to building (V2B) system," *Applied Energy*, vol. 340, pp. 121052, Jun. 2023.
[8] W. J. Choi and S. Lee, "Interpretable deep learning model for load and temperature forecasting: Depending on encoding length, models may be cheating on wrong answers," *Energy and Buildings*, vol. 297, pp. 113410, Oct. 2023.
[9] X. D. Liang, W. Z. Jun, X. J. Jun, and H. Q. Ran, "Data-model hybrid driven topology identification framework for distribution networks," *CSEE Journal of Power and Energy Systems*, vol. 10, no. 4, pp. 1478-90, Jul. 2024.
[10] Z. S. Xun, M. H. Rui, C. L. Jun, W. Bo, W. H. Xia, L. X. Zhu, and G. W. Zhong, "Short-Term Load Forecasting of an Integrated Energy System Based on STL-CPLE with Multitask Learning," *Protection and Control of Modern Power Systems*, vol. 9, no. 6, pp. 71-92, November 2024, doi: 10.23919/PCMP.2023.000101.
[11]R. Z. Lu, R. C. Bai, R. D. Li, L. J. Zhu, M. Y. Sun, F. Xiao, D. Wang, H. M. Wu, and Y. M. Ding, "A novel sequence-to-sequence-based deep learning model for multistep load forecasting," *IEEE Transactions on Neural Networks and Learning Systems*, doi: 10.1109/TNNLS. 2023.3329466.
[12] A. Sharma and S. K. Jain, "A novel two-stage framework for mid-term electric load forecasting," *IEEE Transactions on Industrial Informatics*, vol. 20, no. 1, pp. 247-255, Jan. 2024.
[13] T. J. Pu, J. Qiao, Z. X. Zhao, and Peng Zhao, "Research on interpretable methods of machine learning applied in intelligent analysis of power system (part I): Basic concept and framework," *Proceedings of the CSEE*, vol. 43, no. 18, pp. 7010-30, Sep. 2023.
[14] K. Lee, B. Park, J. Kim and J. Hong, "Day-ahead wind power forecasting based on feature extraction integrating vertical layer wind characteristics in complex terrain," *Energy*, vol. 288, pp. 129713, Feb. 2024.
[15] F. J. Han, T. J. Pu, M. Z. Li, and G, "Taylor Short-term forecasting of individual residential load based on deep learning and K-means clustering," *CSEE Journal of Power and Energy Systems*, vol. 7, no. 2, pp. 261-269, Mar. 2021.
[16] Y. Q. Ji, Y. B. Yan, P. He, X. M. Liu, C. S. Li, C. Zhao, and J. L. Fan, "CNN-LSTM short-term load forecasting based on the K-Medoids clustering and grid method to extract load curve features," *Power System Protection and Control*, vol. 51, no. 18, pp. 81-93, Sept. 2023.
[17] Y. He, F. J. Luo, M. Y. Sun and G. Ranzi, "Privacy-preserving and hierarchically federated framework for short-term residential load forecasting," *IEEE Transactions on Smart Grid*, vol. 14, no. 6, pp. 4409-4423, Nov. 2023
[18] X. Cui, M. Lee, C. Koo, and T. Hong, "Energy consumption prediction and household feature analysis for different residential building types using machine learning and SHAP: Toward energy-efficient buildings," *Energy and Buildings*, vol. 309, pp. 113997, Apr. 2024.
[19] L. P. Joseph, R. C. Deo, D. Casillas-Perez, R. Prasad, N. Raj, and S. Salcedo-Sanz, "Short-term wind speed forecasting using an optimized three-phase convolutional neural network fused with bidirectional long short-term memory network model," *Applied Energy*, vol. 359, pp. 122624, Apr. 2024.
[20] W. J. Chung and C. D. Liu, "Analysis of input parameters for deep learning-based load prediction for office buildings in different climate zones using eXplainable Artificial Intelligence," *Energy and Buildings*, vol. 276, pp. 112521, Dec. 2022.
[21] S. Pelekis, I. K. Seisopoulos, E. Spiliotis, T. Pountridis, E. Karakolis, S. Mouzakitis, and D. Askounis, "A comparative assessment of deep learning models for day-ahead load forecasting: Investigating key accuracy drivers," *Sustainable Energy Grids & Networks*, vol. 36, pp. 101171, Dec. 2023.
[22] R. Y. Yu, K. Zhang, B. Ramasubramanian, S. Jiang, S. Ramakrishna, and Y. H. Tang, "Ensemble learning for predicting average thermal extraction load of a hydrothermal geothermal field: A case study in Guanzhong Basin, China," *Energy*, vol. 296, pp. 131146, Jun. 2024.
[23] L. Liu, X. Hu, and J. Chen, "Embedded scenario clustering for wind and photovoltaic power, and load based on multi-head self-attention," *Protection and Control of Modern Power Systems*, vol. 9, no. 1, pp. 122-132, Jan. 2024.
