An integrated federated learning algorithm for short-term load forecasting

Yang, Yang, Wang, Zijin, Zhao, Shangrui and Wu, Jinran

Yang, Y., Wang, Z., Zhao, S. and Wu, J. (2023). An integrated federated learning algorithm for short-term load forecasting. *Electric Power Systems Research, 214*, Article 108830. https://doi.org/10.1016/j.epsr.2022.108830

An integrated federated learning algorithm for short-term load forecasting

Yang Yang[a], Zijin Wang[a], Shangrui Zhao[b,*], Jinran Wu[c,d]

[a] *College of Automation & College of Artificial Intelligence, Nanjing University of Posts and Telecommunications, Nanjing 210023, PR China*
[b] *School of Science, Wuhan University of Technology, Wuhan 430070, PR China*
[c] *School of Mathematical Sciences, Queensland University of Technology, Brisbane 4001, Australia*
[d] *The Institute for Learning Sciences and Teacher Education, Australian Catholic University, Brisbane 4000, Australia*

Abstract

Accurate power load forecasting plays an integral role in power systems. To achieve high prediction accuracy, models need to extract effective features from raw data, and the training of models needs a large amount of data. However, data sharing will require the disclosure of the private data of the participants. To address this issue, we combined variational mode decomposition (VMD), the federated k-means clustering algorithm (FK), and SecureBoost into a single algorithm, called *VMD-FK-SecureBoost*. First, we used VMD to decompose the original data into several sub-sequences. This enabled us to extract the implied features to separately predict each sub-sequence to improve the prediction accuracy. Second, we use FK to recombine the sub-sequences into several clusters with common characteristics. Finally, with SecureBoost, we use clustering results to realize federated learning with privacy protection. We calculated the prediction values by accumulating the prediction results of the sub-sequences. The results for the examples in the US and Australia showed that the prediction performance of VMD-FK-SecureBoost was better than those of XGBoost and SecureBoost. Particularly, the MAPEs of one-step-ahead forecasting in the Texas and Newcastle CBD from our proposed method are 0.209% and 2.127% respectively, which are the lowest of all the algorithms.

Keywords: Federated learning, Decomposition-ensemble method, Clustering, Load forecasting

1. Introduction

Accurate short-term load forecasting (STLF) is crucial to the security, stable operation, and economic dispatch of a power system [1]. However, power load data are complex due to the uncertain behavior of consumers [2]. Extracting features from complex data to achieve the required STLF accuracy is a challenge. In addition, as the data privacy of society improves the privacy and security of data should be considered in training models [3]. Thus, how to realize STLF through effective feature extraction with privacy protection has become a crucial research topic.

*Corresponding author
Email addresses: yyang@njupt.edu.cn (Yang Yang), 1321058617@njupt.edu.cn (Zijin Wang), zhaosr@whut.edu.cn (Shangrui Zhao), j73.wu@qut.edu.au (Jinran Wu)

1.1. Literature review

Recently, many researchers have designed different STLF [4, 5, 6] due to the immense practicality of STLF in smart grids. STLF methods can be generally divided into two categories: statistical methods [7, 8, 9] and machine learning (ML) methods [10, 11]. Among the statistical methods, Qin et al. [7] proposed a new hybrid model that combines the auto-regressive integrated moving average model and feedforward neural networks. Aly [12] proposed a hybrid optimized model that improves the accuracy when the load data are intermittent, non-linear, and fluctuating. Statistical methods are simple, but they highly require the support of a stable time series. Thus, it is difficult to maintain effective prediction performance for complex power load data with only statistical methods.

For complex data, researchers combine ML methods with decomposition methods to perform forecasting. Yang et al. [13] proposed an effective dimensionality reduction approach with variational mode decomposition (VMD) and a variational autoencoder (VAE) for complex power load data. Yang et al. [14] designed an iterative decompose–cluster–feedback algorithm for load forecasting. Empirical mode decomposition (EMD) is used to decompose the load sequence into subsequences with various amplitude–frequency characteristics, which avoids direct calculation of the load sequence. Yue et al. [15] proposed a mode decomposition–recombination method in which ensemble empirical mode decomposition (EEMD) is utilized and the accuracy of load forecasting is improved. Nguyen et al. [16] proposed a novel self-boosted mechanism for limited data accessibility, in which the original time series is decomposed into multiple time series as additional features. These methods show that combining ML methods with decomposition algorithms can yield more accurate results. Among the different decomposition algorithms, VMD can effectively avoid mode aliasing and shows good decomposition performance on non-stationary and non-linear signals.

On other aspect, traditional ML modeling tends to transmit and store power load data in a data center or a centralized system [17]. There are two main kinds of centralized frameworks. One framework trains the ML model separately for each set of historical load data from different departments. The other framework is aggregated load forecasting, which aggregates all data to train an ML model [18]. However, data transmission and centralized methods cause serious network congestion. Moreover, with the establishment of the General Data Protection Regulation (GDPR), the centralized methods are required to comply with the data regulations [19]. Therefore, centralized STLF has become more expensive computationally and even more impracticable [3]. These issues can be addressed by edge computing and decentralized learning methods, which make central data storage and processing unnecessary [20].

Federated learning (FL) was proposed as a decentralized ML method [21]. This method achieved a global model for sharing updated weights instead of training data using edge devices. These weights are securely processed by the central server, and then, returned to the edge devices [22]. FL performs better than the centralized methods, in terms of scalability and private data security [23]. Fekri et al. [17] trained the model without sharing local data through smart meters and FL strategies. In addition, the model that

adopted federated averaging achieved better prediction performance than a single centralized ML model. Ma et al. [24] designed a federated two-stage learning framework that augments prototypical federated learning. Furthermore, the federated k-means clustering algorithm (FK) is an FL-based unsupervised learning method, unlike the FL supervised learning method [25]. FK can cluster data from multiple participants with privacy protection. Kumar et al. [26] applied FK to a privacy-preserving real-world case for the first time and found that FK provides a possible solution for user data privacy. Wang et al. [27] used FK to extract the electricity consumption pattern while protecting the privacy of the data owners.

