



OPEN ACCESS

EDITED BY

Xingshuo Li,
Nanjing Normal University, China

REVIEWED BY

Yunting Yao,
Nanjing Normal University, China
Yeming Dai,
Qingdao University, China

*CORRESPONDENCE

Yong Li,
✉ liyong19860306@163.com

RECEIVED 21 October 2024

ACCEPTED 25 February 2025

PUBLISHED 18 March 2025

CITATION

Li Y, Wang H, Huang X, Hao J, Lei W and Wang Q (2025) Short-term power load forecasting in distribution networks considering human comfort level. *Front. Energy Res.* 13:1514755. doi: 10.3389/fenrg.2025.1514755

COPYRIGHT

© 2025 Li, Wang, Huang, Hao, Lei and Wang. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](#). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Short-term power load forecasting in distribution networks considering human comfort level

Yong Li*, Hua Wang, Xianbin Huang, Juanyang Hao, Wenting Lei and Quanlin Wang

Country State Grid Tianfu New Area Electric Power Supply Company, Chengdu, China

The growth of power demand and the increase of new energy penetration have resulted in a heightened necessity for the precision of short-term power load forecasting in distribution networks. The majority of current research on short-term load forecasting is focused on the improvement of algorithms, with relatively limited attention paid to meteorological factors. Furthermore, research in this area typically focuses on a single meteorological factor, namely, temperature, and does not sufficiently address the processing of meteorological data features. In light of the aforementioned considerations, this paper puts forth a human comfort model founded upon the ordering relationship analysis method and the entropy weight method. Furthermore, it employs the XGBoost algorithm to construct a short-term load forecasting model, utilizing the human comfort score and historical load data as inputs for forecasting the load. This approach is intended to enhance the precision of the load forecasting. The experimental results demonstrate that the proposed prediction model exhibits superior performance in short-term load forecasting, achieving a significantly higher level of accuracy than the baseline model. This model offers a notable advancement in practical forecasting applications.

KEYWORDS

human comfort, ordering relation analysis, short-term load prediction, XGBoost (extreme gradient boosting), entropy weigh method

1 Introduction

Accurate short-term power load forecasting is essential for the safe and reliable operation of power grids, serving as a crucial guide for the scientific planning of power generation scheduling and the optimization of resource allocation. However, achieving high forecast accuracy remains a significant challenge due to the inherent variability of weather conditions, the intermittency of renewable energy sources such as wind and solar, and the complexity of electricity consumption patterns. Enhancing the precision of short-term load forecasts, particularly under dynamic weather conditions, is pivotal for maintaining stable power system operation and effectively balancing supply and demand. Such improvements can provide more reliable decision support for power system scheduling, facilitate better coordination between renewable energy generation and load demand, and contribute to the efficient and sustainable management of the grid.

A substantial corpus of research findings has been accumulated with the objective of addressing the short-term power load forecasting problem. This body of work has witnessed a transition from the utilization of traditional mathematical statistical methods to the deployment of machine learning methodologies (Chongqing et al., 2017). The mathematical statistical methods that have been employed include regression analysis (Jidong et al., 2013), the autoregressive integral sliding average model (ARIMA) (Zhou et al., 2023), the exponential smoothing method (Ji et al., 2012), and so forth. These methods have the advantage of a simple underlying principle and a faster calculation speed; however, they are not without limitations in terms of their ability to handle nonlinear relationships between variables. In light of the rapid advancement of machine learning technology, scholars have increasingly directed their attention towards the utilization of machine learning in load forecasting. Among these, the Support Vector Machine (SVM) is one of the most commonly employed methods (Jiang et al., 2018). It utilizes a kernel function to address linear indivisible problems, exhibits an efficient processing ability in high-dimensional space, and gained considerable popularity during the 1990s. As deep learning technology has evolved, the recurrent neural network (RNN) has emerged as a prominent approach for processing time-series data, garnering significant attention and adoption in this domain. At present, the enhanced recurrent neural networks, exemplified by Long Short-Term Memory (LSTM) (Dai et al., 2023; Zhu et al., 2023) and Gate Recurrent Unit (GRU) (Dai et al., 2024), are extensively utilized for short-term load forecasting. The Sequence to Sequence (Seq2Seq) models, on the other hand, has been mostly used to deal with time series in recent years due to their flexibility in dynamically determining network step sizes and effectively capturing the relationships between sequences of varying lengths. Literature Dai et al. (2023) proposed an approach for short-term power load prediction by improving the Seq2Seq model based on bidirectional long-short term memory (Bi-LSTM) network, augmented by an attention mechanism to help the decoder focus on the key sequence information that affects the prediction result, thereby improving the load forecasting accuracy. Literature Dai and Yu (2024) employed Temporal Convolutional Network (TCN) as the encoder and decoder within the Seq2Seq framework, further optimizing the model parameters using the Bayesian Optimization (BO) algorithm to enhance model performance. Furthermore, Extreme Gradient Boosting (XGBoost) has demonstrated efficacy in numerous data mining competitions and is a prevalent tool in power system load forecasting (Chen and Guestrin, 2016; Chuanjie et al., 2023; Dai et al., 2022).

Overall, research on short-term load forecasting has mostly focused on algorithmic improvements, with various types of combined forecasting methods emerging and network structures becoming increasingly complex (Zhang et al., 2022; Wang et al., 2023; Wang and Chong, 2022). However, meteorological factors, as key variables with a significant impact on load, have been the subject of relatively limited attention in related studies. Within the extant framework of meteorological data processing, studies are often constrained to the consideration of temperature as an isolated variable. Furthermore, the majority of historical load data and meteorological data are directly introduced into the forecasting model as the input features of the model, which is insufficiently in-depth in the processing of meteorological data

features and fails to fully explore and integrate the synergistic effects of other multidimensional meteorological factors. For this reason, the literature Qin et al. (2006) introduced the concept of human comfort, conducted a comprehensive analysis of the role of temperature, humidity and wind speed on human comfort, and improved the training efficiency and prediction accuracy of the neural network prediction model. However, the universality of this comfort index model in different regions has yet to be verified. Literature GAO et al. (2017) proposed a novel human comfort concept, using hierarchical analysis and entropy weight method to construct a human comfort evaluation model and an improved random forest algorithm to establish a short-term load prediction model. Nevertheless, the evaluation model employed in this approach necessitates the manual rating of diverse meteorological indicators, which is a considerable undertaking.

To address the above problems, this paper quantifies the degree of affiliation of each meteorological indicator to human comfort by constructing a fuzzy affiliation function, and determines the reasonable weights of each meteorological indicator in the evaluation of human comfort by using the combination of the ordering relationship method and entropy weighting method, and then constructs a human comfort evaluation model that integrates multi-dimensional meteorological factors, such as temperature, humidity, wind speed, etc. On this basis, the XGBoost algorithm is employed to construct a short-term load forecasting model. This model takes the human comfort score and historical load data as its core input features, thereby improving the accuracy of the load forecasting model by integrating the human comfort factor. The model proposed in this paper was tested on an actual load dataset and demonstrated superior prediction accuracy in short-term load forecasting compared to the baseline model. This result validates the efficacy of the proposed model in short-term load forecasting.