[24] C. X. Wang, Y. Z. Zhou, Q. S. Wen, and Y. Wang, "Improving load forecasting performance via sample reweighting," *IEEE Transactions on Smart Grid*, vol. 14, no. 4, pp. 3317-20, July 2023.
[25] R. Carmona and X. S. Yang, "Joint granular model for load, solar and wind power scenario generation," *IEEE Transactions on Sustainable Energy*, vol. 15, no. 1, pp. 674-686, Jan. 2024.
[26] C. D. Fan, S. H. Nie, L. Y. Xiao, L. Z. Yi, and G. R. Li, "Short-term industrial load forecasting based on error correction and hybrid ensemble learning," *Energy and Buildings*, vol. 313, pp. 114261, Jun. 2024.
[27] Y. Y. Chu, P. Xu, M. X. Li, Z. Chen, Z. B. Chen, Y. B. Chen and W.L. Li, "Short-term metropolitan-scale electric load forecasting based on load decomposition and ensemble algorithms," *Energy and Buildings*, vol. 225, pp. 110343, Oct. 2020.
[28] M. R. N. Kalhori, I. T. Emami, F. Fallahi and M. Tabarzadi, "A data-driven knowledge-based system with reasoning under uncertain evidence for regional long-term hourly load forecasting," *Applied Energy*, vol. 314, pp. 118975, May. 2022.
[29] J. Q. Shi, C. X. Li, and X. H. Yan, "Artificial intelligence for load forecasting: A stacking learning approach based on ensemble diversity regularization," *Energy*, vol. 262, pp. 125295, Jan. 2023.
[30] J. -W. Xiao, P. Liu, H. L. Fang, X. -K. Liu and Y. -W. Wang, "Short-Term residential load forecasting with baseline-refinement profiles and bi-attention mechanism," *IEEE Transactions on Smart Grid*, vol. 15, no. 1, pp. 1052-1062, Jan. 2024.
[31] J. L. Fan, W. Zhuang, M. Xia, W. X. Fang, and J. Liu, "Optimizing attention in a Transformer for multihorizon, multienergy Load forecasting in integrated energy systems," *IEEE Transactions on Industrial Informatics*, doi: 10.1109/TII.2024.3392278.
[32] C. Wang, Y. Wang, Z. T. Ding, and K. F. Zhang, "Probabilistic multi-energy load forecasting for integrated energy system based on Bayesian Transformer network," *IEEE Transactions on Smart Grid*, vol. 15, no. 2, pp. 1495-1508, March 2024.
[33] L. Z. Ming, W. Y. Xuan, V. Sachin, R. Fabian, H. Janmes, S. Marin, H. Thomas Y., and T. Max, "KAN: Kolmogorov-Arnold Networks," 2024, *arXiv:2404.19756*.
[34] S. Khemraj, T. J. Diego, W. Z. Cheng, Z. Zongren, and K. G. Em, "A comprehensive and FAIR comparison between MLP and KAN for differential and networks," *Computer Methods in Applied Mechanics and Engineering*, vol. 431, pp. 117290, Nov. 2024.
[35] G. N. Li, Y. B. Wu, S. Yoon, and X. Fang, "Comprehensive transferability assessment of short-term cross-building-energy prediction using deep adversarial network transfer learning," *Energy*, vol. 299, pp. 131395, Jul. 2024.
[36] X. Z. Wang and B. Dong, "Physics-informed hierarchical data-driven predictive control for building HVAC systems to achieve energy and health nexus," *Energy and Buildings*, vol. 291, pp. 113088, May. 2023.
[37] J. A. Sward, T. R. Ault, and K. M. Zhang, "Genetic algorithm selection of the weather research and forecasting model physics to support wind and solar energy integration," *Energy*, vol. 254, pp. 124367, Sep. 2022.
[38] B. Yand, T. J. Zhu, P. L. Cao, G. X. Guo, C. Y. Zeng, D. Y. Li, Y. J. Chen, H. Y. Ye, R. N. Shao, H. C. Shu and T. Yu, "Classification and summarization of solar irradiance and power forecasting methods: A thorough review," *CSEE Journal of Power and Energy Systems*, vol. 9, no. 3, pp. 978-95, May. 2023.
[39] P. Borate, J. Riviere, C. Marone, A. Mali, D. Kifer, and P. Kifer, "Using a physics-informed neural network and fault zone acoustic monitoring to predict lab earthquakes," *Nature Communications*, vol. 14, no. 1, pp. 3693, Jun. 2023.
[40] C. J. Li, Z. Y. Dong, L. Ding, H. Petersen, Z. H. Qiu, G. Chen, and D. Prasad, "Interpretable memristive LSTM network design for probabilistic residential load forecasting," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 69, no. 6, pp. 2297-2310, June 2022.
[41] K. Qu, P. Li, Y. H. Huang and G. Q. Si, "Modeling and scenario generation method of annual load series for evaluation of renewable