Tree boosting is a highly effective and widely used ML method that has shown impressive performance in many ML tasks. For example, XGBoost [28] is a scalable tree boosting system. SecureBoost is a novel gradient–tree boosting algorithm based on the FL framework [29]. Ma et al. [30] implemented an example based on SecureBoost. Liang et al. [31] proposed a vertical federated learning (VFL) model with Hetero Secure Boost Tree (HSBT) algorithm, which reduced by 50% the cost of a Chinese bank for data protection. Therefore, SecureBoost can provide effective privacy protection for big data analysis.

1.2. Motivation

According to the literature, STLF with multiple participants has two main challenges. First, effectively extracting features from data sets is challenging. Second, data privacy should be considered when modeling with data from multiple participants.

Considering these challenges, this paper proposes an integration algorithm based on the FL framework for STLF. First, VMD is used to extract features from data in a completely non-recursive way, thus avoiding the error caused by recursive algorithms. Second, FK can group data into several clusters with common characteristics while protecting data privacy [25]. Thus, FK is used to further extract features from the decomposed VMD results. Finally, SecureBoost provides a solution for the ML model training over multiple data sources with privacy [29]. Therefore, SecureBoost is used as the ML module for privacy protection.

1.3. Contributions

This paper proposes the integration algorithm *VMD-FK-SecureBoost* based on VMD, FK, and SecureBoost, for STLF. First, VMD decomposes the data set into several sub-sequences. Second, FK recombines the decomposed sub-sequences to generate new clusters with common characteristics. Finally, SecureBoost realizes the FL with multiple participants and protects data privacy. The experiment results showed that the proposed algorithm can extract features effectively and with data security. This paper makes the following contributions:

(a) This paper proposes an integration algorithm (VMD-FK-SecureBoost) that combines feature extraction and data privacy protection for STLF. VMD-FK-SecureBoost can further extract feature information while protecting data privacy.

(b) This paper uses the secure data interaction method of the FL framework to avoid data leakage.

(c) The effectiveness of VMD-FK-SecureBoost has been verified on two actual power load data sets (from the US and Australia). The comparison of VMD-FK-SecureBoost with XGBoost, VMD-XGBoost, SecureBoost, and VMD-SecureBoost showed that VMD-FK-SecureBoost has the best forecasting performance.

1.4. Structure of this paper

The remaining part of this paper is organized as follows. Section 2 presents some definitions of terms. Section 3 details the proposed integration algorithm in detail. Section 4 reports the experiment results for the integrated algorithm and the results of their verification on power load data from the US and Australia. Section 5 concludes this paper.

2. Definition of terms

2.1. Variational mode decomposition

Decomposition algorithms can effectively extract potential features from data. VMD is a nonrecursive decomposition algorithm [32]. By setting the number of mode decompositions according to the actual situation, VMD can obtain the optimal solution to the variational problem. Therefore, this study uses VMD to decompose the power load data for effective feature extraction. In this study, VMD constructs a variational problem through the power load data x_t. The step-by-step implementation of the VMD algorithm is described as follows.

Step 1: VMD constructed a variational problem through the input signal f. In this problem, f was decomposed into M intrinsic mode functions (IMFs) via Hilbert transform. Therefore, the corresponding constraint variational expression is:

$$\begin{aligned} &\min_{\{u_j\},\{\omega_j\}} \left\{ \sum_{j=1}^{M} \left\| \partial_t \left\{ [\nabla(t) + \frac{i}{t\pi}] * u_j(t) \right\} e^{-i\omega_j t} \right\|_2^2 \right\}, \\ &s.t. \sum_{j=1}^{M} u_j = f, \end{aligned} \tag{1}$$

where j is the number of current IMFs; i is an imaginary number; f is the signal to be decomposed; $\{u_j\}$ and $\{\omega_j\}$ correspond to the jth model component and the centre frequency, respectively; ∇ denotes the Dirac distribution; $*$ is a convolution operator; and M is the number of IMFs.

Step 2: By introducing the Lagrangian operator λ, the constraint problem was transformed into an unconstrained problem using the following equation:

$$\begin{aligned} L(\{u_j\},\{\omega_j\},\lambda) &= \alpha \sum_{j=1}^{M} \left\| \partial_t [(\nabla(t) + \frac{i}{t\pi})] * u_j(t) e^{-i\omega_j t} \right\|_2^2 \\ &+ \|f(t) - \sum_{j=1}^{M} u_j(t)\|_2^2 + \langle \lambda(t), f(t) - \sum_{j=1}^{M} u_j(t) \rangle, \end{aligned} \tag{2}$$

where α is the quadratic penalty factor for reducing the Gaussian noise interference.

Step 3: To optimize each modal component and the center frequency, Fourier isometric transform and the alternate direction method of multipliers (ADMM) were used. After each iteration, u_j, ω_j, and λ were optimized alternately with the saddle point of the augmented Lagrangian function. The equation for the specific optimization process is as follows:

$$\hat{u}_j^{n+1} = \frac{\hat{f}(\omega) - \sum_{j \neq k} \hat{u}_k^n(\omega) + \frac{\hat{\lambda}(\omega)}{2}}{1 + 2\alpha(\omega - \omega_j^n)^2}, \tag{3}$$

$$\hat{\omega}_j^{n+1} = \frac{\int_0^\infty \omega \left|\hat{u}_j^n(\omega)\right|^2 d\omega}{\int_0^\infty \left|\hat{u}_j^n(\omega)\right|^2 d\omega}, \tag{4}$$

and

$$\hat{\lambda}^{n+1}(\omega) = \hat{\lambda}^n(\omega) + \tau(\hat{f}(\omega) - \sum_j \hat{u}_j^{n+1}), \tag{5}$$

where n is the iteration number; $\hat{u}_j^{n+1}$, $\hat{\omega}_j$, and $\hat{f}(\omega)$ correspond to the Fourier transform of the mode u_j^n, the related center frequency, and the Fourier transform of the actual time series f; τ is the noise tolerance.

