Expert Systems With Applications 211 (2023) 118649

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Multi-step ahead forecasting for electric power load using an ensemble model

Yubo Zhao [a,b,1], Ni Guo [a,1,2], Wei Chen [a,c,*,3], Hailan Zhang [a], Bochao Guo [a], Jia Shen [a], Zijian Tian [a]

[a] School of Mechanical Electronic & Information Engineering, China University of Mining and Technology (Beijing), Beijing 100083, China
[b] Shaanxi Coal Industry Company Limited, Xi'an 710076, Shaanxi Province, China
[c] School of Computer Science & Technology, China University of Mining and Technology, Xuzhou 221116, China

ARTICLE INFO

Keywords:
Multi-step prediction
Electric power load
Ensemble models
NeuralProphet
Lightgbm

ABSTRACT

The multi-step prediction of electric power load is a crucial technology to promote power grid intelligence. Precise forecasting of short-term electric power will enhance the meticulous distribution management level of the grid and the dynamic balance between the supply and demand side of the power system. Conversely, a poor forecast of electric power load will lead to the unreasonable power supply to the grid. However, the electric power load is characterized by strong volatility and randomness, and it is a great challenge to precisely grasp the complex non-linear patterns hidden in power load sequences. For making the grid distribution more rational and efficient, a more precise electric power load prediction model needs to be explored. Thus, an ensemble model NeuralProphet-Lightgbm which combines the advanced NeuralProphet model and Lightgbm model is proposed in this paper. The experiments with different prediction horizons of 12, 24 and 48 h were conducted in this paper, and the root mean square error (RMSE), mean squared error (MSE), mean absolute percentage error (MAPE) and standard deviation (SD) are used as assessment metrics for prediction performance. In order to verify the effectiveness of the proposed model, the NeuralProphet-Quantile Regression Forest (QRF), Prophet, Autoregressive Integrated Moving Average model (ARIMA), Prophet-Lightgbm, NeuralProphet, Long Short Term Memory (LSTM) and LazyProphet models are used as benchmark models. The results of the proposed model significantly outperforms than the comparative models and is very robust as prediction horizons increasing. It is a good choice for efficiently forecasting electric power load.

1. Introduction

Electricity is an important source to drive the steady development of industries, featuring instantaneous generation and incapable of being stored in large quantities. To cope with crises like fossil fuel depletion, climate variability and ambient pollution, boosting the utilization of electricity resources has become an important solution to the problem (Kudo, Takeuchi, Nozaki, Endo, & Sumita, 2009).

The application of the Power Internet of Things brings an explosion of power data, and using the big data technology to mine the information value embedded in the data will promote the intelligence of the power grid (Bedi, Venayagamoorthy, Singh, Brooks, & Wang, 2018; Zhong, Chen, Dan, & Rezaeipanah, 2022). Electric power forecasting plays a crucial role in modern power supply systems, it is the basis for rational dispatch of electric power (Ahmad, Ghadi, Adnan, & Ali, 2022). In addition, the power generated by new energy sources will be integrated into the grid system on a large scale and applied to people's daily life, and these changes break the traditional fixed power supply pattern. These new changes bring more complexity and randomness to the grid, which poses a huge challenge to the traditional power supply and requires the enhancement of prediction technologies to respond in a quick manner to these new changes.

* Corresponding author at: School of Mechanical Electronic & Information Engineering, China University of Mining and Technology (Beijing), Beijing 100083, China.
E-mail address: chenwdavior@163.com (W. Chen).
[1] Co-first authors.
[2] ORCID: 0000-0002-5687-3973.
[3] ORCID: 0000-0002-7663-278X.

https://doi.org/10.1016/j.eswa.2022.118649
Received 27 June 2022; Received in revised form 17 August 2022; Accepted 19 August 2022
Available online 24 August 2022

Multi-step prediction is always a challenge in forecasting area due to the relevance of historical data to future trends decreases as the forecast horizon increases and the temporal features expressed gradually weaken, which results in higher forecast errors (Shamshirband et al., 2019). The publications in this field present a multitude of prediction approaches, mainly categorized into two main groups. One is a multivariate regression model based on the varying factors influencing the electric power, and this method often leads to high forecast precision and good robustness, but often requires the assistance of astronomical meteorological services (Abdel-Aal, 2008). The other group is the univariate time series prediction, i.e., mining potential forecasting information hidden in historical data that has occurred and existed. The large-scale univariate time series with high frequency is more difficult due to the high volatility and complex trend components mixed in time series, such as seasonality and periodicity, as well as the customers' behavioral habits (Liu, Ding, & Jia, 2020). Short-term forecasting of electric power load is of great importance in power sector, and an ensemble model namely NeuralProphet-Lightgbm is proposed in this paper for 12 h, 24 h and 48 h ahead forecasting of univariate electric power load.

There are two common ways of combining models on univariate time series, one is incorporating decomposition techniques to decouple the original sequence and then sum up the predictions of the sub-series separately to produce the final predicted values (Xia, Zhang, & Cao, 2018; Yan, Yu, & Bai, 2021), the other is the two-layer structure model, where the predicted values of the model are used as input set of another model (Lopez-Martin, Sanchez-Esguevillas, Hernandez-Callejo, Arribas, & Carro, 2021).

This paper provides a new way of feature reconstruction on univariate time series which differs from the aforementioned ways. Firstly, it decouples the original sequence using the advanced NeuralProphet model to yield a clear sub-sequence for the purpose of noise reduction, and then using the sub-sequence and predicted values of base model together as the high-level feature input to the second layer model for further feature extraction by employing the Lightgbm model. This way of feature reconstruction is rare in forecasting field.

The main contributions of this paper are mainly in the following aspects:

- The NeuralProphet model is firstly employed in the field of electric power load forecasting. It enables to decompose the time series into quality model components to reduce the complexity of univariate time series and could achieve both decoupling and prediction of sequences compared to a single decomposition technique, whereas a single decomposition technique usually needs to be incorporated with a prediction model to obtain future predicted values.
- This paper provides a new way of feature reconstruction that the sub-sequences of the raw data decoupled by the NeuralProphet model are used as feature input in the second layer model, significantly reducing the information loss of the original data.
- The model sub-sequences are taken as features in this paper rather than being further predicted, which retains the original information in a large degree, ensuring the prediction efficiency and accuracy of the model simultaneously.
- The structure of the proposed model is highly interpretable and its predictive performance greatly outperforms that of traditional models. In addition, the proposed model has high robustness as the prediction horizon increases.

The chapters of this paper are structured as follows: the second part presents the relevant publications in prediction domain, the principles of the relevant methodologies are demonstrated in the third part. The next section describes the procedure of model establishment and hyper-parameter tuning, and the experiment results are discussed in section 5. The conclusion and the outlook on the future research are summarized in the last section.

2. Related literatures

This section illustrates the latest research developments in the field of time series forecasting, which is popular with deep learning models, statistical models, machine learning models and ensemble models. These models can be applied not only to electrical power prediction but also to other different fields.

Among the deep learning models, the long short-term memory network (LSTM) is one of the most popular time series prediction methods, which was proposed by Hochreiter and Schmidhuber (Hochreiter & Schmidhuber, 1997) in 1997 and has been improved and optimized by many researchers afterward. Kim et al. proposed an model named long short-term memory-recurrent neural network (LSTM-RNN) based on very short term -photovoltaic power generation forecasting (VST-PVGF) framework to estimate the photovoltaic (PV) power generation for smart city energy management and the results imply that the multiscale LSTM is outperformed the other comparison forecasting models (Kim, Kwon, Park, Kim, & Cho, 2020). Wang et al. proposed a multi-step time series prediction model based on the dilated convolution network and long short-term memory, the proposed dilated convolution-long short-term memory (DC-LSTM) model provides a new method for the prediction of chaotic time series and lays a foundation for scientific data analysis of chaotic time series monitoring systems (R. Wang, Peng, Gao, Gao, & Jiang, 2020). The deep learning models are good options in cases where computational power allows, but they are not suitable to be considered when computational power is limited, as the model requires huge calculate memory consumption and time to wait.

Statistical models comprise widely used ARIMA model (Che & Zhai, 2022; Eldali, Hansen, Suryanarayanan, & Chong, 2016), Holt-Winters model (Jiang, Wu, Gong, Yu, & Zhong, 2020; Ma et al., 2020) etc.. Zou et al. proposes a combined method of electric load prediction based on BP neural networks and ARIMA model, the results has yield better prediction precision than single BP neural networks and ARIMA model (Zou, Wu, Zhao, Wang, & Zhou, 2018). Nevertheless, the employment of the ARIMA model often requires to knowledge of statistics for users. In recent years, the Prophet model has been a focus for many researchers (Shohan, Faruque, & Foo, 2022; Yenidoğan, Çayir, Kozan, Dağ, & Arslan, 2018), its remarkable attributes are the high interpretability and its validity in prediction. Almazrouee and Almeshal et al. conduct long-term forecasting of electrical loads in Kuwait using Prophet and Holt-Winters model, the Prophet model shows more accurate predictions than the Holt-Winters (Almazrouee, Almeshal, Almutairi, Alenezi, & Alhajeri, 2020). Moreover, excellent variants of Prophet model have emerged, like NeuralPophet and LazyProphet models (Löning & Király, 2020; Triebe et al., 2021). ChikkaKrishna and Rachakonda et al. used Prophet and NeuralProphet model for short-term traffic prediction, the performance of NeuralProphet model outperforms than Prophet model (ChikkaKrishna, Rachakonda, & Tallam, 2022). These statistical models mentioned above have great application in the prediction domain and can be taken as good reference models.