The remaining sections of this paper are organized as follows. Section 2 introduces the fundamental concepts of the fuzzy affiliation function, ordinal relationship method and entropy weight method, which collectively form the basis of the human comfort evaluation method. Section 3 presents an overview of the fundamental concepts underlying the XGBoost algorithm and outlines the specific steps involved in the prediction method proposed in this paper. Section 4 presents the case study, including an overview of the experimental data, the evaluation criteria selected, and a comparison of the model results. Section 5 provides a summary of the paper and presents the conclusions.

2 Human comfort evaluation model based on ordering relationship analysis and entropy weight method

2.1 Traditional human comfort index

A notable correlation exists between temperature and load fluctuations, which are characterized by a relatively continuous, regular and less volatile pattern. In contrast, humidity and wind speed exhibit a high degree of randomness and significant fluctuations, which presents a challenge in directly studying their impact on load. Consequently, existing studies typically do not analyze the direct impact of each meteorological factor on load

separately. Instead, they tend to measure the coupling effect of multiple meteorological factors by introducing a meteorological composite index. The meteorological composite index is expressed in various forms, with the human comfort index being the most prevalent in studies of power systems.

The Human Comfort Index is a meteorological indicator that is employed to evaluate the extent to which humans can tolerate various weather conditions. It is constructed on the basis of the complex mechanism of heat exchange between the human body and the surrounding atmospheric environment, and comprehensively considers the joint influence of key meteorological parameters, such as temperature, relative humidity, wind speed, etc., in order to reflect the human body's perception of these meteorological elements. There are various forms of specific calculations of human comfort, and Equation 1 are used in Beijing, Nanjing and Hangzhou (Qin et al., 2006):

$$Conf = 1.8Temp + 0.55(1 - Hmd) - 3.2\sqrt{Wsp} + 27 \quad (1)$$

It should be noted, however, that the human comfort index has a specific scope of application. Furthermore, the universality of the model remains to be verified, due to the considerable meteorological differences that exist between different regions.

2.2 Human comfort evaluation model based on ordering relationship method and entropy weight method

The definition of human comfort is inherently multifaceted, encompassing subjective perceptions and objective environmental factors. This complexity renders its portrayal through a purely quantitative model challenging, as it is difficult to achieve precision and unambiguity in such a context. In this context, fuzzy evaluation theory, as an advanced mathematical tool, can achieve a scientific evaluation of the complex concept of comfort by combining qualitative and quantitative approaches.

2.2.1 Fuzzy affiliation function

The initial step in fuzzy evaluation is to ascertain the degree of affiliation of the indicator in question. This is achieved through the standardization of the indicator.

The fuzzy affiliation function represents a pivotal concept within the theoretical framework of fuzzy logic. It is employed to quantify the degree of attribution of elements within a fuzzy set to a specific feature or attribute. The construction of a fuzzy affiliation function entails the quantification of the degree of affiliation of a fuzzy concept, achieved through the mapping of the input value to a continuous range between 0 and 1. This range expresses the degree of fuzzy attribution of a thing or concept to a certain characteristic. In practice, the fuzzy affiliation function can be modelled based on a variety of mathematical curves or distributions, including the rectangular distribution function, trapezoidal distribution function, exponential distribution function, and Gaussian function. The adjustment of function parameters, including peak position, standard deviation, and slope, enables the capture and quantification of the affiliation of fuzzy concepts, thus facilitating the application of fuzzy logic to real-world problems with fuzzy, uncertain, or ambiguous boundaries.

In consideration of the characteristics exhibited by meteorological data, this paper employs a Gaussian combined affiliation function and a gradient affiliation function as the affiliation function for fuzzy evaluation, as illustrated in Figures 1A, B respectively.

The trapezoidal membership function can be expressed as Equations 2:

$$A = \max\left(\min\left(\frac{x - v_1}{v_2 - v_1}, 1, \frac{v_4 - x}{v_4 - v_3}\right), 0\right) \quad (2)$$

where v_1 and v_4 are the lower and upper limits of the variable, and v_2 and v_3 are the values at the ends of the interval, respectively.

The Gaussian combination membership can be expressed as Equations 3:

$$A = \begin{cases} e^{-\frac{(x-c_1)^2}{2\sigma_1^2}}, & x \leq c_1 \\ 1, & c_1 < x \leq c_2 \\ e^{-\frac{(x-c_2)^2}{2\sigma_2^2}}, & x \geq c_2 \end{cases} \quad (3)$$

where c_1 and σ_1 are the mean and standard deviation of the left Gaussian function, and c_2 and σ_2 are the mean and standard deviation of the right Gaussian function, respectively.

2.2.2 Ordering relationship analysis method

The ordering relationship analysis method is a technique for determining the relative importance of evaluation indicators based on the evaluator's subjective judgement. The method establishes the ordering relationship between evaluation indicators through the comparison of the relative importance of the indicators by the evaluator, thereby determining the weight of each indicator accordingly. In comparison with the hierarchical analysis method, which requires consistency checking, the ordering relationship method is more straightforward, requiring significantly less calculation and offering a simpler, more intuitive approach (Bin et al., 2022). Accordingly, this paper employs the ordering relationship method to ascertain the subjective weights, with the calculation process outlined as follows:

2.2.2.1 Determine the order relationship

In order to calculate the value of human comfort at m moments, it is necessary to consider m evaluation objects involving n meteorological indicators. In this context, the set of evaluation indicators is noted as $G = \{G_1, G_2, \dots, G_n\}$.

When the importance of G_i indicators is not lower than that of G_j indicators, it is noted as $G_i \geq G_j$. In the event that the aforementioned n indicators are in a state of relationship as Equations 4:

$$G_1 \geq G_2 \geq \dots \geq G_n \quad (4)$$

Then the indicator set $G = \{G_1, G_2, \dots, G_n\}$ is said to have established the ordering relationship in accordance with the ' \geq ' operator.

The ordering relationship is primarily determined through the evaluation of experts in relevant fields. Initially, experts select the first important evaluation indicators based on experience, which are recorded as G_1^* . Subsequently, experts select the first important evaluation indicators among the remaining indicators, and the ordering relationship can be determined as $G^* = \{G_1^*, G_2^*, \dots, G_n^*\}$.