energy accommodation capacity," *Automation of Electric Power Systems*, vol. 45, no. 1, pp. 123-131, Jan. 2021.
[42] C. Q. Kang, Q. Xia, N. D. Xiang, and M. Liu, "The study on long-term daily load curve forecasting," *Automation of Electric Power Systems*, vol. 20, no. 6, pp. 16-20, June. 1996.
[43] Y. Y. He, C. J. Cao, and J. L. Xiao, "Day-ahead peak load probability density forecasting based on QRLSTM-DF considering exogenous factors," *IEEE Transactions on Industrial Informatics*, vol. 19, no. 9, pp. 9447-9456, Sept. 2023.
[44] Terminology of energy saving for electric power, DL/T 1365-2014.
[45] C. Wan, W. K. Cui, and Y. H. Song, "Machine learning-based probabilistic forecasting of wind power generation: A combined bootstrap and cumulant method," *IEEE Transactions on Power Systems*, vol. 39, no. 1, pp. 1370-1383, Jan. 2024.
[46] L. M. Fan, H. L. Li, and X. P. Zhang, "Load prediction methods using machine learning for home energy management systems based on human behavior patterns recognition," *CSEE Journal of Power and Energy Systems*, vol. 6, no. 3, pp. 563-71, Sep. 2020.
[47] Y. M. Dai, Y. X. Wand, M. M. Leng, X. Y. Yang, and Q. Z, "LOWESS smoothing and Random Forest based GRU model: A short-term photovoltaic power generation forecasting method," *Energy*, vol. 256, pp. 124661, Oct. 2022.
[48] G. Z. Yu, L. Lu, B. Tang, S. Y. Wang, and Q. Dong, "Research on ultra-short-term subsection forecasting method of offshore wind power considering transitional weather," *Proceedings of the CSEE*, vol. 42, no. 13, pp. 4859-4871, Jul. 2022.
[49] Y. K. Semero, J. H. Zhang and D. H. Zheng, "PV power forecasting using an integrated GA-PSO-ANFIS approach and Gaussian process regression based feature selection strategy," *CSEE Journal of Power and Energy Systems*, vol. 4, no. 2, pp. 210-218, Jun. 2018.
[50] J. S. Zhang, L. Z. Ji, and M. Wang, "Information theoretical importance sampling clustering and its relationship with fuzzy c-means," *IEEE Transactions on Fuzzy Systems*, vol. 32, no. 4, pp. 2164-2175, April 2024.
[51] J. F. Liu, Z. H. Zhang, X. H. Fan, Y. Y. Zhang, J. Q. Wang, K. Zhou, S. Liang, X. Y. Yu, and W. Zhang, "Power system load forecasting using mobility optimization and multi-task learning in COVID-19," *Applied Energy*, vol. 310, pp. 118303, Mar. 2022.