Machine learning models play an active role in regression tasks, the most popular machine learning models in recent years involve support vector regression (SVR) and tree integration models like quantile regression forest (QRF), Gradient Boosting Regression (GBR), and Lightgbm etc.. Cao and Wu applied support vector regression model with fruit fly optimization algorithm for handling nonlinear time series prediction, the results verified the proposed approach is a viable option for electricity forecasting (Cao & Wu, 2016). However, support vector machine model performs well on small sample data sets and is not suitable for regression on large sample sets (Q. Liu et al., 2021). Quantile regression forest (QRF) is firstly proposed by Meinshausen and Ridgeway in 2006 (Meinshausen & Ridgeway, 2006). It is a combination of quantile regression (QR) and random forest (RF) algorithms, which retains the advantages of both. The QRF is developed for surface water quality prediction by AI-Sulttani et al., the results showed that the QRF attained the superior performance (Al-Sulttani et al., 2021). Lightgbm is a variation of GBR which incorporates optimization techniques such as histogram algorithm and leaf-wise to deliver a significant improvement in accuracy and speed (Ke et al., 2017). The superior performance of the Lightgbm has been reflected in numerous studies (Cui et al., 2021; Ju et al., 2019).

Ensemble models have been recently attracting a lot of interest in the prediction areas, which integrates two or more models to cope with serial data (Afan et al., 2021; Alizadeh, Jafari Nodoushan, Kalarestaghi, & Chau, 2017; Wang et al., 2021). Generally, the serial data comprises a variety of characteristics in the real world, such as trends and periodic characteristics. Whereas, the typical single model is not capable sufficiently to process such intricate data. The ensemble model could benefit from multiple single models to mine the features embedded in the data, which is the primary cause to make it so appealing. Many scholars have contributed researches regarding ensemble models (Laib, Khadir, & Mihaylova, 2019; Z. Liu, Wang, Zhang, & Huang, 2019; Piotrowski, Kopyt, Baczyński, Robak, & Gulczyński, 2021; Qin & Li, 2020; Sun & Wang, 2018; D. Wang, Wang, Song, & Liu, 2018). The study (Khairalla, Ning, Al-Jallad, & El-Faroug, 2018) outlines and investigates the Stacking Multi-Learning Ensemble (SMLE) model for time series prediction and demonstrates that the ensemble model is an extremely encouraging methodology for complex time series forecasting. An optimized system namely jellyfish search-seasonal autoregressive integrated moving average-least squares support vector regression (JS-LSSVR) is preferred in the study (Chou & Truong, 2021) and the analytical results show that the proposed system can forecast energy consumption 1 week ahead efficiently. The research (Kao, Nawata, & Huang, 2020) used an integrated framework called ensemble empirical mode decomposition-autoregressive integrated moving average-genetic algorithm-support vector regression (EEMD-ARIMA-GA-SVR) to predict the primary energy consumption of an economy, the results demonstrate the feasibility and efficiency of the proposed framework.

In the aforementioned publications, a variety of intelligent ensemble models are put forward to address various practical problems, whereas, few studies have performed multi-step forecasting for large-scale univariate time series, and most of them require extra features as input, which would result in decreasing flexibility of the model. This paper proposed an ensemble model that subtly incorporates a superior statistical model NeuralProphet and machine learning model Lightgbm, and the high-quality components extracted from the NeuralProphet model will be further extracted by Lightgbm model. The results show that the proposed model achieves outstanding accuracy and strong robustness than other benchmark models in this paper.

3. Methodologies

A varies of methodologies have been presented for electric power forecasting in recent years, each with its own characteristics and applicable to various dataset. It is challenging for general models to accurately capture the development patterns embedded in large and highly frequented data. The related prediction methodologies are roughly divided into three categories, respectively are deep learning models, statistical models and ensemble models. Deep learning models has excellent prediction performance but always with poorly interpretability and very high calculate cost. Statistical models have been well developed recently, such as the widely used Prophet model and its variants structures NeuralProphet and LazyProphet models. Lightgbm model is the validated outperforming machine learning model, with superior performance in various competitions. These mentioned models will be considered for experiment testing in this paper.

3.1. Prophet algorithm

Prophet model is a popular time processing algorithm which is created by Facebook in 2017 (Taylor & Letham, 2018). It is not sensitive for missing values and outliers in the time series and decomposes the time series into components such as trend, seasonal, holiday and residual terms, the mathematical expressions are shown as follows:

$$y(t) = g(t) + s(t) + h(t) + \varepsilon(t) \tag{1}$$

Where the $y(t)$ is the timeseries, $g(t)$ is the trend term, $s(t)$ is the seasonal term, $h(t)$ is the holiday term and $\varepsilon(t)$ is the residual term.

The equation of $g(t)$ can be expressed as below:

$$g(t) = \frac{C}{1 + e^{-k(t-b)}} \tag{2}$$

Where C is the upper limit of the trend, k and b are the growth rate and offset respectively, and t is the time.

The equation of $s(t)$ can be expressed as below:

$$s(t) = \sum_{n=1}^{N}\left[a_n \cos\left(\frac{2\pi nt}{T}\right) + b_n \sin\left(\frac{2\pi nt}{T}\right)\right] \tag{3}$$

Where T is the period and n is half the number of periods.

3.2. NeuralProphet algorithm

NeuralProphet is an optimized forecasting framework that improves on Prophet model. It holds the same model components as Prophet model and blends other optimal techniques, for example, it employs the Gradient Descent technique of PyTorch, which makes its training process faster than the Prophet model, In addition, it used AR-Net (Auto-Regressive Feed-Forward Neural Network) for model autoregression, and it also comprises lagged regressors and future regressors. Triebe et al. validate the effectiveness of NeuralProphet and illustrate that the model can produce more qualitative interpretable predictive components (Triebe et al., 2021).

3.3. LazyProphet algorithm

LazyProphet is another powerful time series prediction method (Löning & Király, 2020), it works by connecting the first point in the time series to another point midway, and then connecting the midway point to the last point, repeating this process several times, and simultaneously changes the position of the “kink”(intermediate node). Moreover, the ‘decay’ factor is introduced to penalize the slope of each line from the midpoint to the final point. Hence, this model is called ‘LazyProphet’ due to its less demanding nature. In this paper, the predictive performance of LazyProphet is only second to the proposed model in this paper. It is a good reference model on high frequency and large-scale data.

3.4. Lightgbmt algorithm

The Lightgbm model (Ke et al., 2017) is a remarkable machine learning algorithm, which is optimized on Gradient Boosting Regression (GBR) model (Friedman, 2001). The gradient boosting regression (GBR) is such one of boosting algorithm which were employed in this study. Like most boosting algorithm, gradient boosting regression was integrated by many base learners and the precise results could be obtained by iteratively training. Allow base learner more focus on the wrong training samples in each iteration is the common point in boost algorithm. In GBR algorithm, negative gradient was used to represent the errors of previous round, and the errors was minimized by increasing base learner. The training process of GBR is as follows:

Assumed that the training set is $\{xi,yi\}^{n}{}_{i=1}$ and the number of iterations is M.

1. Initialize the model with a constant value:

$$F_0(x) = \underset{\gamma}{\operatorname{argmin}} \sum_{i=1}^{n} L(y_i, \gamma) \tag{4}$$

Table 1
The statistic summaries of the sample data.

Abbreviation	Definition	Values
Count	The number of samples	26,304
Mean	The average of sample data	1.089
Std	Standard deviation	0.917
Min	Minimum value	0.124
Max	Maximum value	6.561

Table 2
The grid search range of parameters for the benchmark models in prediction experiments.

Models	Parameters	Grid search range for parameters
NeuralProphet-QRF	quantile	[0.1, 0.2, 0.3, 0.4,0.5, 0.6, 0.7, 0.8, 0.9]
	nthreads	[1, 2, 3, 4, 5, 6, 7, 8]
	n_estimators	[200, 400, 600, 800, 1000]
	min_samples_leaf	[2, 4, 6, 8, 10, 12, 14, 16, 18, 20]
	min_samples_split	[2, 4, 6, 8, 10, 12, 14, 16, 18, 20]
Prophet-Lightgbm	max_depth	[2, 3, 4, 5, 6, 7, 8, 9, 10]
	learning_rate	[0.01, 0.05, 0.1,0.15, 0.2, 0.25, 0.3]
	n_estimators	[100, 200, 300, 400, 500, 600, 700, 800,900,1000,1500,2000]
	bagging_fraction	[0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9]
	feature_fraction	[0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9]
	learning_rate	[0.01, 0.05, 0.1,0.15, 0.2, 0.25, 0.3]
LazyProphet	min_child_samples	[1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11,12, 13, 14, 15, 16, 17, 18, 19, 20]
	num_leaves	[10, 20, 30, 40, 50, 60, 70, 80, 90, 100]
	num_iterations	[100, 200, 300, 400, 500, 600, 700, 800, 900, 1000]
LSTM	No. of nodes in the first LSTM layer	[50, 100, 150]
	No. of nodes in the second LSTM layer	[50, 100, 150]
	No. of nodes in the third LSTM layer	[50, 100, 150]
NeuralProphet-Lightgbm (proposed)	max_depth	[2, 3, 4, 5, 6, 7, 8, 9, 10]
	learning_rate	[0.01, 0.05, 0.1,0.15, 0.2, 0.25, 0.3]
	n_estimators	[100, 200, 300, 400, 500, 600, 700, 800,900,1000,1500,2000]
	bagging_fraction	[0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9]
	feature_fraction	[0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9]

Where the l yi is the expression of loss function in model, specifically $\frac{1}{2}(y - F_m(x))^2$ the F_m represents the $\mathrm{m_{th}}$ iteration training model and the $F_m(x)$ is the prediction. The F_{m+1} was obtained by increasing the base learner h to $F_m(x)$ where exist function $F_{m+1}(x) = F_m(x) + h(x) = y$, We trained the model iteratively with the aim at minimizing the loss function expressed as the equation above.

2. The iterations m is from 1 to M:

(1) Compute the pseudo-residuals:

$$r_{im} = -\left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)}\right]_{F(x)=F_{m-1}(x)} \tag{5}$$

$for \quad i = 1, \cdots\cdots n$

Where the r_{im} is the negative gradient which also named pseudo-residuals denotes the deviation between the prediction and measure data.