2.2. Federated k-means clustering algorithm

FK is a clustering method based on the FL framework that can cluster data with privacy protection [25]. Considering that clustering performs excellently with a large number of data sets [14], in this study, this method was combined with SecureBoost. The developed algorithm was trained according to the clusters with the same characteristics to improve its forecasting performance.

First, the k-means clustering algorithm established several cluster centers, μ_c $(c = 1, 2, \cdots, C)$, where C is the number of clusters. Second the distance between the time series x_t $(t = 1, 2, \cdots, T)$ and the cluster center was calculated using the following L, as follows:

$$L = \sum_{c=1}^{C} \sum_{j=1}^{\eta_c} \left\| x_j^c - \mu_c \right\|^2, \tag{6}$$

where η_c represents the number of clusters, and x_j^c is the time series that belongs to the cluster c.

Then, x_t $(t = 1, 2, \cdots, T)$ was divided into C clusters. Finally, the cluster centers μ_c' $(c = 1, 2, \cdots, C)$ were iteratively updated with the corresponding time series x_j $(j = 1, 2, \cdots, \eta_c)$ to minimize the loss function L. This process can be formulated as:

$$\mu_c' = \frac{1}{\eta_c} \sum_{j=1}^{\eta_c} x_j. \tag{7}$$

Fig. 1 shows the clustering progress of FK. In this clustering method, the arbiter securely aggregates the distance to the centroid from each participant instead of directly transmitting private data.

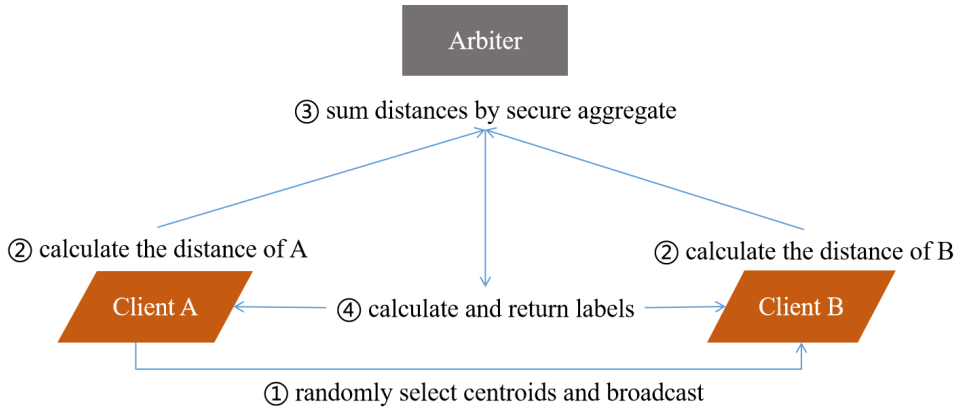

Figure 1: Framework of FK. (For clarity, Client A and Client B were chosen to represent the participants. Actually, however,there can be more than two participants).

2.3. SecureBoost

SecureBoost is a novel gradient–tree boosting algorithm based on the FL framework [29]. To ensure the confidentiality of training data, SecureBoost cooperatively learns a sharing gradient-tree boosting model through multi-party data under privacy constraints.

First, SecureBoost determines inter-database intersections with a privacy-preserving protocol [33]. Second, it trains a shared gradient-tree boosting model with the collaboration of multiple parties, without violating data privacy. The training progress can be divided into the following steps.

Step 1: The local models of SecureBoost download the current global model from the server.

Step 2: The local models update the current model based on their local data.

Step 3: The updated information is encrypted and sent back to the server.

Step 4: The server updates the global model with the aggregated updates.

These steps are shown in Fig. 2.

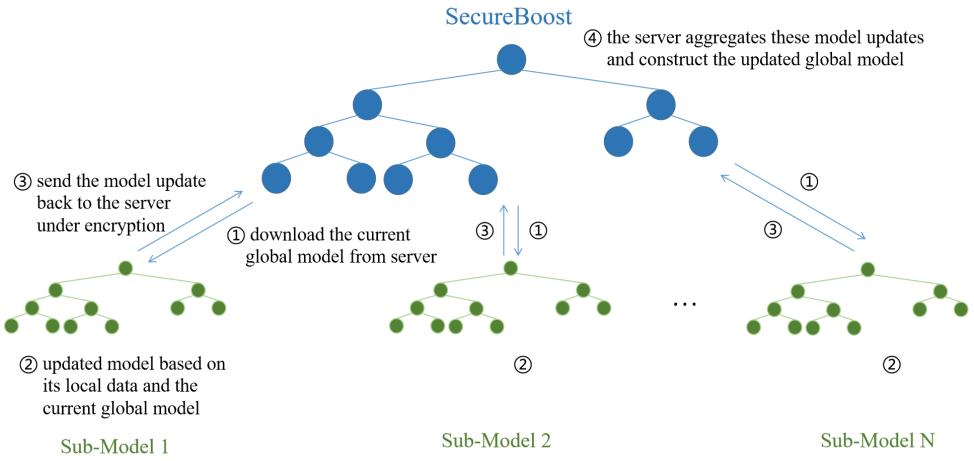

Figure 2: Framework of SecureBoost.

3. The proposed model

This study proposes VMD-FK-SecureBoost, which considers the data privacy of different participants and extracts more potential characteristics of power load data. This algorithm effectively improves the prediction accuracy of STLF. The steps of the proposed algorithm are as follows.

Step 1: Series decomposition. VMD extracts features from the power load data X_i $(i = 1, 2, 3, \cdots, T)$, where T is the data size, M is the manually set number of subsequences, and V_j $(j = 1, 2, 3, \cdots, M)$ are decomposed by VMD as described in Sec. 2.1, and M is set manually.

Step 2: Federated clustering. FK recombines the sub-sequences V_j $(j = 1, 2, 3, \cdots, M)$ from different participants into new clusters C_k $(k = 1, 2, 3, \cdots, C)$, where the number of clusters C is an artificial parameter, as mentioned in Sec. 2.2.

Step 3: Model training. From the clustering results in Step 2, the data sets are divided into training sets and testing sets. For the target of multi-step-ahead forecasting, these training sets and testing sets are split into different sample sets and label sets using the rolling window method. The forecasting results V_j' $(j = 1, 2, 3, \cdots, M)$ are obtained by using these data sets to train SecureBoost.