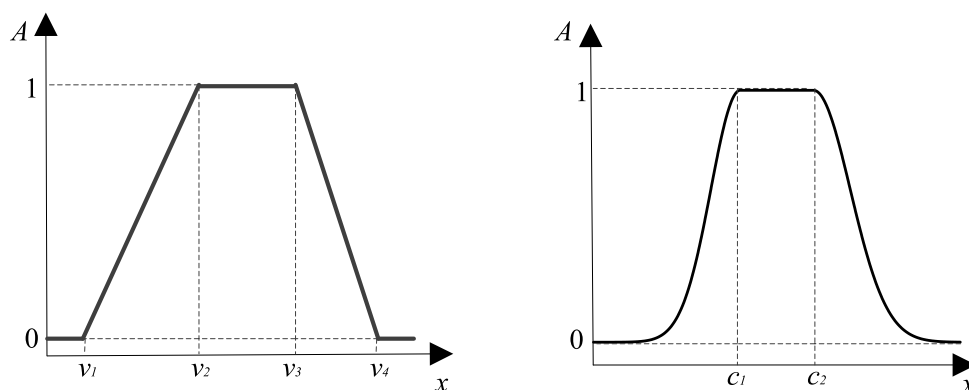


FIGURE 1
(A) Trapezoidal membership function (B) Gaussian combination membership function.

TABLE 1 Ordering relation assignment method.

r_j	Principle
1.0	U_{j-1} is as important as U_j
1.3	U_{j-1} is slightly more important than U_j
1.6	U_{j-1} is slightly more important than U_j
2.0	U_{j-1} is extremely more important than U_j

2.2.2.2 Quantitative analysis of differences in the importance of indicators

The importance of indicator G_j is expressed in terms of a subjective weight, w_{sj} , and the ratio of the importance of indicators G_{j-1} and G_j is denoted as r_j is calculated from Equation 5.

$$r_j = \frac{w_{s,j-1}}{w_{sj}}, j = 2, 3, \dots, n-1, n \quad (5)$$

The ordering relations of r_j can be assigned in accordance with the specifications outlined in Table 1.

2.2.2.3 Calculation of subjective weights

The subjective weights of the indicators are calculated based on the assignment of the experts to the ordering relations. Their relationship is shown in Equation 6

$$\begin{cases} w_{sn} = \frac{1}{1 + \sum_{j=2}^n \prod_{k=j}^n r_k} \\ w_{s,j-1} = r_j w_j, j = n, n-1, \dots, 3, 2 \end{cases} \quad (6)$$

According to the above equation, w_1 to w_{n-1} can be calculated to obtain the subjective weight vector $W_s = [w_{s1}, w_{s2}, \dots, w_{sn}]^T$ for each meteorological indicator.

2.2.3 Entropy weight method

Entropy weight method is an objective assignment method based on the information entropy theory, which determines the

weight of each indicator in the comprehensive evaluation by calculating the information entropy value of each evaluation indicator to measure the size of the information it contains.

The main steps for determining the objective weights through the entropy weighting method are as follows:

- For the set of meteorological indicators G , record the value of the j -th meteorological indicator at the i -th moment as u_{ij} , and construct the attribute matrix $U = (u_{ij})_{m \times n}$. Standardize the evaluation indicator values based on Equation 7, and construct the standardization matrix of the evaluation indicators:

$$q_{ij} = \frac{u_{ij} - u_{\min,j}}{u_{\max,j} - u_{\min,j}} \quad (7)$$

where q_{ij} is the indicator value after normalization of u_{ij} , $u_{\max,j}$ and $u_{\min,j}$ are the maximum and minimum values of the G_j indicators of all evaluation objects, respectively.

Thus, the normalization matrix of evaluation indicators is obtained from Equation 8:

$$Q = \begin{pmatrix} q_{11} & q_{12} & \dots & q_{1n} \\ q_{21} & q_{22} & \dots & q_{2n} \\ \dots & \dots & \dots & \dots \\ q_{m1} & q_{m2} & \dots & q_{mn} \end{pmatrix} \quad (8)$$

- Calculate the weight of the indicator value of the i -th evaluation object under the j -th indicator through Equation 9:

$$p_{ij} = q_{ij} / \sum_{i=1}^m q_{ij} (i = 1, 2, \dots, m; j = 1, 2, \dots, n) \quad (9)$$

Thus, the following matrix is obtained:

$$P = \begin{pmatrix} p_{11} & p_{12} & \dots & p_{1n} \\ p_{21} & p_{22} & \dots & p_{2n} \\ \dots & \dots & \dots & \dots \\ p_{m1} & p_{m2} & \dots & p_{mn} \end{pmatrix} \quad (10)$$

(iii) Calculate the entropy value of the j -th indicator in Equation 11.

$$e_j = -\frac{1}{\ln n} \sum_{i=1}^m p_{ij} \ln p_{ij}, (j = 1, 2, \dots, n) \quad (11)$$

(iv) Calculate the coefficient of variation of the j -th indicator in Equation 12.

$$g_j = 1 - e_j, (j = 1, 2, \dots, n) \quad (12)$$

(v) Calculate the weight of the j -th indicator in Equation 13.

$$w_{oj} = g_j / \sum_{j=1}^n g_j = (1 - e_j) / \left(n - \sum_{j=1}^n e_j \right), (j = 1, 2, \dots, n) \quad (13)$$

Thus, the objective weight vector $W_o = [w_{o1}, w_{o2}, \dots, w_{on}]^T$ is obtained.

2.2.4 Comprehensive scoring

There are obvious differences between the sequential relationship method and entropy weight method in the guiding ideology and the results of weight calculation. The sequential relationship method focuses on starting from the experts' experience and emphasizes the experts' judgement on the relative importance between the indicators, whereas the entropy weight method is more inclined to using data information and assigning values by calculating the information entropy and the weights between the indicators. In practical applications, the weights obtained by a single method may have limitations. Therefore, to comprehensively consider the subjective and objective weights and obtain more comprehensive and accurate information about the weights of the indicators, the Lagrange multiplier method is used for the combination of the assignment to obtain the comprehensive weights w_j as in Equation 14:

$$w_j = \frac{\sqrt{w_{sj} \cdot w_{oj}}}{\sum_{j=1}^n \sqrt{w_{sj} \cdot w_{oj}}} (j = 1, 2, \dots, n) \quad (14)$$

The ordinary multiplication and addition operator is employed to synthesize and obtain the human comfort evaluation score b_i at the moment being evaluated, as in Equation 15:

$$b_i = \sum_{j=1}^n (w_j \bullet A_j) \quad (15)$$

Subsequently, the human comfort scores $B = [b_1, b_2, \dots, b_n]$ is obtained for each moment.

3 Xgboost-based prediction model

This section may be divided by subheadings. It should provide a concise and precise description of the experimental results, their interpretation, as well as the experimental conclusions that can be drawn.

3.1 XGBoost

XGBoost (eXtreme Gradient Boosting) is one of the boosting algorithms. Its core idea is to integrate numerous weak classifiers into a single, robust classifier.

The fundamental elements of XGBoost are decision trees. These decision trees are the 'weak learners,' which together form the XGBoost. In the iterative process, the generation of the latter decision tree will take into account the prediction results of the previous decision tree. This entails that the bias of the previous decision tree will be taken into account, so that the training samples that were wrong in the previous tree will receive more attention in the subsequent period. This is followed by the training of the next tree, which is based on the adjusted sample distribution. The tree model employed is the Classification and Regression Tree (CART) model.