(2) Fit a base learner $h_m(x)$ to pseudo-residuals, train it using the training set $\{(x_i, r_{im})\}_{i=1}^{n}$.

(3) Compute multiplier γ_m by solving the following one-dimensional optimization problem:

$$\gamma_m = \underset{\gamma}{argmin} \sum_{i=1}^{n} L(y_i, F_{m-1}(x_i) + \gamma h_m(x_i)) \tag{6}$$

(4) Update the model:

$$F_m(x) = F_{m-1}(x) + \gamma_m h_m(x) \tag{7}$$

3. Output Fm(x)

The Lightgbm model introduces four improvements to the traditional Gradient Boosting Regression, respectively, Gradient-based one-side sampling (GOSS), Exclusive Feature Bundling (EFB), Histogram algorithm and the Leaf-wise strategy (Ke et al., 2017). The GOSS algorithm sorts the gradient size of the training samples, discarding the samples with small gradient at random, while retaining the samples with large gradients to reduce memory usage. The EFB algorithm achieves dimension reduction by bundling mutually exclusive features of different dimensions, which greatly improve the training speed of the model. The basic idea of the Histogram algorithm is to discretize the continuous eigenvalues into k integers and construct a histogram with the width of k. The discrete values are used as an index to iterate through samples to find the optimal segmentation point, speeding up the training of the model by losing little precision. The Lightgbm uses the Leaf-wise decision tree growth strategy with depth restriction, which finds the leaf with the largest split gain each time and then splits it. The Leaf-wise causes lower computational cost than the traditional level-wise strategy in tree models, and it prevents overfitting by setting a maximum depth limit. The Lightgbm model could deliver a high calculate efficiency while ensuring a high level of accuracy (Ke et al., 2017).

4. Model establishment

4.1. Data pre-processing

The electric power data is obtained from the UCI.

public dataset with a one-minute sampling rate over a period of almost 4 years (Hebrail & Berard, 2012). Due to the large time span and highly intensive frequency of acquisition, missing values exist in the data set. It is crucial to guarantee the quality and completeness of the experimental data. In this paper, we firstly interpolated the missing values of the data by averaging the adjacent before and after values, and then resampled the data at an hourly frequency. The dataset spans the period from December 2006 to November 2010 and the complete three-year data from 2007 to 2009 are selected as experimental data, with a total of 26,304 data, and the statistic summaries of the data is shown in Table 1. As can be seen from the table, the experimental data fluctuates ranging from 6.561 to 0.124, and the mean of the data set is 1.089.

In this paper, the approach of Min-Max normalization is utilized to scale the data to the interval [0,1], so as to minimize the error caused by the data itself. The mathematical formula is presented as follows:

$$x_t' = \frac{x_t - x_{min}}{x_{max} - x_{min}} \tag{8}$$

Where x_t denotes the raw input value at time t and x_t' is the normalized value. x_{min} and x_{max} represent the minimum and maximum value in the sample data.

4.2. Construction of the proposed model

Fig. 1 is the flow chart of the proposed model, the overall structure is mainly composed of NeuralProphet model and Lightgbm model, the connected way of them has learned the blending method but are not identical. In the first layer of the structure, the NeuralProphet will be utilized as base model for predicting the univariate electric power data.

Table 3
The optimal parameter set for the benchmark models in different prediction experiments.

models	parameters	Prediction steps 12	24	48
NeuralProphet-QRF	quantile	0.1	0.1	0.1
	nthreads	4	7	4
	n_estimators	200	200	1000
	min_samples_leaf	20	20	4
	min_samples_split	20	20	4
Prophet-Lightgbm	objective	regression	regression	regression
	random_state	42	42	42
	max_depth	6	7	4
	learning_rate	0.01	0.15	0.25
	n_estimators	600	600	400
	bagging_fraction	0.2	0.2	0.7
	feature_fraction	0.2	0.2	0.7
LazyProphet	objective	regression	regression	regression
	metric	rmse	rmse	rmse
	verbosity	−1	−1	−1
	booting_type	gbdt	gbdt	gbdt
	seed	42	42	42
	linear_tree	False	False	False
	learning_rate	0.15	0.15	0.15
	min_child_samples	7	14	8
	num_leaves	30	20	20
	num_iterations	900	900	500
LSTM	No. of nodes in the first LSTM layer	100	150	100
	No. of nodes in the second LSTM layer	100	150	100
	No. of nodes in the third LSTM layer	100	150	100
	batch size	64	64	64
NeuralProphet-Lightgbm (proposed)	objective	regression	regression	regression
	random_state	42	42	42
	max_depth	2	3	3
	learning_rate	0.05	0.1	0.1
	n_estimators	200	100	200
	bagging_fraction	0.1	0.9	0.1
	feature_fraction	0.6	0.9	0.5

The NeuralProphet model has an inherent advantage for handling the large-scale and high-frequency univariate data (Khurana et al., 2022), thus the input of the NeuralProphet model only requires a pre-processed univariate time series data which without any additional features. As the NeuralProphet model introduced more optimized elements on the basis of Prophet model, it produces higher quality model components than Prophet model (Triebe et al., 2021). In this paper, the NeuralProphet model adopts the default parameter settings and the acquired model components are specifically trend, season_yearly, season_weekly, season_daily, residual1 and yhat1, where yhat1 is the predicted value of the NeuralProphet model. The higher-level feature terms are produced by using NeuralProphet model, and each feature term has its distinct developmental characteristics. It is notable that both the feature terms and predicted values of the NeuralProphet model are connected as feature set of the second layer, rather than the single feature terms or predicted values of the base model, which maximum the information extraction from the raw data.

The Lightgbm model is an excellent machine learning model which delivers a high accuracy (Ke et al., 2017). In this paper, the powerful Lightgbm model is used to perform further information extraction from the high-lever features yield in the first layer. In order to fully exploit the predictive potential of the Lightgbm model, the grid search method is used to test the key parameters of the model, and the optimal parameter set is determined as Table 3. It can be seen from Fig. 1, the complexity and interpretability of the proposed model are more acceptable, and the accuracy of the model is significantly superior to the benchmark models in this paper. In addition, the proposed model also maintains strong robustness in prediction tests with different prediction steps of 12, 24 and 48, which are described in Table 8.

4.3. Hyperparameter tuning

The hyperparameter tuning process is conducted to mine the potential of the model so as to improve its performance. The optimal parameter sets allow the benchmark models achieve the best predictive

Table 4
The mathematical expressions of four evaluation metrics.

Evaluation metrics	Mathematical expression
RMSE	$\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}$
MSE	$\frac{1}{n}\sum_{i=1}^{n}\lvert y_i-\hat{y}_i\rvert^2$
MAPE	$\frac{100\%}{n}\sum_{i=1}^{n}\left\lvert\frac{\hat{y}_i-y_i}{y_i}\right\rvert$
SD	$\sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}}$

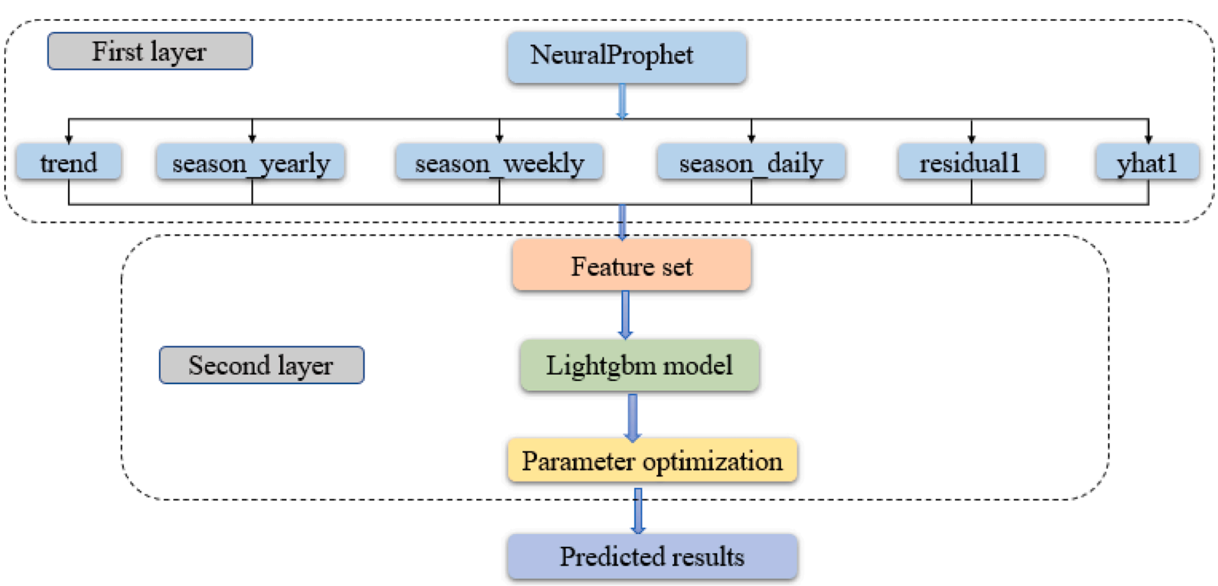

Fig. 1. The flow chart of the NeuralProphet-Lightgbm model.

Table 5
The means and standard deviations of prediction errors for different models in experiments with different prediction steps.