Step 4: Prediction results. The final prediction results X_i' $(i = 1, 2, 3, \cdots, T)$ are calculated according to the different participants by accumulating the predicted results V_j' $(j = 1, 2, 3, \cdots, M)$.

The steps of VMD-FK-SecureBoost are shown in Fig. 3. The specific implementation process of the proposed algorithm is provided in Algorithm 1.

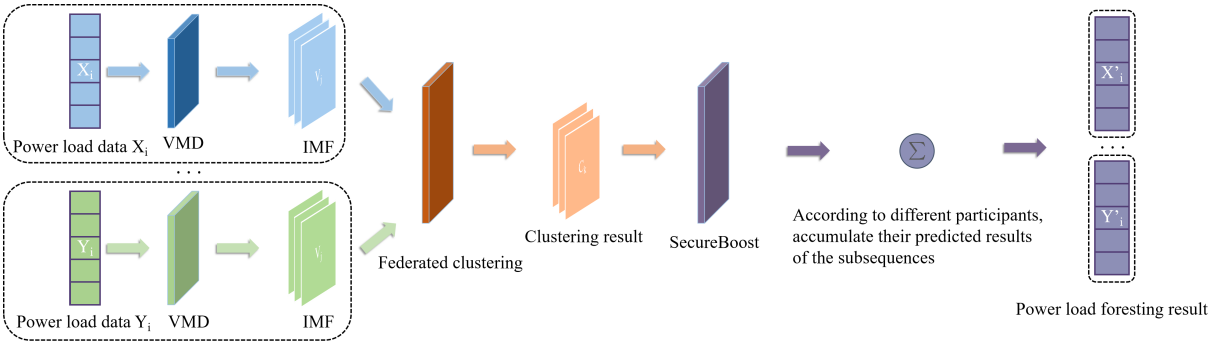

Figure 3: Flowchart of the proposed algorithm.

Remark 1. It is noteworthy that the manual parameters M and C in Step 1 and Step 2 would influence the feature extraction efficiency. Those parameters are tuned by the empirical method in this paper. Several state-of-the-art optimization algorithms can improve the further effect of the proposed algorithm, such as the Bayesian optimization algorithm [34].

4. Case study

This section comprehensively analyzes the STLF results by studying the following questions.

(1) Can the features extracted by VMD be effective for STLF?

Algorithm 1 VMD-FK-SecureBoost

Input: Power load data X_i $(i = 1, 2, 3, \cdots, T)$

Output: Final prediction value X_i' $(i = 1, 2, 3, \cdots, T)$

1: Set the value of the expected patterns M.
2: Obtain the decomposed sub-sequences V_j $(j = 1, 2, 3, \cdots, M)$.
3: Set C as the number of clusters C_k $(k = 1, 2, 3, \cdots, C)$.
4: Cluster V_j into cluster C_k for each j $(j = 1, 2, 3, \cdots, M)$.
5: Initialize the global model of SecureBoost.
6: **for all** $k = 1, 2, 3, \cdots, C$ **do**
7: Train and update local models;
8: Send updated weights to the central server; and
9: aggregate the weights and update the global model.
10: **end for**
11: Output the forecasting results V_j' $(j = 1, 2, 3, \cdots, M)$.
12: Sum V_j' to obtain the final prediction value X_i' $(i = 1, 2, 3, \cdots, T)$.

(2) How was the forecasting performance affected by the FK results?

(3) Can the FL with data privacy protection improve the accuracy of STLF in VMD-FK-SecureBoost?

The two data sets used in this study are from the US (Energy Information Administration: https://www.eia.gov/beta/) and Australia (Australia distribution zone substations: https://www.ausgrid.com.au/). The entire validation experiment is carried out on Matlab R2020a and PyCharm Community Edition 2022.1 x64 environment with Windows 10 and a 2.30 GHz Intel Core i5-8300H CPU, with 64-bit support and 16GB RAM.

In this study, a comparative experiment was also designed to further analyze the performance of VMD-FK-SecureBoost. Four regression evaluation metrics are introduced for quantitative analysis of the prediction results. The simulation effect and fitting degree of the different models are measured by the following indicators:

$$R^2 = 1 - \sum_{i=1}^{T}(Y_i - \hat{Y}_i)^2 / \sum_{i=1}^{T}(Y_i - \frac{1}{T}\sum_{i=1}^{T} Y_i),$$

$$\text{MAPE} = \frac{1}{T}\sum_{i=1}^{T} |\frac{100 \times (Y_i - \hat{Y}_i)}{Y_i}|,$$

$$\text{MAE} = \frac{1}{T}\sum_{i=1}^{T} |Y_i - \hat{Y}_i|,$$

and

$$\text{RMSE} = \sqrt{\frac{1}{T}\sum_{i=1}^{T}(Y_i - \hat{Y}_i)^2},$$

where T is the data size, and Y_i and $\hat{Y}_i$ are the ith observed value and the prediction, respectively.

The experiment results of the two examples are analyzed from three aspects: the advantages of the FL framework-based models over the centralized method in forecasting accuracy, the effectiveness of the decomposition algorithm for feature extraction, and the forecasting performance of the clustering algorithm in VMD-FK-SecureBoost.

4.1. Example 1: The data set of the power load in the US

Example 1 considered four US regions (Southeast, Tennessee, Texas, and Central) for STLF. The data were recorded every hour. Each region had $8,760$ data points (from 00:00 on January 1, 2021 to 23:00 on December 31, 2021). This data set was divided into the training set (80%) and the testing set (20%). There were $7,008$ data points (from 00:00 on January 1, 2021 to 23:00 on October 19, 2021) in the training set, and $1,752$ data points (from 00:00 on October 20, 2021 to 23:00 on December 31, 2021) in the testing set. Tables 1–3 show the error indicators for Example 1. Fig. 4 shows the curves for the one-, three-, and five-step-ahead forecasting, respectively. The experimental analysis of Example 1 was carried out as follows.