In the context of short-term load forecasting, the inputs to the forecasting model are P, representing historical load data; B, denoting human comfort values; and D, indicating holiday characteristics (represented by 0/1, with 0 denoting weekdays and 1 representing holidays). Using the prediction model $\theta(-)$, the following XGBoost model can be constructed by constructing the mapping from the prediction input x to the prediction object, as Equation 16:

$$\hat{y} = \theta(x) = \sum_{k=1}^K f_k(x), f_k \in F \quad (16)$$

where: K represents the number of trees, f_k denotes the function model of the k th tree, and F signifies the function space, which is constituted by decision trees.

The objective function of XGBoost is as Equations 17, 18:

$$X_{obj} = l(y, \hat{y}) + \sum_{k=1}^K \Omega(f_k) \quad (17)$$

$$\Omega(f_k) = \gamma T + \lambda \frac{1}{2} \|\omega\|^2 \quad (18)$$

where: $l(y, \hat{y})$ is the loss function; $\Omega(f_k)$ It is a regular term to prevent overfitting; T is the number of leaf nodes; a ω vector composed of regression values for the output of the leaf node; λ , is a hyperparameter. γ

During iteration, the objective function is updated to Equation 19:

$$\tau^{(k)} = l(y, \hat{y}^{(k-1)} + f_k(x)) + \Omega(f_k) \quad (19)$$

At $f_k = 0$, the Taylor second-order expansion of the loss function is performed, and the objective function is approximate as Equation 20

$$\tau^{(k)} \approx l(y, \hat{y}^{(k-1)} + f_k(x)) + \frac{1}{2} h f_k^2(x) + \Omega(f_k) \quad (20)$$

It can be inferred Equation 21:

$$\begin{aligned} X_{obj} &\approx \sum_{k=1}^K \left[g_j f_k(x) + \frac{1}{2} h_j f_k^2(x) \right] + \Omega(f_k) \\ &= \sum_{j=1}^T \left[g_j w_j + \frac{1}{2} (h_j + \lambda) w_j^2 \right] + \lambda T \end{aligned} \quad (21)$$

where: $g = \frac{\partial L(y, \hat{y}^{(k-1)})}{\partial \hat{y}^{(k-1)}}$, is the first derivative; $h = \frac{\partial^2 L(y, \hat{y}^{(k-1)})}{\partial (\hat{y}^{(k-1)})^2}$, which is the second derivative.

The optimal ω and objective function values are as Equations 22, 23:

$$\omega_j^* = -\frac{g_j}{h_j + \lambda} \quad (22)$$

$$X_{\text{obj}} = -\frac{1}{2} \sum_{j=1}^T \frac{g_j}{h_j + \lambda} + \lambda T \quad (23)$$

In the training process, the objective function gain is calculated, the leaf node with the largest gain loss is selected, and the complete decision tree is constructed until the limited number of layers is reached.

3.2 Overall process

The overall flow of this paper is shown in Figure 2. Firstly, a fuzzy affiliation function is constructed in order to quantify the degree of affiliation of each meteorological indicator with regard to human comfort. Subsequently, the combination of the ordering relationship method and the entropy weight method is employed in order to determine the reasonable weight of each meteorological indicator in the evaluation of human comfort, thus enabling the calculation of the value of the human comfort score. The data set comprising the human comfort score and the standardized load, when considered together, constitutes a training set. The XGBoost algorithm is then employed for model training. After the model training is completed, model inference is performed based on the weather forecast data and other feature data to derive the load prediction results on the forecast day.

4 Case analysis

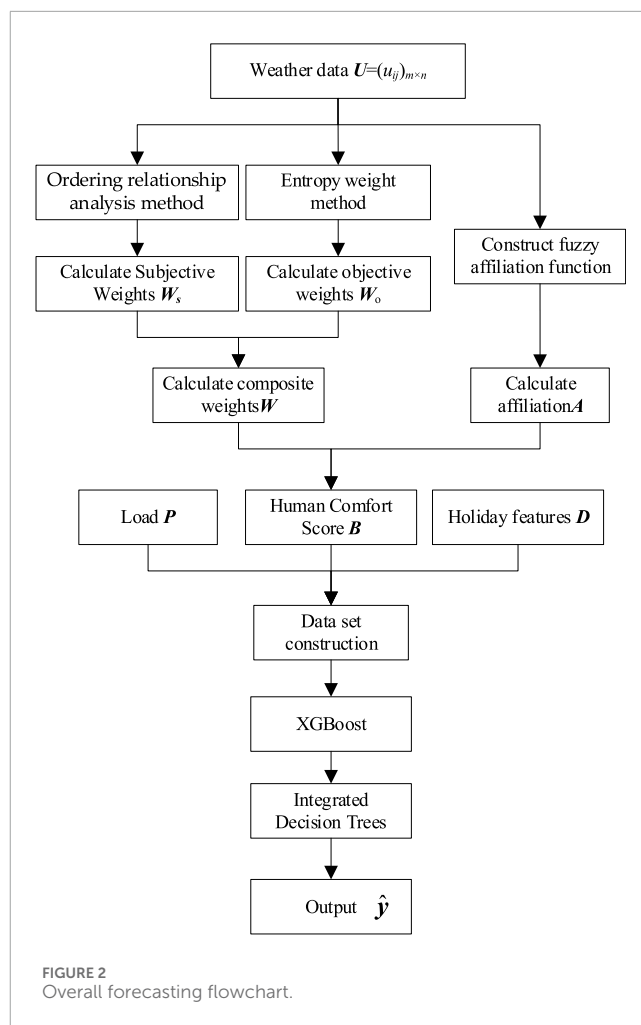
4.1 Case setting

The data set used for illustrative purposes is derived from a local power grid in Sichuan. It encompasses daily load data from July 2021 to October 2023, with a collection frequency of 15 min and a total of 96 measurement points. The meteorological data are sourced from Open-Meteo (Zippenfenig, 2023), with a time resolution of 1 h. Any missing values in the data set are supplemented by linear interpolation, and the load data are normalized by z-score.

The experimental platform employed in this study is an Intel(R) Core(TM) i7-7700 CPU with a main frequency of 3.60 GHz. The proposed method was implemented using the Python language, with the py-XGBoost framework for XGBoost and the PyTorch framework for FNN.

4.2 Human comfort evaluation

In consideration of the comfort criteria set forth by ASHRAE-55 (American National Standards Institute, 2023) and ISO 7730 (ISO, 2005), the affiliation of temperature was selected to adhere to the Gaussian combination affiliation function, with a mean and standard deviation of the left Gaussian function of 22 and 10, respectively, and a mean and standard deviation of the right



Gaussian function of 24. Affiliation functions for humidity and wind speed were also considered. The affiliation function for humidity was found to be a trapezoidal function with the ends of the interval at 30 and 60, while the Gaussian combined affiliation function was selected for wind speed, with a mean and standard deviation of 0. The values for the left Gaussian function are 2 and 0.8, while the values for the right Gaussian function are 5.4 and 8.