Models	Metrics (Mean ± SD)		
Prediction step: 12			
	RMSE	MSE	MAPE
NeuralProphet-QRF	0.8662 ± 0.3062	0.8347 ± 0.6711	63.4548 ± 6.4494
Prophet	0.5615 ± 0.2776	0.3846 ± 0.451	46.1604 ± 8.1442
ARIMA	0.5115 ± 0.2148	0.3031 ± 0.3144	41.0602 ± 12.0951
Prophet-Lightgbm	0.5661 ± 0.3192	0.4122 ± 0.4936	40.3853 ± 18.4421
NeuralProphet	0.4934 0.1928	0.2759 ± 0.2782	41.4419 ± 16.363
LSTM	0.5368 ± 0.2018	0.3248 ± 0.2089	39.0896 ± 21.7805
LazyProphet	0.4015 ± 0.2001	0.1973 ± 0.2192	28.6728 ± 13.5159
NeuralProphet-Lightgbm (proposed)	**0.2563 ± 0.1367**	**0.0825 ± 0.1099**	**21.162 ± 6.523**
Prediction step: 24			
Models	Metrics (Mean ± SD)		
	RMSE	MSE	MAPE
NeuralProphet-QRF	1.0635 ± 0.3813	1.2620 ± 0.9238	65.1274 ± 3.7745
Prophet	0.6325 ± 0.2054	0.4380 ± 0.3057	56.6731 ± 13.4434
ARIMA	0.7269 ± 0.3114	0.6156 ± 0.5084	51.9794 ± 22.9317
Prophet-Lightgbm	0.6932 ± 0.1593	0.5034 ± 0.2231	51.1251 ± 15.4299
NeuralProphet	0.6759 ± 0.302	0.5389 ± 0.4998	47.6625 ± 14.5379
LSTM	0.6339 ± 0.2013	0.4383 ± 0.2765	37.5743 ± 14.2588
LazyProphet	0.5228 ± 0.1834	0.3036 ± 0.2073	32.2936 ± 8.9902
NeuralProphet-Lightgbm (proposed)	**0.4162 ± 0.2366**	**0.2236 ± 0.2536**	**25.0750 ± 4.8044**
Prediction step: 48			
Models	Metrics (Mean ± SD)		
	RMSE	MSE	MAPE
NeuralProphet-QRF	1.0692 ± 0.2667	1.2073 ± 0.6388	66.9180 ± 3.9146
Prophet	0.6577 ± 0.1566	0.4546 ± 0.2338	66.3899 ± 13.0664
ARIMA	0.7528 ± 0.2116	0.6070 ± 0.3567	56.0199 ± 13.5602
Prophet-Lightgbm	0.7017 ± 0.1039	0.5022 ± 0.1538	57.0228 ± 11.208
NeuralProphet	0.6951 ± 0.2101	0.5229 ± 0.3546	53.8824 ± 13.4809
LSTM	0.7114 ± 0.171	0.5324 ± 0.2676	50.0182 ± 15.2777
LazyProphet	0.5837 ± 0.1628	0.3646 ± 0.1967	40.9436 ± 11.6973
NeuralProphet-Lightgbm (proposed)	**0.4081 ± 0.1773**	**0.1948 ± 0.1808**	**25.4181 ± 5.5484**

performance in the search range and help to avoid differences between the performance of the models resulting from the poor parameters. In this paper, the optimization procedure is implemented by the Grid Search method. The Grid Search method is widely used in the process of hyperparameter optimization and the search range for the parameters can be customized. Additionally, the method is flexible and easy to be implemented, and provides a maximum possibility to find the global optimal set of parameters. In this paper, the parameters of the benchmark models are tuned, and the search scope is specified in Table 2, where the key parameters of the LazyProphet model mainly includes learning_rate, min_child samples, num_leaves and num_iterations, the learning rate impacts the training rate of the model, too large value will result in the model skipping the best results, and too small value may

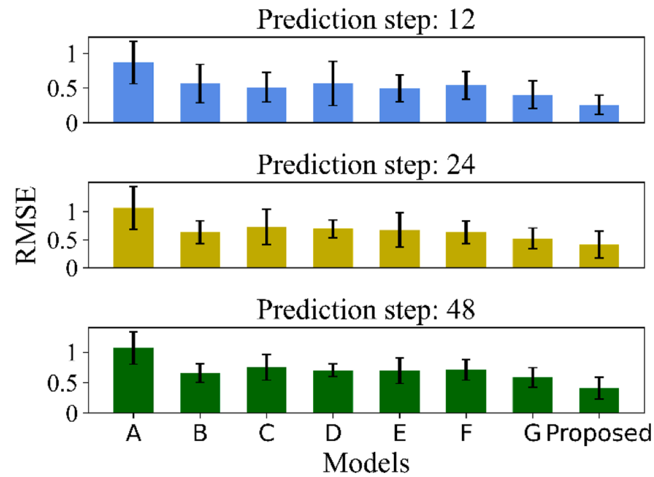

Fig. 2. The bar chart for mean and standard deviation of prediction model's RMSE in experiments with different prediction steps.

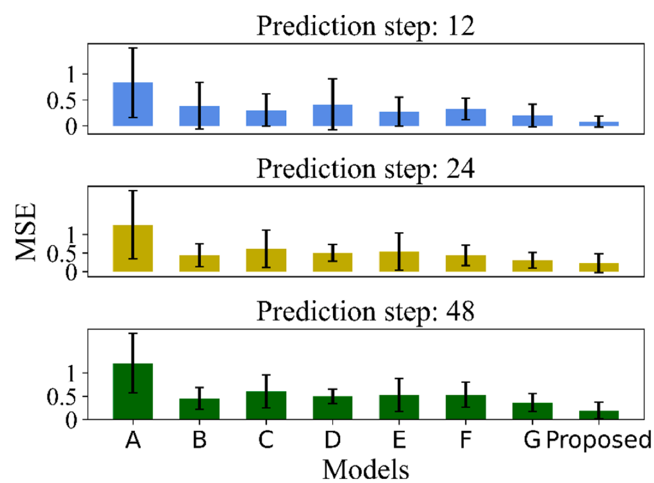

Fig. 3. The bar chart for mean and standard deviation of prediction model's MSE in experiments with different prediction steps.

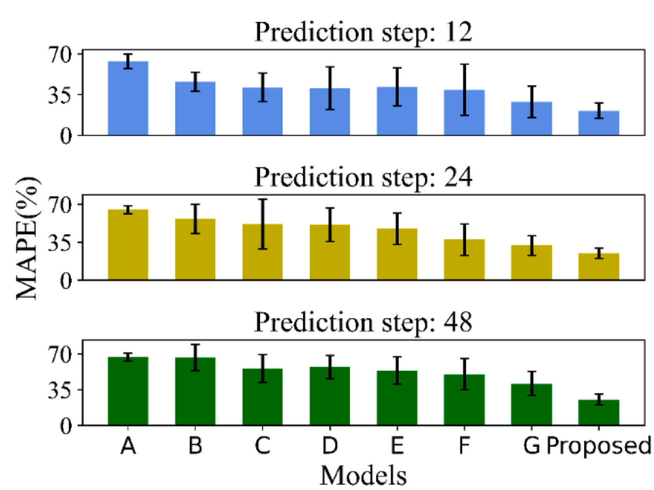

Fig. 4. The bar chart for mean and standard deviation of prediction model's MAPE in experiments with different prediction steps.

Table 6
The statistical results of error term MAPE for different prediction models on 10 sets of test data.

	NeuralProphet-QRF	Prophet	ARIMA	Prophet-Lightgbm	NeuralProphet	LSTM	LazyProphet	proposed
Prediction step: 12								
Median	63.7040	48.2804	38.5996	39.5468	41.5211	32.4984	24.6358	**22.4303**
Min	55.9396	33.3453	22.8277	12.9816	20.6022	15.7451	15.6032	8.5613
Max	71.7655	58.0157	60.9995	65.4260	80.0339	90.0721	59.1412	28.6907
Range	15.8259	24.6704	38.1718	52.4444	59.4317	74.327	43.538	**20.1294**
Prediction step: 24								
Median	64.1959	55.0697	46.5956	52.9883	45.5851	32.1177	31.897	**24.6238**
Min	59.3877	34.0028	32.0385	18.3304	31.895	26.1839	19.8399	19.405
Max	71.8699	78.222	97.5061	76.3729	81.3756	64.9725	48.8344	32.4056
Range	12.4822	44.2192	65.4676	58.0425	49.4807	38.7886	28.9945	**13.0007**
Prediction step: 48								
Median	66.5262	65.4402	54.4802	54.9699	50.9531	51.1844	39.8409	**25.0084**
Min	61.9927	43.7074	38.5394	42.1450	39.3003	29.1549	25.3061	15.4860
Max	76.4118	91.3836	78.6757	73.7785	81.2906	73.5755	66.5082	34.1310
Range	14.4191	47.6762	40.1363	31.6335	41.9903	44.4206	41.2021	**18.645**

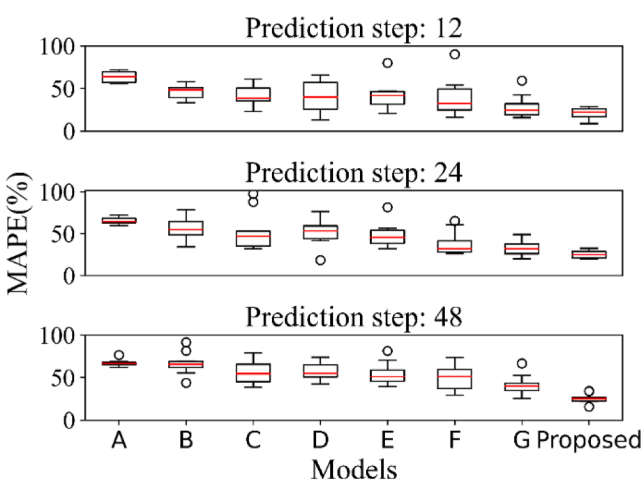

Fig. 5. The box plot of different model's MAPE in experiments with different prediction steps.

cause over-fitting of the model, the other three parameters are designed to set the model tree size to avoid overfitting and underfitting. LSTM is a typical neural network model for time series prediction, in which the number of the hidden layers and the neural nodes in each layer have a great influence on the performance of the model. In this paper, three hidden layers are designed, and the search grid for the neural nodes in each hidden layer is [50, 100, 150]. The number of neural nodes in each hidden layer is determined to 100 at prediction steps 12 and 48, and the number of neural nodes in each hidden layer is set to 150 when the prediction step is 24, the batch size is set to 64. Tree model is a kind of regression model with excellent performance, Quantile Regression Forest (QRF) and Lightgbm are involved in this paper. The key parameters and search grid for the QRF and Lightgbm models are detailed in Table 2, other parameters of the model use default values. Moreover, the parameters of the ARIMA model, NeuralProphet model and the Prophet model also applied with the default parameters. The optimal parameter sets for the benchmark models in different prediction horizon experiments yielded by the Grid Search method are displayed in Table 3.