Table 1: One-step-ahead forecasting results of VMD-FK-SecureBoost and the contrast models in four US regions.

Region	Model	R^2	MAPE	MAE	RMSE
Southeast	XGBoost	0.910	2.545	639.202	867.060
	VMD-XGBoost	0.950	1.936	466.386	643.912
	SecureBoost	0.954	1.742	442.068	620.042
	VMD-SecureBoost	0.975	1.549	366.912	459.515
	VMD-FK-SecureBoost	**0.974**	**1.408**	**334.741**	**461.240**
Tennessee	XGBoost	0.914	2.688	472.991	671.903
	VMD-XGBoost	0.943	2.097	377.815	547.848
	SecureBoost	0.946	1.926	344.302	532.484
	VMD-SecureBoost	0.968	1.706	303.248	407.735
	VMD-FK-SecureBoost	**0.977**	**1.359**	**240.390**	**348.112**
Texas	XGBoost	0.863	3.503	1336.478	1741.291
	VMD-XGBoost	0.941	2.242	892.465	1144.090
	SecureBoost	0.924	2.961	1150.597	1298.418
	VMD-SecureBoost	0.967	1.945	752.177	860.296
	VMD-FK-SecureBoost	**0.997**	**0.209**	**83.788**	**104.145**
Central	XGBoost	0.907	1.872	530.278	696.572
	VMD-XGBoost	0.963	1.213	336.092	438.163
	SecureBoost	0.925	1.594	457.647	626.357
	VMD-SecureBoost	0.972	1.101	316.628	385.619
	VMD-FK-SecureBoost	**0.980**	**0.898**	**249.048**	**322.609**

NOTE. The power load data of the US were from 00:00 on October 20, 2021 to 23:00 on December 31, 2021.

4.1.1. Comparative experiment between XGBoost, VMD-XGBoost, SecureBoost, and VMD-SecureBoost

This subsection describes the design of a comparative experiment in the decentralized and centralized frameworks to verify the improvement of VMD-FK-SecureBoost. The number of the mode decompositions

Table 2: Three-step-ahead forecasting results of VMD-FK-SecureBoost and the contrast models in four US regions.

Region	Model	R^2	MAPE	MAE	RMSE
Southeast	XGBoost	0.772	4.102	1028.050	1376.608
	VMD-XGBoost	0.911	2.598	630.430	861.878
	SecureBoost	0.816	3.646	920.043	1236.686
	VMD-SecureBoost	0.965	1.787	425.825	536.319
	VMD-FK-SecureBoost	**0.969**	**1.543**	**368.752**	**505.936**
Tennessee	XGBoost	0.753	4.512	796.770	1141.531
	VMD-XGBoost	0.900	2.783	499.527	726.780
	SecureBoost	0.790	4.158	737.146	1052.222
	VMD-SecureBoost	0.957	2.012	357.951	475.330
	VMD-FK-SecureBoost	**0.974**	**1.473**	**256.792**	**373.187**
Texas	XGBoost	0.776	4.434	1742.970	2224.887
	VMD-XGBoost	0.936	2.261	931.869	1193.317
	SecureBoost	0.847	3.824	1487.240	1837.927
	VMD-SecureBoost	0.966	1.825	709.786	864.017
	VMD-FK-SecureBoost	**0.970**	**1.608**	**662.004**	**819.366**
Central	XGBoost	0.743	3.124	890.539	1159.265
	VMD-XGBoost	0.936	1.626	450.360	578.306
	SecureBoost	0.755	2.889	833.292	1133.404
	VMD-SecureBoost	0.971	1.067	305.552	386.687
	VMD-FK-SecureBoost	**0.975**	**1.018**	**281.688**	**360.609**

NOTE. The power load data of the US were from 00:00 on October 20, 2021 to 23:00 on December 31, 2021.

of VMD was determined using the empirical method. The experiment results in Tables 1–3 show that the models that used VMD had a better prediction effect than the other models without the decomposition algorithm. Compared with XGBoost and SecureBoost, VMD-XGBoost and VMD-SecureBoost showed improved fitting effects. For example, the MAPEs of VMD-XGBoost and VMD-SecureBoost were 1.936 and 1.549, respectively, lower than those of XGBoost and SecureBoost, which were 2.545 and 1.742, respectively, in the Southeast. These show that VMD can extract potential feature information from data and effectively improve prediction accuracy.

SecureBoost produces a more accurate forecasting result than XGBoost because SecureBoost can train a global model using the data from multiple participants. Compared with the other contrast models in Table 1, VMD-SecureBoost (R^2: 0.975, MAPE: 1.549, MAE: 366.912, and RMSE: 459.515) had a better fitting degree to the actual electric load demand than VMD-XGBoost (R^2: 0.950, MAPE: 1.936, MAE: 466.386, and RMSE: 643.912) in the one-step-ahead forecasting results for the Southeast region. Therefore, VMD-SecureBoost showed better forecasting performance than VMD-XGBoost.

4.1.2. Comparative experiment between VMD-SecureBoost and VMD-FK-SecureBoost

To verify the forecasting performance improvement of FK in VMD-FK-SecureBoost, a contrast model was designed in this experiment. In Fig. 4(b), the one-step-ahead forecasting curve of VMD-FK-

Table 3: Five-step-ahead forecasting results of VMD-FK-SecureBoost and the contrast models in four US regions.