The sequential relationship method was employed to determine the subjective weights for the temperature, humidity, and wind speed of the three evaluation indicators. The expert scoring method was used to sort the indicators, and it was determined that the sequential relationship for the temperature, humidity, and wind speed is in accordance with the degree of importance, with $r_2 = 2$, $r_3 = 1$. The calculation of the subjective weights of the indicators yielded the following results: $W_s = [0.5517, 0.2759, 0.1724]$. Based on the entropy weight method to calculate the objective weights, the final result is $W_o = [0.3974, 0.3163, 0.2862]$. Then the combined weight $W = [0.4750, 0.2997, 0.2253]$.

Combining the affiliation and composite weights gives a score for human comfort, and the results of human comfort calculations at 12.PM on selected dates are extracted in Table 2.

As illustrated in the table, it can be observed that dates with specific meteorological conditions, such as 14 April and 21 October,

TABLE 2 Human comfort results on selected dates.

Date	Temperature/°C	Humidity/%	Wind speed/(m/s)	Human comfort score
2023.1.5	9.1	78	0.2	0.5779
2023.4.14	25.9	35	3.57	0.9768
2023.6.12	33.8	27	0.71	0.6202
2023.7.9	35.8	38	1.55	0.5936
2023.10.21	19.8	64	0.92	0.9699
2023.12.25	8.5	68	1.27	0.6341

have higher human comfort scores, as the temperature, humidity and wind speed are within the optimal range. In contrast, the results for 5 January and 25 December illustrate the impact of winter conditions on human comfort. On both days, moderate wind speeds were recorded, with the temperature being slightly higher on 5 January. However, the high humidity levels resulted in a greater sensation of coldness, leading to lower human comfort scores and a perception of greater discomfort on 5 January compared to 25 December. The results for 12 June and 9 July demonstrate that high temperatures, even when accompanied by relatively suitable humidity and wind speed conditions, can significantly impair human comfort in a hot summer environment. The proposed human comfort evaluation model is effective in quantifying and accurately describing the complex and ambiguous concept of human comfort by integrating and unifying meteorological indicators of different scales.

The Pearson Correlation Coefficient (PCC) is employed as a feature selection method to quantify the correlation between load and each meteorological index, thereby identifying the features with a strong correlation with load. The correlation coefficient between two random variables $X = [x_1, x_2, \dots, x_n]$ and $Y = [y_1, y_2, \dots, y_n]$, is calculated using Equation 24:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (24)$$

where, \bar{x} and \bar{y} are the sequence mean values of sequence X and Y respectively, x_i and y_i are the i -th element of sequence X and Y respectively, r denotes the correlation coefficient of sequence X and Y , the value of which ranges from -1 to 1 , $r > 0$ means that sequence X and Y show positive correlation, and $r < 0$ means that sequence X and Y show negative correlation, and the closer the absolute value of r is to 1 , it means that the correlation relationship between sequence X and Y is stronger.

The correlation coefficients between each meteorological data set and load data were calculated separately in order to obtain the heat map, as illustrated in Figure 3. As illustrated in the figure, the correlation coefficient between load and human comfort score is the highest, at 0.7002 . This value is higher than that of any single meteorological indicator and also higher than that of the traditional

human comfort index. This indicates that the proposed human comfort evaluation model effectively combines each meteorological index and has the strongest correlation with load, and the training algorithm can more easily capture the high correlation between load and human comfort score, which is more conducive to the training of the model.

4.3 Comparison of results

In order to provide a quantitative evaluation of the deterministic prediction performance of the model, this paper employs three indicators: mean absolute percentage error (MPAE), root mean square error (RMSE), and coefficient of determination (R^2). The first two are evaluation indices for the effect of error, with a smaller value indicating a higher level of prediction accuracy. The latter is an evaluation index for the effect of fitting, with a larger value indicating a stronger prediction model. The specific calculation formula is as Equations 25–27:

$$E_{MAPE} = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| / y_i \quad (25)$$

$$E_{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (26)$$

$$R^2 = 1 - \sum_{i=1}^N (y_i - \hat{y}_i)^2 / \sum_{i=1}^N (y_i - \bar{y})^2 \quad (27)$$

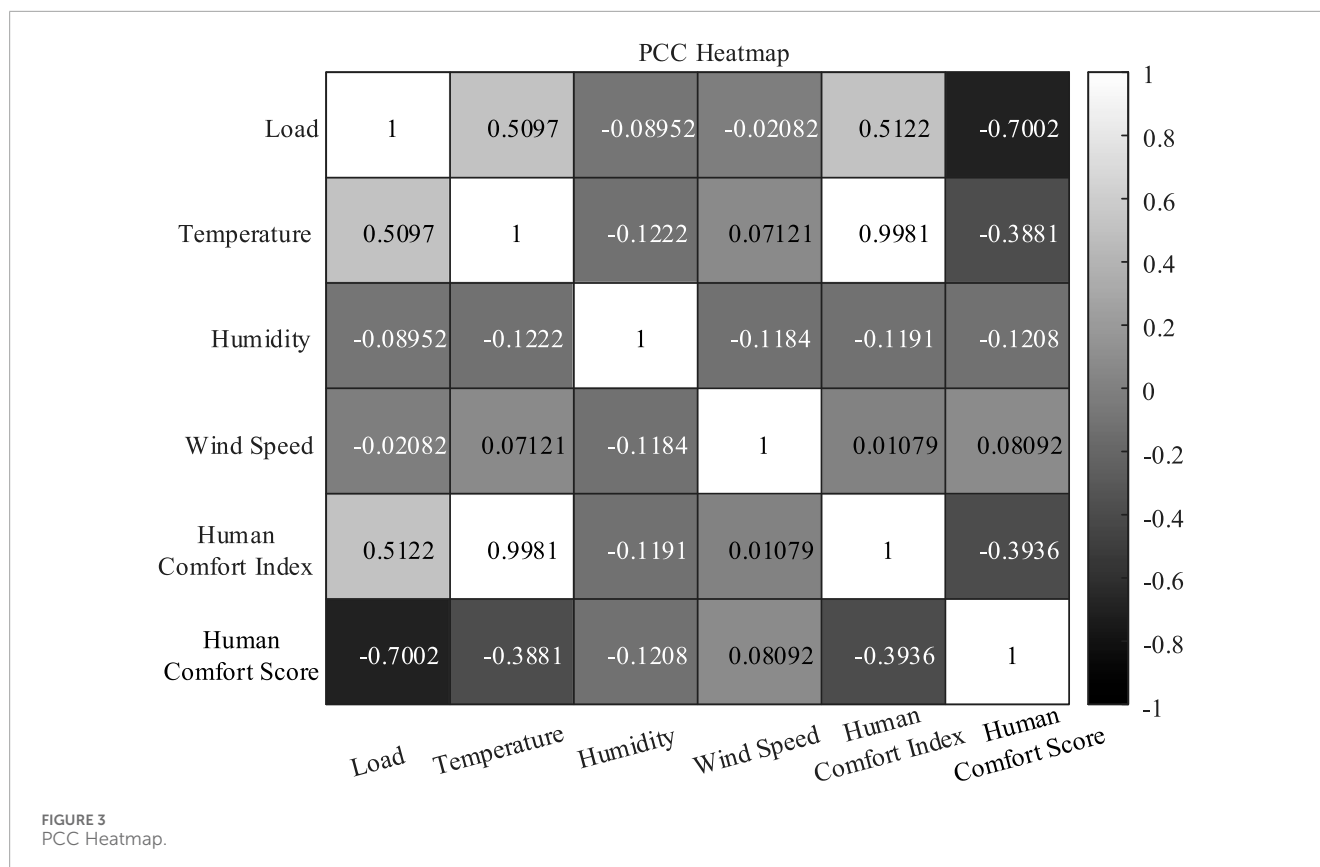
Meanwhile, in order to verify the effectiveness of the model proposed in this paper, FNN (Feed-forward Neural Network) is selected as the baseline model to compare and analyze with the model in this paper.