5. Discussion and results

5.1. Evaluation indicator

The root mean square error (RMSE), the mean squared error (MSE) and the mean absolute percentage error (MAPE) are the mostly commonly evaluated indicators in regression analysis and widely adopted in many regression studies to assess the results produced by the prediction model. The smaller value of the RMSE, MSE and MAPE implies the smaller gap between the predicted and actual values. The standard deviation (SD) can reflect the dispersion of sample data. A smaller standard deviation indicates more concentrated data while a larger standard deviation suggests the data appear more discrete. In this paper, the standard deviation is taken to assess the dispersion of the test set prediction error, where smaller standard deviations signify a more stable prediction performance of the model. The mathematical expressions of RMSE, MSE, MAPE and SD are listed in Table 4.

Where y_i and $\hat{y}_i$ respectively denotes the observed and predicted value, $\overline{y}_i$ is the average of the observed values, x_i is the value of sample data, $\overline{x}$ is the mean of sample data, and n is the number of the testing sample data.

5.2. Experiment analysis

The univariate multi-step prediction in time series is a challenging task in forecasting field, and the prediction error of the model grows large as the prediction steps increases. The multi-step technique is usually employed in the power industry for electric power load forecasting. In this paper, the short-term electric power load forecasts are carried out for a 12-hour, 24-hour and 48-hour period respectively, which have a significant impact on the safe operation of the power system. The proposed NeuralProphet-Lightgbm model is dedicated to univariate multi-step prediction. In order to verify the effectiveness of the proposed model, the experiments are undertaken with forecasting windows of 12-hour, 24-hour and 48-hour, each prediction step was performed on 10 stets of test data and the results was the average of the 10 sets of test data. Totally-seven models respectively NeuralProphet-QRF, Prophet model, ARIMA model, Prophet-Lightgbm model, NeuralProphet model, LSTM model, and LazyProphet model are involved in this paper for model comparison. Among the benchmark models, the single model ARIMA is a typical time series model, Prophet model, NeuralProphet model and LazyProphet model are excellent time series in recent years, LSTM is a representative deep neural network model with good prediction effect for time series, and the combined models NeuralProphet-QRF and Prophet-Lightgbm incorporate excellent individual models. These benchmark models are more representative and can effectively validate the predictive performance of the proposed model. The prediction performance of different models with regard to the mean and standard deviation of the RMSE, MSE and MAPE are exhibited in Table 5, where the predicted results of our proposed method are highlighted in bold. As can be seen from the table, the proposed model presented better performance at prediction horizon of 12-hour, 24-hour and 48-hour, with RMSE, MSE and MAPE of 0.2563, 0.0825 and 21.162, 0.4162, 0.2236 and 25.0750, 0.4081, 0.1948 and 25.4181 respectively. The MAPE of the NeuralProphet-QRF model, NeuralProphet model and NeuralProphet-Lightgbm model at prediction

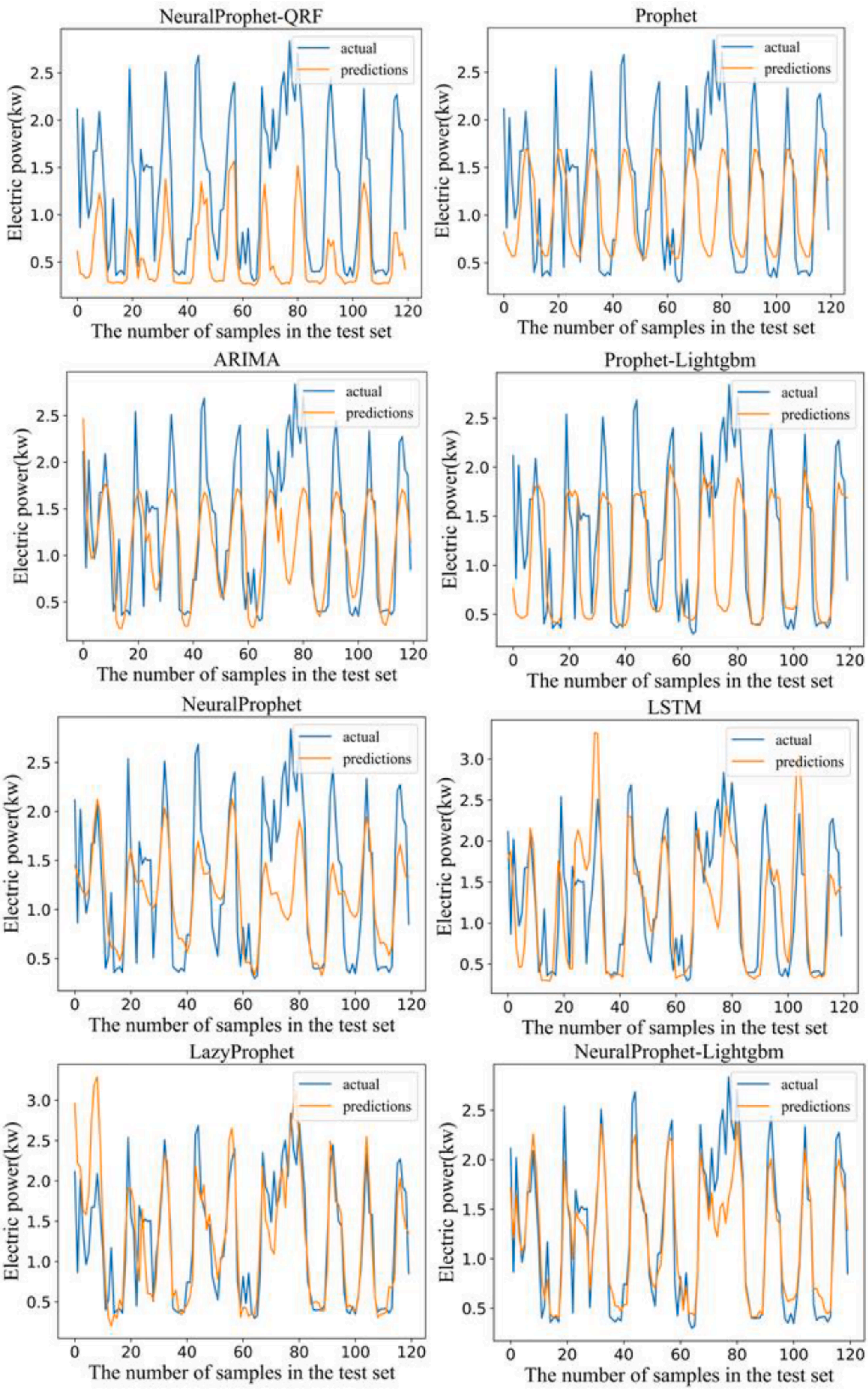

Fig. 6. The fitting plots of predicted and actual values for different models at prediction step 12.

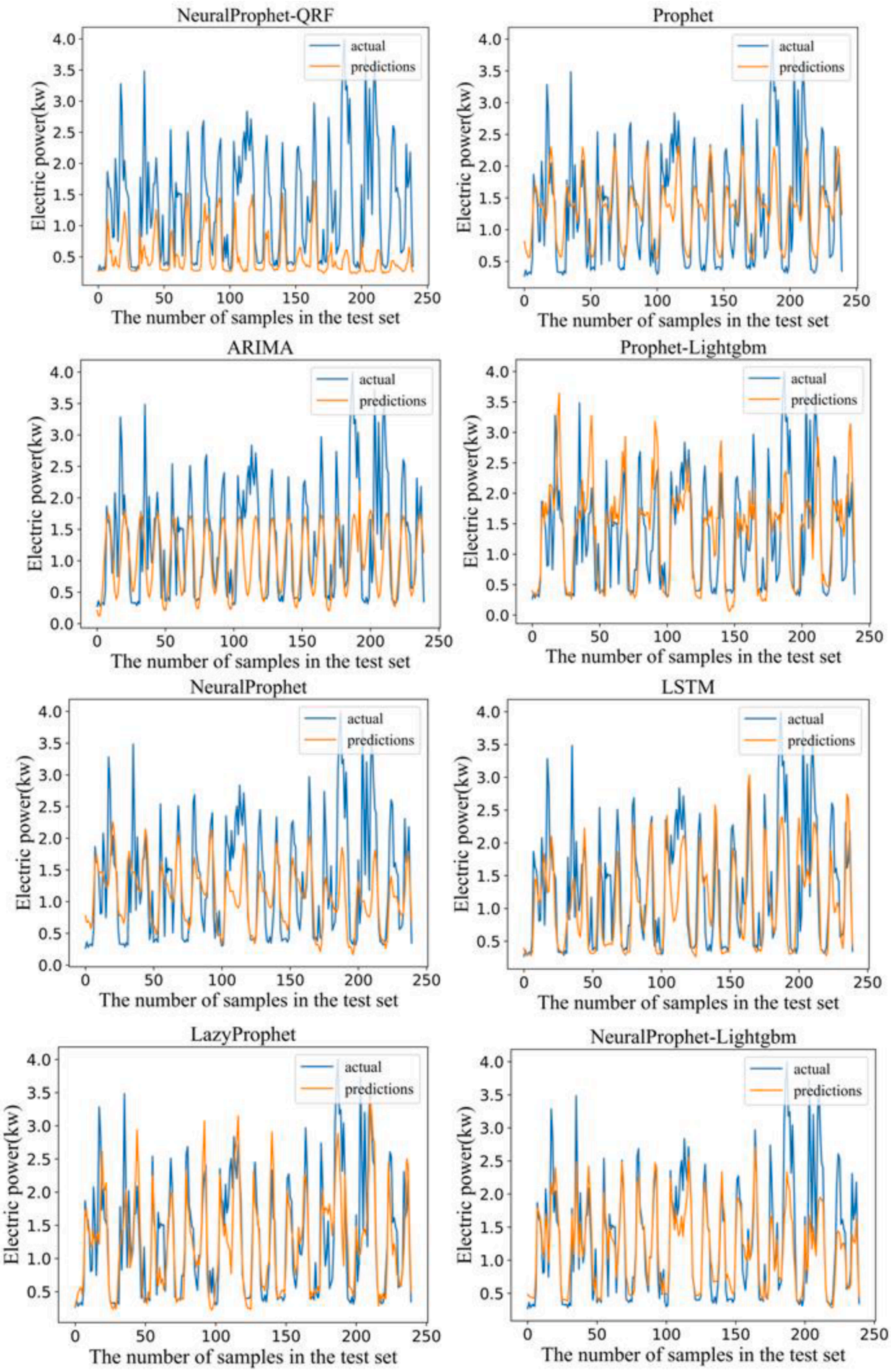

Fig. 7. The fitting plots of predicted and actual values for different models at prediction step 24.