Region	Model	R^2	MAPE	MAE	RMSE
Southeast	XGBoost	0.696	4.764	1189.634	1588.745
	VMD-XGBoost	0.889	2.909	708.667	961.614
	SecureBoost	0.731	4.476	1117.290	1493.966
	VMD-SecureBoost	0.957	1.962	467.598	594.430
	VMD-FK-SecureBoost	**0.965**	**1.673**	**398.979**	**542.335**
Tennessee	XGBoost	0.672	5.198	911.733	1314.607
	VMD-XGBoost	0.867	3.263	589.738	835.508
	SecureBoost	0.684	5.147	903.142	1290.422
	VMD-SecureBoost	0.946	2.332	413.231	531.260
	VMD-FK-SecureBoost	**0.944**	**2.465**	**435.988**	**544.872**
Texas	XGBoost	0.740	4.724	1859.436	2400.754
	VMD-XGBoost	0.928	2.403	991.066	1265.228
	SecureBoost	0.775	4.480	1740.534	2232.450
	VMD-SecureBoost	0.965	1.764	693.658	878.624
	VMD-FK-SecureBoost	**0.951**	**2.095**	**866.632**	**1040.279**
Central	XGBoost	0.705	3.393	962.871	1242.046
	VMD-XGBoost	0.921	1.823	504.947	644.666
	SecureBoost	0.698	3.280	940.372	1255.208
	VMD-SecureBoost	0.973	1.021	291.365	378.783
	VMD-FK-SecureBoost	**0.970**	**1.125**	**312.361**	**396.770**

NOTE. The power load data of the US were from 00:00 on October 20, 2021 to 23:00 on December 31, 2021.

SecureBoost is closer to the power data curve of the Tennessee region than VMD-SecureBoost. Therefore, the model that used the clustering algorithm had better prediction results.

The experiment results showed that VMD-FK-SecureBoost had a better prediction effect than the models without the clustering algorithm. Tables 1–3 show that VMD-FK-SecureBoost had a lower MAPE than VMD-SecureBoost in the one-, three-, and five-step-ahead prediction. For example, the results of VMD-FK-SecureBoost (R^2: 0.977, MAPE: 1.359, MAE: 240.390, and RMSE: 348.112) all show a more accurate forecasting performance than the results of VMD-SecureBoost (R^2: 0.968, MAPE: 1.706, MAE: 303.248, and RMSE: 407.735) in the one-step-ahead forecasting study in Texas. The results further show that compared with VMD-SecureBoost, SecureBoost with the decomposition and clustering algorithm had a better prediction effect.

4.1.3. Comparative experiment between XGBoost, VMD-XGBoost, and VMD-FK-SecureBoost

This sub-section describes the design of a group of comparative experiments to verify the effectiveness of data privacy protection for STLF. VMD-FK-SecureBoost securely aggregated data from different participants and cooperatively trained a global model. As shown in Table 1, the one-step-ahead forecasting result of VMD-FK-SecureBoost (MAPE: 0.209) is better than those of XGBoost (MAPE: 3.503) and VMD-XGBoost (MAPE: 2.242) in Texas. As can be seen in Fig. 4(c), VMD-FK-SecureBoost had a

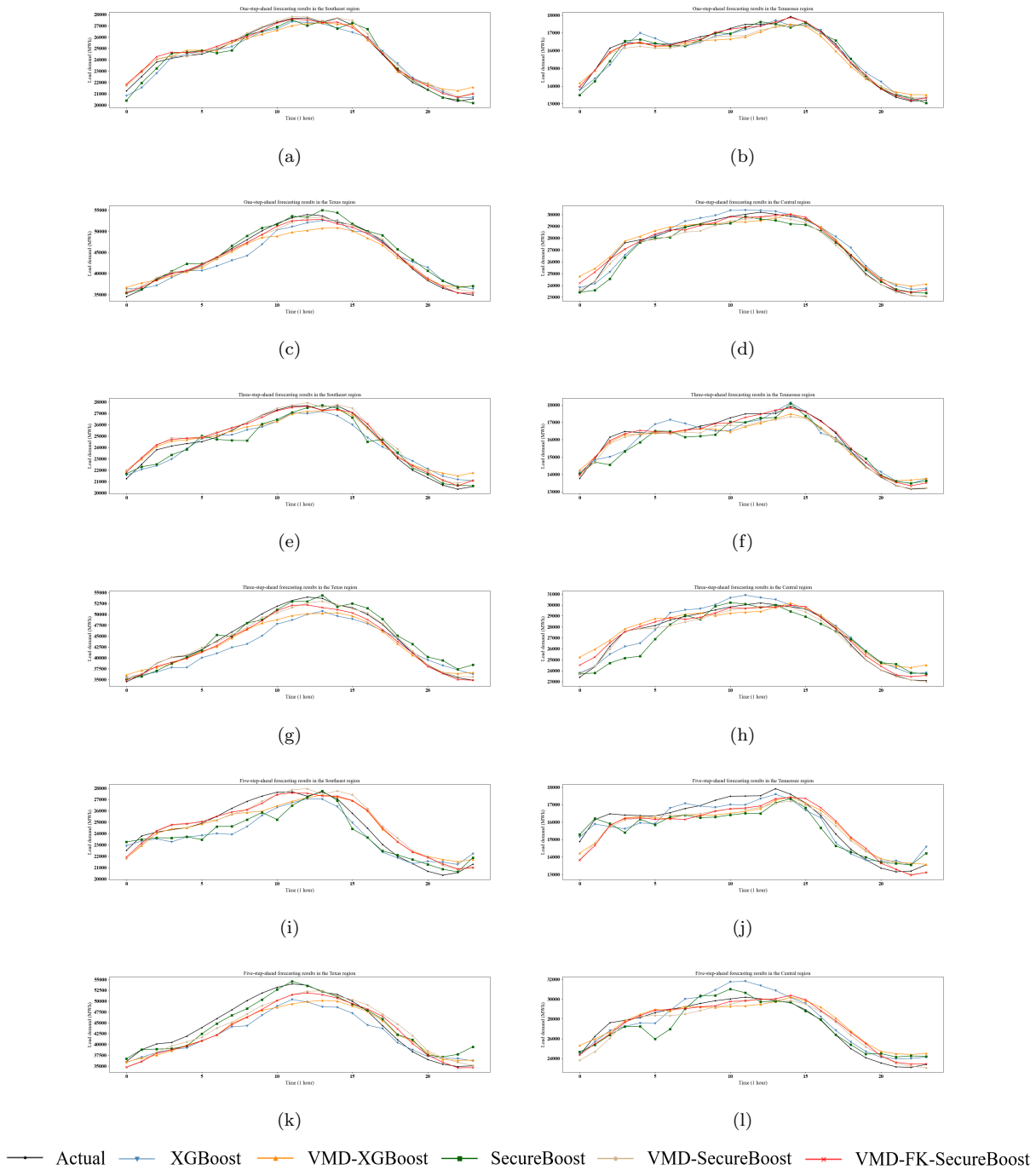

Figure 4: One-, three- and five-step-ahead forecasting of various hybrid models in four different regions of US (Power load data in US is from 00:00 on Oct 20, 2021 to 00:00 on Oct 21, 2021).

better fitting effect than XGBoost and VMD-XGBoost. Therefore, VMD-FK-SecureBoost can improve prediction accuracy while ensuring data privacy.