For this purpose, the prediction model 1 is set: the training algorithm adopts XGBoost algorithm, and the dataset is only load data and daily feature data.

Prediction model 2: The training algorithm adopts XGBoost algorithm, and the dataset is load data and temperature data.

Predictive model 3: The training algorithm adopts XGBoost algorithm, and the dataset is load data and traditional human comfort index data.

Predictive model 4: The model of this paper, the training algorithm adopts XGBoost algorithm, the dataset is load data and human comfort score proposed in this paper.



Predictive model 5: FNN model is used, the dataset is load data and human comfort score proposed in this paper.

The experiments are divided into training set and test set according to 0.8:0.2, and the XGBoost learning rate is set to 0.1, the number of iterations is set to 100, the range of random sampling ratio is set to 0.3, the maximum depth of the tree is set to 5, the minimum weight of the leaves is set to 10, the sample sampling ratio is set to 0.3, and the number of iterations is set to 100. The training algorithm of the FNN model is BP back-propagation, which is made up of three fully-connected layers, with 1,024 nodes of the first hidden layer, and the number of nodes of the first hidden layer is 1,024. Hidden layer nodes is 1,024 and the second hidden layer nodes is 512, ReLU is used for the activation function, Adam optimizer is used for the optimizer, the learning rate is 0.001, and the maximum training period is 1,000. The results of each prediction model on the test set are shown in Table 3.

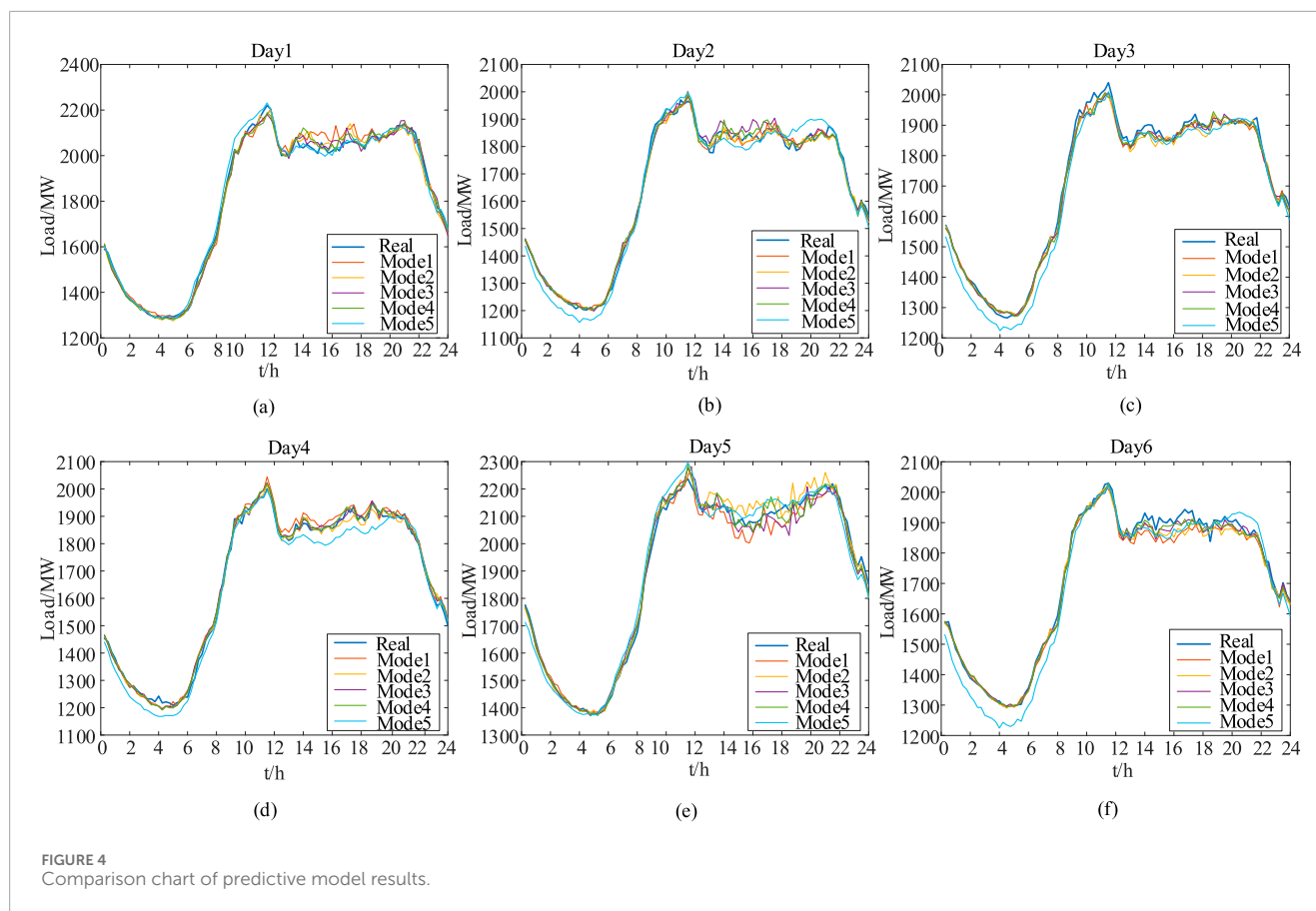
As evidenced by the data presented in the table, the model proposed in this study demonstrates the highest level of prediction accuracy. In comparison to the baseline model, the mean absolute percentage error (MAPE) was reduced by 1.6877, the root mean square error (RMSE) was reduced by 56.5554, and the coefficient of determination (R^2) was improved by 0.06. These results demonstrate a notable enhancement in the prediction accuracy. Furthermore, an evaluation of the same algorithm for Models 1, 2, 3, and 4 demonstrated that the human comfort assessment model exerted a beneficial influence on the enhancement of prediction accuracy to a certain extent. Model 2 reduced the error compared to model 1, indicating that temperature has a significant influence

TABLE 3 Predictive Model comparison results.