horizon of 12-hour, 24-hour and 48-hour respectively are 63.4548, 41.4419 and 21.162, 65.1274, 47.6625 and 25.0750, 66.9180, 53.8824 and 25.4181, the proposed model NeuralProphet-Lightgbm model outperforms the single model NeuralProphet model, whereas the ensemble model NeuralProphet-QRF model is inferior to the single NeuralProphet model, the performances of the three models are ranked as NeuralProphet-QRF < NeuralProphet < NeuralProphet-Lightgbm it suggests that the Lightgbm model is superior to the QRF model. It also can be seen from the Table 5 that the NeuralProphet-Lightgbm model performs better than the Prophet-Lightgbm model, indicating that the

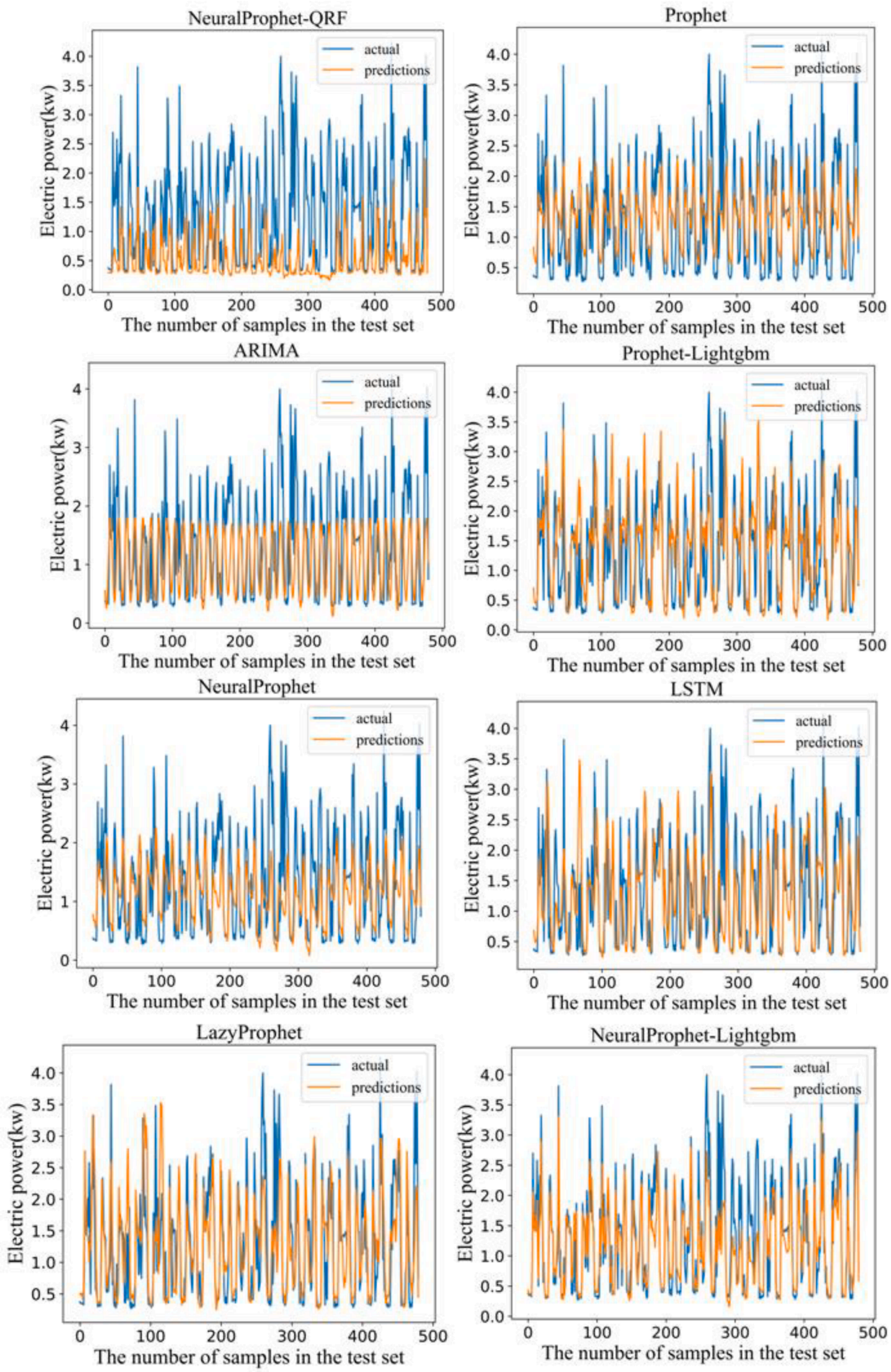

Fig. 8. The fitting plots of predicted and actual values for different models at prediction step 48.

NeuralProphet model outperforms the Prophet model, which is consistent with the experimental results. Moreover, the results reveal that both the ensemble models Prophet-Lightgbm and NeuralProphet-Lightgbm deliver a degree of improvement over the single Prophet and NeuralProphet model. The proposed model surpasses the other seven comparative models with the lowest prediction error and small standard deviation, regardless of the forecast window of 12-hour, 24-hour or 48-hour. The second-best performer is the LazyProphet model, its

Table 7
The comparison results of MAPE performance between the proposed model and other models.

Models	MAPE (%)		
	Prediction steps		
	12	24	48
Proposed/NeuralProphet-QRF	−66.65	−61.5	−62.02
Proposed/prophet	−54.16	−55.76	−61.71
Proposed/ARIMA	−48.46	−51.76	−54.63
Proposed/ Prophet-Lightgbm	−47.6	−50.95	−55.42
Proposed/ NeuralProphet	−48.94	−47.39	−52.83
Proposed/ LSTM	−45.86	−33.27	−49.18
Proposed/ LazyProphet	−26.19	−22.35	−37.92

Table 8
The MAPE of models in experiments with different prediction steps.

	MAPE (%)		
	Prediction steps		
	12	24	48
NeuralProphet-QRF	63.4548	65.1274	66.9180
Prophet	46.1604	56.6731	66.3899
ARIMA	41.0602	51.9794	56.0199
Prophet-Lightgbm	40.3853	51.1251	57.0228
NeuralProphet	41.4419	47.6625	53.8824
LSTM	39.0896	37.5743	50.0182
LazyProphet	28.6728	32.2936	40.9436
proposed	**21.162**	**25.0750**	**25.4181**

performance surpasses that of the LSTM model. In addition, as the number of prediction steps increases, the error terms of the prediction models have all increased to some extent, while the performance of the proposed model remains steadily.

The visualization of Table 5 is plotted in Fig. 2, Fig. 3 and Fig. 4, which depicts the error bars of different models in terms of RMSE, MSE and MAPE at different prediction horizons. The horizontal axes labels of the error bars A, B, C, D, E, F, G, Proposed respectively signify the NeuralProphet-QRF, Prophet, ARIMA, Prophet-Lightgbm, NeuralProphet, LSTM, LazyProphet and NeuralProphet-Lightgbm. The black lines in the bar plots indicate the magnitude of the standard deviation of the prediction error, and a shorter line reflects the model's performance is more stable and reliable. From the figure, it can be intuitively seen that the proposed model has the lowest mean error and a smaller volatility range, implying that the proposed model is pretty outstanding and robust. The second-best results are achieved by the LazyProphet model, and the worst performance is obtained by NeuralProphet-QRF model.

Table 6 provides the statistical results of error term MAPE for different prediction models on test data. Totally-four statistical indicators respectively Median, Min, Max and the Range of the MAPE are summarized in the table, which reflect the stability of the model's predictive performance on the test set. It can be seen that the proposed model has the lowest median and the second lowest Range in comparison to the other models, which reveals that the proposed model is extremely robust and shows an absolute advantage over other models, and the superiority of the proposed model becomes more obvious as the number of prediction steps increases. Fig. 5 presents a box plot of the model's MAPE on test data, where the red line marks the position of the median value, and the horizontal labels A, B, C, D, E, F, G and Proposed refer to the models of NeuralProphet-QRF, Prophet, ARIMA, Prophet-Lightgbm, NeuralProphet, LSTM, LazyProphet and NeuralProphet-

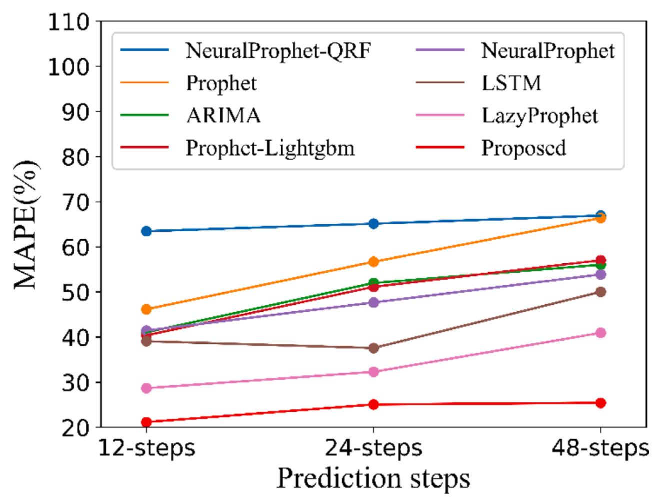

Fig. 10. The fold line plot of the models' MAPE in the experiments with different prediction steps.