4.2. Example 2: The data set of the power load in Australia

In Example 2, as an additional evaluation experiment, the integrated algorithm was applied to different distribution zone substations (Berkeley Vale, Camperdown, Newcastle CBD, and North Sydney) in Australia. Example 2 adopted the load data of the total power consumption of those substations. Those load data were collected every 15 mins. Each substation had 34,656 data points (from 00:00 on May 1, 2020 to 24:00 on April 30, 2021). This data set was split into the training set with 27,724 data points (from 00:00 on May 1, 2020 to 19:00 on February 12, 2021) and the testing set with 6,932 data points (from 19:15 on February 12, 2021 to 24:00 on April 30, 2021). Tables 4–6 show the prediction error indicators from the different distribution zone substations in Example 2. Fig. 5 shows the multiple-step-ahead forecasting curves. The following sub-sections will analyze the experiment in detail.

Table 4: One-step-ahead forecasting results of VMD-FK-SecureBoost and the contrast models in the four distribution zone substations in Australia.

Substation	Model	R^2	MAPE	MAE	RMSE
Berkeley Vale	XGBoost	0.628	7.757	1.581	2.101
	VMD-XGBoost	0.778	6.457	1.258	1.624
	SecureBoost	0.718	6.913	1.381	1.829
	VMD-SecureBoost	0.905	4.853	0.951	1.062
	VMD-FK-SecureBoost	**0.964**	**2.547**	**0.505**	**0.656**
Camperdown	XGBoost	0.635	5.329	0.824	1.113
	VMD-XGBoost	0.760	4.362	0.655	0.902
	SecureBoost	0.653	5.223	0.800	1.084
	VMD-SecureBoost	0.936	2.562	0.383	0.466
	VMD-FK-SecureBoost	**0.939**	**2.508**	**0.375**	**0.455**
Newcastle CBD	XGBoost	0.693	5.747	1.014	1.441
	VMD-XGBoost	0.866	4.177	0.702	0.953
	SecureBoost	0.706	5.763	1.024	1.412
	VMD-SecureBoost	0.967	2.252	0.374	0.473
	VMD-FK-SecureBoost	**0.971**	**2.127**	**0.354**	**0.440**
North Sydney	XGBoost	0.784	5.094	1.365	1.869
	VMD-XGBoost	0.833	5.005	1.276	1.643
	SecureBoost	0.872	3.863	1.040	1.440
	VMD-SecureBoost	0.893	4.857	1.220	1.318
	VMD-FK-SecureBoost	**0.960**	**2.560**	**0.649**	**0.808**

NOTE. The power load data of Australia were from 19:15 on February 12, 2021 to 24:00 on April 30, 2021.

4.2.1. Comparative experiment between XGBoost, VMD-XGBoost, SecureBoost, and VMD-SecureBoost

The effectiveness of VMD was verified on the data set of four distribution zone substations in Australia. The empirical method was used in VMD to set the number of mode decompositions. Fig. 5(h) shows that VMD-SecureBoost had the best fitting performance in the three-step-ahead forecasting of the North Sydney zone substation power load data. Due to the periodic and regular sub-sequences decomposed by VMD, VMD-SecureBoost had advantages in prediction.

Table 5: Three-step-ahead forecasting results of VMD-FK-SecureBoost and the contrast models in the four distribution zone substations in Australia.

Region	Model	R^2	MAPE	MAE	RMSE
Berkeley Vale	XGBoost	0.577	8.175	1.679	2.239
	VMD-XGBoost	0.659	7.870	1.551	2.010
	SecureBoost	0.639	7.890	1.578	2.069
	VMD-SecureBoost	0.928	3.747	0.742	0.926
	VMD-FK-SecureBoost	**0.935**	**3.168**	**0.666**	**0.881**
Camperdown	XGBoost	0.453	6.533	1.006	1.362
	VMD-XGBoost	0.625	5.444	0.829	1.128
	SecureBoost	0.478	6.411	0.977	1.330
	VMD-SecureBoost	0.903	2.845	0.439	0.572
	VMD-FK-SecureBoost	**0.907**	**2.809**	**0.431**	**0.562**
Newcastle CBD	XGBoost	0.447	9.282	1.564	1.935
	VMD-XGBoost	0.773	5.537	0.930	1.241
	SecureBoost	0.607	6.517	1.164	1.631
	VMD-SecureBoost	0.943	2.899	0.487	0.624
	VMD-FK-SecureBoost	**0.948**	**2.612**	**0.446**	**0.595**
North Sydney	XGBoost	0.770	5.273	1.414	1.928
	VMD-XGBoost	0.822	5.091	1.307	1.698
	SecureBoost	0.807	4.789	1.287	1.769
	VMD-SecureBoost	0.901	4.410	1.109	1.268
	VMD-FK-SecureBoost	**0.950**	**2.633**	**0.689**	**0.897**

NOTE. The power load data of Australia were from 19:15 on February 12, 2021 to 24:00 on April 30, 2021.

As shown in the evaluation metrics in Tables 4–6, VMD-SecureBoost had the highest prediction accuracy in all four distribution zone substations. For the example of the Newcastle CBD in Table 4, the MAPE of VMD-SecureBoost (2.252) was lower than those of XGBoost (5.747), VMD-XGBoost (4.177), and SecureBoost (5.763). These show that VMD can effectively improve the learning efficiency of the model.

4.2.2. Comparative experiment between VMD-SecureBoost and VMD-FK-SecureBoost

VMD-FK-SecureBoost had a more effective forecasting result than VMD-SecureBoost. Therefore, the proposed model is further compared with VMD-SecureBoost in this sub-section.