Tian Ni (IEEE Student Member) received the B.S. degree in Measurement and Control Technology and Instruments from Wuhan University of Technology, Wuhan, Hubei,China, in 2019. He is currently working toward the M.S. degree in Electrical Engineering from School of Automation, Wuhan University of Technology, Wuhan, Hubei,China. His research interests include artificial intelligence, load forecasting, etc.

Xixiu Wu (IEEE Menber) received the Ph.D. degree in Electrical Engineering from the Huazhong University of science and technology, China. During 2012 and 2013, she was a visiting professor with the University of Liverpool University, UK for her work on gas insulated substation simulation. She is currently an Associate Professor with the School of Automation, Wuhan University of Technology, Wuhan, China. Her research interests include electromagnetic field simulation and power system digital twin.

Hui Hou (IEEE Senior Member, CSEE Senior Member) received the B. S. degree from Wuhan University, China, in 2003, and the Ph. D degree from the Huazhong University of Science and Technology, China, in 2009. She is currently an associate professor at School of Automation in Wuhan University of Technology. Her research interests include power system risk assessment, protection and security automatic control, etc.

Jingqi Xu (IEEE Student Member) received the B.S. degree in electrical and electronic engineering from Wuhan University of Technology, Wuhan, China. He is currently pursuing the M.S. degree in electrical and computer engineering in the University of Southern California, Los Angeles, USA. His research interests mainly include generated artificial intelligence and load forecasting.

Zhao Yang Dong (IEEE Fellow) received the Ph.D. degree in electrical engineering from the University of Sydney, Sydney, NSW, Australia, in 1999.He is the Chair Professor and the Head of Department of Electrical Engineering, the City University of Hong Kong, Hong Kong. His previous roles include the SPG Endowed Professor of Power Engineering and the Co-Director of SPG-NTU Joint Lab, Nanyang Technological University, Singapore, SHARP Professor and the Inaugural Director of UNSW Digital Grid Futures Institute, the Director of ARC Research Hub for Integrated Energy Storage Solutions, the Ausgrid Chair Professor, and the Director of Ausgrid Centre for Intelligent Electricity Networks providing research and development support for the AU$500m Smart Grid, Smart City National Demonstration Project of Australia. His research interest includes power system planning, load modeling, smart grid, smart cities, energy market, renewable energy and its grid connection, and computational methods and their application in power system analysis. Dr. Dong has been serving as the Editor/Associate Editor for several IEEE transactions and IET journals. He has won many research and industry grants nationally and internationally. Prof Dong is a Fellow of IEEE for his contributions in computational methods in power system planning and stability.

Luo Chao received the B.S. degree and Ph.D. degree from Wuhan University, Wuhan, China, in 2010 and 2018 respectively, then he joined Central Southern China Electric Power Design Institute (CSEPDI). His current research interests include power system planning and electrical power market.