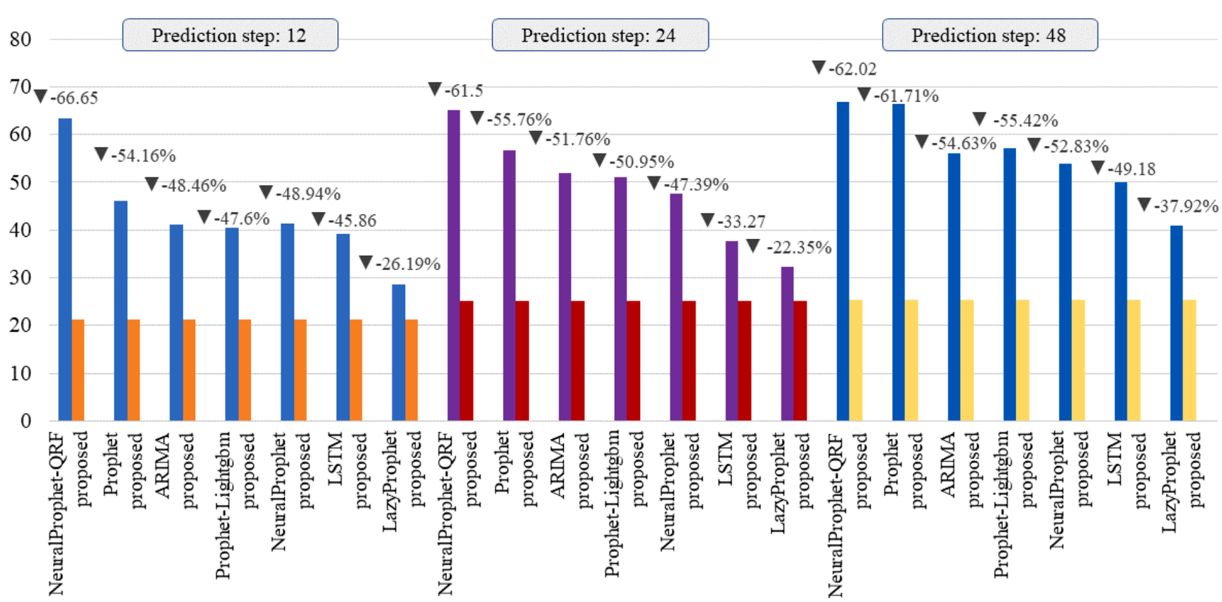

Fig. 9. The bar chart of the error term MAPE between the proposed model and other models.

Lightgbm respectively. The height of the box in the graph indicates the fluctuations range of MAPE. It is apparent from the graph that the NeuralProphet-QRF model and the proposed model both have very small box heights, suggesting that the MAPE values of these two models are highly concentrated, but the median of the NeuralProphet-QRF is the highest, while the median of the proposed model is the lowest, meaning that the proposed model offers the highest forecasting accuracy and is very stable. Also, the fluctuation range of MAPE for the proposed model respectively are 20.1294, 13.0007 and 18.645 at the prediction horizon of 12-hour, 24-hour and 48-hour, the second-best model LazyProphet are 43.538, 28.9945 and 41.2021 respectively, where the differences of MAPE fluctuation range between the proposed model and second-best model is more evident. As a result, the model proposed in this paper has a significant advantage over other involved models in terms of prediction accuracy and stability. Fig. 6, Fig. 7 and Fig. 8 depict the fitting plots of predicted and actual values for different models at prediction steps 12, 24 and 48 respectively. As can be seen from the graphs, the time series fluctuates more frequently and exhibits stronger irregularities. The proposed model has strong capability of prediction and performs well in most wave peaks and valleys in curves, while the other models have the insufficient capability for prediction and the deviation between actual and predicted values is relatively large, making it difficult to meet prediction demand. The difficulty of model prediction increases as the predicted steps increase, the proposed model performs extremely stable while the predictive errors of other models tend to be larger.

5.3. Experiment results

Table 7 gives the comparison results of MAPE performance between the proposed model and other models. As can be seen from the table, the proposed model yields a great superiority over the other models, and the larger the prediction horizon, the more obvious the advantage of the proposed model. The predictive performance of NeuralProphet-QRF model was the poorest, and the performance of LazyProphet model was only inferior to the proposed model, but as the prediction horizon increasing, the gap of model's performance between the LazyProphet and the proposed model became larger. Fig. 9 is the bar chart of error term MAPE between the proposed model and other models. it can be seen from the figure, as the prediction horizon increases, the model's MAPE incrementally grows, while the proposed model is more stable with less obvious increase in MAPE values.

To further validate the stability of the proposed model, the model's MAPE at different prediction steps are analyzed. Table 8 presents the model's MAPE in the experiments with different prediction steps. Fig. 10 displays a fold line plot regarding Table 8, the vertical coordinates of the graph indicate the MAPE values. It can be seen that the MAPE fold line of the NeuralProphet-QRF model is in top position, and the MAPE fold line of the ARIMA and Prophet-Lightgbm model are almost overlap, which implies the predictive performance of these two models is comparable. The purple fold line is the MAPE of the LazyProphet model, and fold line towards upward as the predicted steps increasing. The red fold line is the MAPE of the proposed NeuralProphet-Lightgbm model, which goes up very flatly as the prediction horizon increases, the line between the predicted steps 24 and 48 is nearly a level line and it is always located at the lowest position in the graph. Fig. 10 reveals that the propose model outperforms other comparative models in term of the prediction precision and stability, and it has good value for application in time series forecasting.

6. Conclusion

The hourly frequency time series of electric power load is featured with high volatility and randomness, therefore the multi-step prediction for univariate electric power time series is always a huge challenge. The experiment results show that the prediction performance of the proposed model is much superior to the comparative models, which with excellent accuracy and stability. What's more, the proposed model also has satisfactory performance in multi-step forecasting, which assists decision-makers in planning electricity management and helps to reduce energy waste effectively.

The limitation of this study is in the parameter tuning segment, the grid search method is trained on each combination of arguments in the parameter list until the optimal parameter set is found, which is time-consuming. Coupling with a superior and faster automatic parameter optimization algorithm is an important direction to improve the model performance. Incorporating multivariate factors into the model is also a focus of future research.

CRediT authorship contribution statement

Yubo Zhao: Conceptualization, Data curation, Formal analysis. **Ni Guo:** Conceptualization, Methodology, Software, Writing – original draft. **Wei Chen:** Conceptualization, Methodology, Validation, Supervision, Formal analysis. **Hailan Zhang:** Data curation, Visualization. **Bochao Guo:** Data curation, Investigation. **Jia Shen:** Data curation, Visualization. **Zijian Tian:** Writing – review & editing, Supervision, Formal analysis.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

The present study is supported by the National Natural Science Foundation of China under Grant 52074305, in part by the National Natural Science Foundation of China under Grant 51874300 and 51874299, in part by the National Natural Science Foundation of China and Shanxi Provincial People's Government Jointly Funded Project of China for Coal Base and Low Carbon under Grant U1510115 and in part by the Open Research Fund of Key Laboratory of Wireless Sensor Network and Communication, Shanghai Institute of Microsystem and Information Technology, Chinese Academy of Sciences, under Grant 20190902.

References

Abdel-Aal, R. (2008). Univariate modeling and forecasting of monthly energy demand time series using abductive and neural networks. *Computers & Industrial Engineering, 54*(4), 903–917. https://doi.org/10.1016/j.cie.2007.10.020

Afan, H. A., Ibrahem Ahmed Osman, A., Essam, Y., Ahmed, A. N., Huang, Y. F., Kisi, O., . . . El-Shafie, A. (2021). Modeling the fluctuations of groundwater level by employing ensemble deep learning techniques. *Engineering Applications of Computational Fluid Mechanics, 15*(1), 1420-1439. 10.1080/19942060.2021.1974093.

Ahmad, N., Ghadi, Y., Adnan, M., & Ali, M. (2022). Load forecasting techniques for power system: Research challenges and survey. *IEEE Access*. https://doi.org/10.1109/ACCESS.2022.3187839

Al-Sulttani, A. O., Al-Mukhtar, M., Roomi, A. B., Farooque, A. A., Khedher, K. M., & Yaseen, Z. M. (2021). Proposition of new ensemble data-intelligence models for surface water quality prediction. *IEEE Access, 9*, 108527–108541. https://doi.org/10.1109/ACCESS.2021.3100490

Alizadeh, M. J., Jafari Nodoushan, E., Kalarestaghi, N., & Chau, K. W. (2017). Toward multi-day-ahead forecasting of suspended sediment concentration using ensemble models. *Environmental Science and Pollution Research, 24*(36), 28017–28025. https://doi.org/10.1007/s11356-017-0405-4

Almazrouee, A. I., Almeshal, A. M., Almutairi, A. S., Alenezi, M. R., & Alhajeri, S. N. (2020). Long-term forecasting of electrical loads in kuwait using prophet and holt–winters models. *Applied Sciences, 10*(16), 5627. https://doi.org/10.3390/app10165627

Bedi, G., Venayagamoorthy, G. K., Singh, R., Brooks, R. R., & Wang, K.-C. (2018). Review of internet of things (iot) in electric power and energy systems. *IEEE Internet of Things Journal, 5*(2), 847–870. https://doi.org/10.1109/JIOT.2018.2802704

Cao, G., & Wu, L. (2016). Support vector regression with fruit fly optimization algorithm for seasonal electricity consumption forecasting. *Energy, 115*, 734–745. https://doi.org/10.1016/j.energy.2016.09.065

Che, J., & Zhai, H. (2022). Wt-arima combination modelling for short-term load forecasting. *IAENG International Journal of Computer Science, 49*(2).