Predictive models	$E_{MAPE}/\%$	E_{RMSE}	R^2
1	3.5651	108.5776	0.9186
2	3.2242	94.1996	0.9408
3	2.9363	81.3866	0.9574
4	2.9095	79.1073	0.9600
5	4.5972	135.6627	0.9005

in load forecasting, and increasing temperature data can improve the accuracy of load forecasting. Model 3 has a smaller error compared to model 2, indicating that the traditional human comfort index is effective in comprehensively evaluating meteorological data, which is more effective than the single use of temperature data. Model 4, based on Model 3, not only consolidates the aforementioned trend but also achieves a further improvement in prediction accuracy through the introduction of the human comfort evaluation model proposed in this study. The performance of this model exceeds that of the traditional human comfort index model, thereby verifying the effectiveness of the model presented in this paper.

Figures 4A-F, 5A-F show the prediction results and error comparisons for each prediction model for the random 6 days in the validation set, respectively. As illustrated in the figure, each model



learnt the shape of the load curve better, and the trend of load change was basically consistent with the true value. Among them, the FNN model deviates most significantly from the true value in the load trough period, and its performance is the most unsatisfactory. In contrast, the model proposed in this paper has the highest fit with the real value in most of the time, and the best prediction accuracy. On the whole, the model in this paper performs well in learning the shape of load curve and predicting the trend of load change, and achieves better prediction accuracy, which verifies the effectiveness of this paper's model in short-term load forecasting. The model proposed in this paper can more accurately capture the dynamic characteristics of load change, provide more reliable prediction data for grid scheduling, so as to reasonably allocate various types of generation resources, optimise the scheduling strategy of renewable energy, improve the capacity of renewable energy consumption, reduce frequency fluctuations and power imbalance caused by load fluctuations, improve the reliability and flexibility of the power grid, and have a strong practical application value.

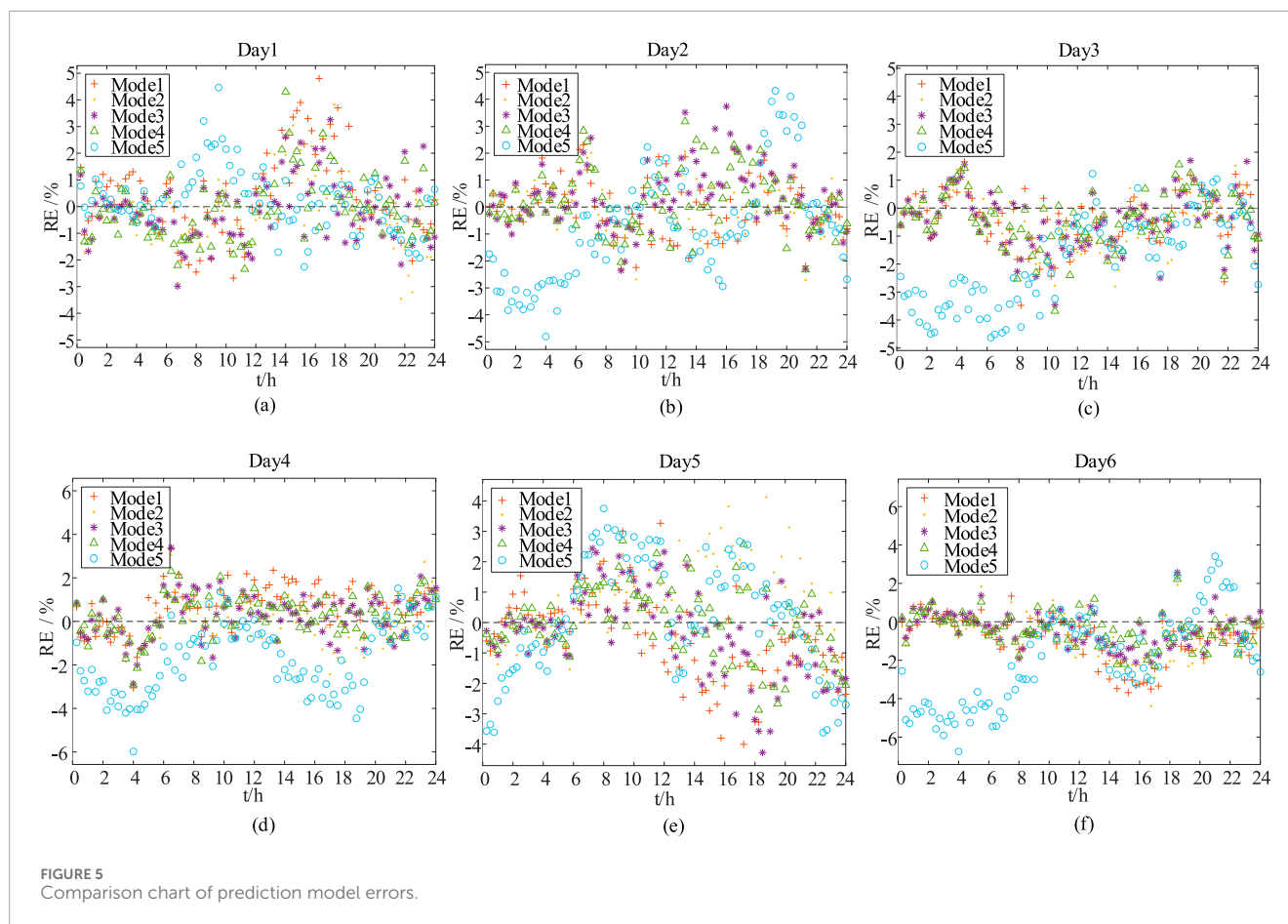
5 Conclusion

This paper proposes a human comfort model that comprehensively considers various meteorological indicators such as temperature, humidity and wind speed, constructs a human comfort model based on the combination of ordering relationship method and entropy weight method, and establishes a short-term

load prediction model based on the XGBoost algorithm. The following conclusions are obtained by analyzing the examples:

- (1) The human comfort evaluation model based on the ordering analysis method and entropy weighting method proposed in this paper integrates the expert experience and data information, and the index weighting information is more comprehensive and accurate. Compared with the traditional human comfort index prediction model, the prediction model in this paper achieves higher prediction accuracy, indicating that the human comfort evaluation model proposed in this paper has stronger applicability in load forecasting.
- (2) The model in this paper performs well in learning the shape of the load curve and predicting the trend of load change, and achieves better prediction accuracy, which verifies the effectiveness of this paper's model in short-term load prediction.