ChikkaKrishna, N. K., Rachakonda, P., & Tallam, T. (2022). Short-term traffic prediction using fb-prophet and neural-prophet. *Paper presented at the 2022 IEEE Delhi Section Conference (DELCON)*.

Chou, J. S., & Truong, D. N. (2021). Multistep energy consumption forecasting by metaheuristic optimization of time-series analysis and machine learning. *International Journal of Energy Research, 45*(3), 4581–4612. https://doi.org/10.1002/er.6125

Cui, Z., Qing, X., Chai, H., Yang, S., Zhu, Y., & Wang, F. (2021). Real-time rainfall-runoff prediction using light gradient boosting machine coupled with singular spectrum analysis. *Journal of Hydrology, 603*, Article 127124. https://doi.org/10.1016/j.jhydrol.2021.127124

Eldali, F. A., Hansen, T. M., Suryanarayanan, S., & Chong, E. K. (2016). Employing arima models to improve wind power forecasts: A case study in ercot. *Paper presented at the 2016 North American Power Symposium (NAPS)*.

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of statistics*, 1189–1232. https://www.jstor.org/stable/2699986.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation, 9*(8), 1735–1780. https://doi.org/10.1007/978-3-642-24797-2_4

Hebrail, G., Berard, A. (2012). Individual household electric power consumption. https://archive.ics.uci.edu/ml/datasets/Individual+household+electric +power+consumption.php.

Jiang, W., Wu, X., Gong, Y., Yu, W., & Zhong, X. (2020). Holt–winters smoothing enhanced by fruit fly optimization algorithm to forecast monthly electricity consumption. *Energy, 193*, Article 116779. https://doi.org/10.1016/j.energy.2019.116779

Ju, Y., Sun, G., Chen, Q., Zhang, M., Zhu, H., & Rehman, M. U. (2019). A model combining convolutional neural network and lightgbm algorithm for ultra-short-term wind power forecasting. *IEEE Access, 7*, 28309–28318. https://doi.org/10.1109/ACCESS.2019.2901920

Kao, Y.-S., Nawata, K., & Huang, C.-Y. (2020). Predicting primary energy consumption using hybrid arima and ga-svr based on eemd decomposition. *Mathematics, 8*(10), 1722. https://doi.org/10.3390/math8101722

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W Liu, T.-Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems, 30*.

Khairalla, M. A., Ning, X., Al-Jallad, N. T., & El-Faroug, M. O. (2018). Short-term forecasting for energy consumption through stacking heterogeneous ensemble learning model. *Energies, 11*(6), 1605. 10.3390/en11061605.

Khurana, S., Sharma, G., Miglani, N., Singh, A., Alharbi, A., Alosaimi, W., … Goyal, N. (2022). *An intelligent fine-tuned forecasting technique for covid-19 prediction using neuralprophet model* (pp. 629–649). Materials and Continua: Computers.

Kim, D., Kwon, D., Park, L., Kim, J., & Cho, S. (2020). Multiscale lstm-based deep learning for very-short-term photovoltaic power generation forecasting in smart city energy management. *IEEE Systems Journal, 15*(1), 346–354. https://doi.org/10.1109/JSYST.2020.3007184

Kudo, M., Takeuchi, A., Nozaki, Y., Endo, H., & Sumita, J. (2009). Forecasting electric power generation in a photovoltaic power system for an energy network. *Electrical Engineering in Japan, 167*(4), 16–23. https://doi.org/10.1002/eej.20755

Laib, O., Khadir, M. T., & Mihaylova, L. (2019). Toward efficient energy systems based on natural gas consumption prediction with lstm recurrent neural networks. *Energy, 177*, 530–542. https://doi.org/10.1016/j.energy.2019.04.075

Liu, Q., Liu, W., Mei, J., Si, G., Xia, T., & Quan, J. (2021). A new support vector regression model for equipment health diagnosis with small sample data missing and its application. *Shock and Vibration, 2021*. https://doi.org/10.1155/2021/6675078

Liu, Y., Ding, S., & Jia, W. (2020). A novel prediction method of complex univariate time series based on k-means clustering. *Soft Computing, 24*(21), 16425–16437. https://doi.org/10.1007/s00500-020-04952-2

Liu, Z., Wang, X., Zhang, Q., & Huang, C. (2019). Empirical mode decomposition based hybrid ensemble model for electrical energy consumption forecasting of the cement grinding process. *Measurement, 138*, 314–324. https://doi.org/10.1016/j.measurement.2019.02.062

Löning, M., & Király, F. (2020). Forecasting with sktime: Designing sktime's new forecasting api and applying it to replicate and extend the m4 study. *arXiv preprint arXiv:2005.08067*. 10.48550/arXiv.2005.08067.

Lopez-Martin, M., Sanchez-Esguevillas, A., Hernandez-Callejo, L., Arribas, J. I., & Carro, B. (2021). Novel data-driven models applied to short-term electric load forecasting. *Applied Sciences, 11*(12), 5708. https://doi.org/10.3390/app11125708

Ma, Y., Wang, J., Song, J., Zheng, Z., Huang, L., & Zhang, J. (2020). Holt-winters predicting model of energy meter based on optimal smoothing coefficient. *Paper presented at the 2020 IEEE 3rd Student Conference on Electrical Machines and Systems (SCEMS)*.

Meinshausen, N., & Ridgeway, G. (2006). Quantile regression forests. *Journal of machine learning research, 7*(6).

Piotrowski, P., Kopyt, M., Baczyński, D., Robak, S., & Gulczyński, T. (2021). Hybrid and ensemble methods of two days ahead forecasts of electric energy production in a small wind turbine. *Energies, 14*(5), 1225. https://doi.org/10.3390/en14051225

Qin, L., & Li, W. (2020). A combination approach based on seasonal adjustment method and echo state network for energy consumption forecasting in USA. *Energy Efficiency, 13*(7), 1505–1524. https://doi.org/10.1007/s12053-020-09897-x

Shamshirband, S., Jafari Nodoushan, E., Adolf, J. E., Abdul Manaf, A., Mosavi, A., & Chau, K.-W. (2019). Ensemble models with uncertainty analysis for multi-day ahead forecasting of chlorophyll a concentration in coastal waters. *Engineering Applications of Computational Fluid Mechanics, 13*(1), 91–101. https://doi.org/10.1080/19942060.2018.1553742

Shohan, M. J. A., Faruque, M. O., & Foo, S. Y. (2022). Forecasting of electric load using a hybrid lstm-neural prophet model. *Energies, 15*(6), 2158. https://doi.org/10.3390/en15062158

Sun, W., & Wang, Y. (2018). Short-term wind speed forecasting based on fast ensemble empirical mode decomposition, phase space reconstruction, sample entropy and improved back-propagation neural network. *Energy Conversion and Management, 157*. https://doi.org/10.1016/j.enconman.2017.11.067

Taylor, S. J., & Letham, B. (2018). Forecasting at scale. *The American Statistician, 72*(1), 37–45. https://doi.org/10.1080/00031305.2017.1380080

Triebe, O., Hewamalage, H., Pilyugina, P., Laptev, N., Bergmeir, C., & Rajagopal, R. (2021). Neuralprophet: Explainable forecasting at scale. *arXiv preprint arXiv:2111.15397*. 10.48550/arXiv.2111.15397.

Wang, D., Wang, Y., Song, X., & Liu, Y. (2018). Coal overcapacity in china: Multiscale analysis and prediction. *Energy Economics, 70*. https://doi.org/10.1016/j.eneco.2018.01.004

Wang, R., Peng, C., Gao, J., Gao, Z., & Jiang, H. (2020). A dilated convolution network-based lstm model for multi-step prediction of chaotic time series. *Computational and Applied Mathematics, 39*(1), 1–22. https://doi.org/10.1007/s40314-019-1006-2

Wang, W.-C., Du, Y.-J., Chau, K.-W., Xu, D.-M., Liu, C.-J., & Ma, Q. (2021). An ensemble hybrid forecasting model for annual runoff based on sample entropy, secondary decomposition, and long short-term memory neural network. *Water Resources Management, 35*(14), 4695–4726. https://doi.org/10.1007/s11269-021-02920-5

Xia, C., Zhang, M., & Cao, J. (2018). A hybrid application of soft computing methods with wavelet svm and neural network to electric power load forecasting. *Journal of Electrical Systems and Information Technology, 5*(3), 681–696. https://doi.org/10.1016/j.jesit.2017.05.008

Yan, G., Yu, C., & Bai, Y. (2021). Wind turbine bearing temperature forecasting using a new data-driven ensemble approach. *Machines, 9*(11), 248. https://doi.org/10.3390/machines9110248

Yenidoğan, I., Çayir, A., Kozan, O., Dağ, T., & Arslan, Ç. (2018). Bitcoin forecasting using arima and prophet. *Paper presented at the 2018 3rd international conference on computer science and engineering (UBMK)*.

Zhong, Y., Chen, L., Dan, C., & Rezaeipanah, A. (2022). A systematic survey of data mining and big data analysis in internet of things. *The Journal of Supercomputing, 1–49*. https://doi.org/10.1007/s11227-022-04594-1

Zou, Z., Wu, X., Zhao, Z., Wang, Q., & Zhou, M. (2018). Prediction of short term electric load based on bp neural networks & arima combination. *Paper presented at the 2018 IEEE 4th Information Technology and Mechatronics Engineering Conference (ITOEC)*.