Although our proposed short-term load forecasting method has improved the forecasting accuracy to a certain extent, it still faces a number of key issues that need to be addressed. The first challenge lies in the method's strong dependence on accurate weather forecast data, and whether it can effectively identify errors and maintain high forecast accuracy when faced with inaccurate weather forecast data needs to be further explored. Secondly, although we validated the effectiveness of the proposed method by training and testing it on a real dataset in a region of Sichuan, the broad applicability and stability of the method in more diversified application scenarios



still need to be thoroughly verified. For this reason, in order to comprehensively evaluate and demonstrate the general applicability of the proposed method, we plan to extend it to a wider range of datasets and other related application areas such as photovoltaic, wind energy prediction, and electricity price prediction in our future work. In addition, we will work on further optimizing the proposed prediction model by integrating more advanced intelligent optimization algorithms to make it more flexible and adaptable to changing and complex scenarios, thus improving its predictive performance and robustness.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

YL: Formal Analysis, Methodology, Resources, Validation, Writing—original draft, Writing—review and editing. HW: Funding acquisition, Resources, Writing—review and editing. XH: Methodology, Project administration, Software, Writing—review and editing. JH: Conceptualization, Project administration,

Visualization, Writing—review and editing. WL: Data curation, Investigation, Project administration, Supervision, Writing—review and editing. QW: Conceptualization, Formal Analysis, Writing—review and editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This research was funded by State Grid Sichuan Electric Power Company Science and Technology Project (No. 512920230001).

Acknowledgments

The authors sincerely acknowledge the contribution of all individuals, reviewers, and editors for their contribution towards the production of this manuscript.

Conflict of interest

Authors YL, HW, XH, JH, WL, QW were employed by State Grid Tianfu New Area Electric Power Supply Company. The authors declare that this study received funding from State Grid Sichuan

Electric Power Company. The funder had the following involvement in the study: study design, data collection, analysis, interpretation of data, and the preparation of the manuscript.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

References

- American National Standards Institute (2023). *ANSI/ASHRAE standard 55-2023: thermal environmental conditions for human occupancy*. Atlanta: American Society of Heating Refrigerating and Air-conditioning Engineers, inc.
- Bin, N. A. N., Dong, S., Chengsi, X. U. T., and Ang, K. (2022). Comprehensive identification of critical lines in power grid based on improved maximizing dispersions method. *Power Syst. Technol.* 46 (10), 4076–4084. doi:10.13335/j.1000-3673.pst.2021.2405
- Chen, T., and Guestrin, C. (2016). “XGBoost: A scalable tree boosting System,” in *ACM SIGKDD international conference on Knowledge Discovery and data mining* (San Francisco, USA: ACM), 785–794.
- Chongqing, K., Qing, X. I. A., and Mei, L. I. U. (2017). *Power system load forecasting*. Beijing: China Electric Power Press.
- Chuanjie, M. A., Sun, Y., Peng, D., and Zhao, H (2023). Multivariate load forecasting for integrated energy system based on XGBoost-MTL. *Electr. Power Eng. Technol.* 42 (5), 158–166.
- Dai, Y., Yang, X., and Leng, M. (2023). Optimized Seq2Seq model based on multiple methods for short-term power load forecasting. *Appl. Soft Comput.* 142, 110335. doi:10.1016/j.asoc.2023.110335
- Dai, Y., and Yu, W. (2024). Short-term power load forecasting based on Seq2Seq model integrating Bayesian optimization, temporal convolutional network and attention. *Appl. Soft Comput.* 166, 112248. doi:10.1016/j.asoc.2024.112248
- Dai, Y., Yu, W., and Leng, M. (2024). A hybrid ensemble optimized BiGRU method for short-term photovoltaic generation forecasting. *Energy* 299 (15), 131458. doi:10.1016/j.energy.2024.131458
- Dai, Y., Zhou, Q., Leng, M., Yang, X., and Wang, Y. (2022). Improving the Bi-LSTM model with XGBoost and attention mechanism: a combined approach for short-term power load prediction. *Appl. Soft Comput.* 130, 109632. doi:10.1016/j.asoc.2022.109632
- Gao, Y., Sun, Y., Yang, W., Chuai, B., and Liang, H. (2017). Weather-sensitive load's short-term forecasting research based on new human body amenity indicator. *Proc. CSEE* 37 (07), 1946–1955. doi:10.13334/j.0258-8013.pcsee.160278
- ISO Ergonomics of the thermal environment -Analytical determination and interpretation of thermal comfort using calculation of the PMV and PPD indices and local thermal comfort criteria (2005).
- Ji, P., Xiong, D., Wang, P., and Chen, J. (2012). “Study on Exponential Smoothing Model for Load Forecasting,” in 2012 Asia-Pacific Power and Energy Engineering Conference, Shanghai, China, 1–4. doi:10.1109/APPEEC.2012.6307555
- Jiang, M., Gu, D., Kong, J., and Tian, Y. (2018). Short-term load forecasting model based on online sequential extreme support vector regression. *Power Syst. Technol.* 42 (7), 2240–2247.
- Jidong, L. I. U., Han, X., Chengbo, C. H. U., and Zhang, L. (2013). Cooling load of summer grid considering non-meteorological factors. *Electr. Autom. Equip.* 33 (2), 40–46.
- Qin, H., Wang, W., Hui, Z., and Wang, L. (2006). Short-term electric load forecast using human body amenity indicator. *Proc. CSU-EPSA* 18 (2), 63–66.
- Shu, Y. B., Chen, G. P., He, J. B., and Zhang, F. (2021). Building a new electric power system based on new energy sources. *Strategic Study CAE* 23, 61–69. doi:10.15302/j-sscae-2021.06.003
- Wang, J., and Chong, D. U. (2022). Short-term load prediction model based on Attention-BiLSTM neural network and meteorological data correction. *Electr. Power Autom. Equip.* 42 (04), 172–177+224.
- Wang, W., Bin, F., Gang, H., et al. (2023). Short-term net load forecasting based on self-attention encoder and deep neural network. *Proc. CSEE* 43 (23), 9072–9084.
- Zhang, S., Jun, L. I., Anqi, J., Huang, J., and Liu, H. (2022). A novel two-stage model based on FPA-VMD and BiLSTM neural network for short-term power load forecasting. *Power Syst. Technol.* 46 (08), 3269–3279. doi:10.13335/j.1000-3673.pst.2021.0969
- Zhou, Y., Kun, S. H. I., Dezhi, L. I., Chen, S., and Dou, X. (2023). Calculation of load aggregator potential and peak regulation strategy based on longitudinal modified ARIMA. *Electr. Power Eng. Technol.* 42 (2), 2–10.
- Zhu, J., Miao, Y., Dong, Z., Chen, H., and Li, S. (2023). Short-term load forecasting method based on Attention-LSTM and multi-model integration. *Electr. Power Eng. Technol.* 42 (5), 138–147.
- Zippenfenig, P. (2023). Open-Meteo.com weather API. *Zenodo*. [Computer software]. doi:10.5281/ZENODO.7970649

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.