Fig. 5(l) shows that the five-step-ahead forecasting result of VMD-FK-SecureBoost was closer to the power load data of the North Sydney zone substation than that of VMD-SecureBoost. The results of the one-step-ahead forecasting study of the Newcastle CBD in Table 4 show that VMD-FK-SecureBoost (R^2: 0.971, MAPE: 2.127, MAE: 0.354, and RMSE: 0.440) had a more effective forecasting result than VMD-SecureBoost (R^2: 0.967, MAPE: 2.252, MAE: 0.374, and RMSE: 0.473). This means that FK provided SecureBoost a better strategy for training, due to which the prediction performance of SecureBoost improved. FK recombined the subsequences obtained by VMD into clusters with the same characteristics through unsupervised learning. Therefore, the forecasting performance of SecureBoost was improved by

Table 6: Five-step-ahead forecasting results of VMD-FK-SecureBoost and the contrast models in the four distribution zone substations in Australia.

Region	Model	R^2	MAPE	MAE	RMSE
Berkeley Vale	XGBoost	0.578	8.204	1.683	2.238
	VMD-XGBoost	0.648	7.997	1.579	2.044
	SecureBoost	0.644	7.832	1.567	2.053
	VMD-SecureBoost	0.923	3.736	0.740	0.954
	VMD-FK-SecureBoost	**0.919**	**3.531**	**0.733**	**0.979**
Camperdown	XGBoost	0.397	6.870	1.056	1.430
	VMD-XGBoost	0.596	5.681	0.866	1.171
	SecureBoost	0.464	6.579	1.000	1.349
	VMD-SecureBoost	0.862	3.427	0.525	0.682
	VMD-FK-SecureBoost	**0.878**	**3.201**	**0.493**	**0.644**
Newcastle CBD	XGBoost	0.455	9.164	1.547	1.921
	VMD-XGBoost	0.764	5.592	0.942	1.263
	SecureBoost	0.610	6.577	1.171	1.625
	VMD-SecureBoost	0.929	2.979	0.511	0.695
	VMD-FK-SecureBoost	**0.932**	**2.812**	**0.492**	**0.681**
North Sydney	XGBoost	0.622	7.796	2.008	2.474
	VMD-XGBoost	0.814	5.176	1.333	1.734
	SecureBoost	0.848	5.499	1.383	1.567
	VMD-SecureBoost	0.799	4.983	1.335	1.804
	VMD-FK-SecureBoost	**0.937**	**2.889**	**0.765**	**1.011**

NOTE. The power load data of Australia were from 19:15 on February 12, 2021 to 24:00 on April 30, 2021.

these clustering results.

In conclusion, VMD-FK-SecureBoost realized STLF with data privacy protection on the four distribution zone substations in Australia without centralized data storage. Furthermore, compared with other contrast models, VMD-FK-SecureBoost achieved the best prediction results on these data sets.

4.2.3. Comparative experiment between XGBoost, VMD-XGBoost and VMD-FK-SecureBoost

In this sub-section the degrees of effectiveness of the centralized and decentralized models for data privacy protection in STLF are compared. VMD-FK-SecureBoost securely aggregated data from the four distribution zone substations in Australia to train the same global model cooperatively.

As shown in Tables 4, 5, and 6, the MAPEs for one-, three-, and five-step-ahead forecasting of VMD-FK-SecureBoost (2.127, 2.612, and 2.812, respectively) are all lower than those of XGBoost (5.747, 9.282, and 9.164, respectively) and VMD-XGBoost (4.177, 5.537, and 5.592, respectively) in the Newcastle CBD. As can be seen in Fig. 5(c), 5(g), and 5(k), VMD-FK-SecureBoost has the closest curves to the actual data compared to XGBoost and VMD-XGBoost in the one-, three-, and five-step-ahead forecasting for the Newcastle CBD. Therefore, VMD-FK-SecureBoost achieved the best prediction result with privacy protection.

Figure 5: One-, three- and five-step-ahead forecasting of various hybrid models in four different distribution zone substations from Australia (Power load data of Australia is from 19:15 on Feb 12, 2021 to 23:15 on Feb 12, 2021).

5. Conclusion

This paper proposed the use of VMD-FK-SecureBoost for STLF with data privacy protection. VMD was used to decompose the original data into several sub-sequences. We verified that unlike the models that do not consider decomposition, VMD can improve the accuracy of the forecasting results. FK was

designed to recombine the aforementioned sub-sequences into new clusters. Compared with the models without a clustering algorithm, VMD-FK-SecureBoost provided the best forecasting accuracy. Finally, the SecureBoost component provided a safe and reliable data interaction platform for several participants and realized the security collaboration of the data sets from different participants. To demonstrate the effectiveness of the proposed algorithm, we tested its forecasting performance in two data sets of actual power load and achieved impressive predictions with small error indicators.

Nevertheless, as FL aims to train the global model by using the parameters and updates of local models, the communication bottleneck has become a key challenge. We will conduct further research to ease the communication restrictions of FL, referring to [24]. In addition, accurate modeling of power is an important challenge for energy control. In the future, we will refer to the novel hybrid modeling method proposed by [4] which combines both recurrent neural networks(RNNs) and Ornstein-Uhlenbeck process, and extend FL to a stochastic optimal control problem to achieve energy control.

Declaration of interests

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.

Credit authorship contribution Statement

Yang Yang: Writing - review & editing, Funding acquisition. Zijin Wang: Software, Visualization, Formal analysis, Writing-original draft. Shangrui Zhao: Writing-review & editing. Jinran Wu: Supervision, Formal analysis, Writing-original draft, Writing-review & editing.

Acknowledgements

The work is supported by the Australian Research Council project (grant number DP160104292), the Australian Research Council Centre of Excellence for Mathematical and Statistical Frontiers (ACEMS) (grant number CE140100049) and the Chinese Fundamental Research Funds for the Central Universities (WUT: 213114009). This work is supported in part by the National Natural Science Foundation of China under Grant 61873130 and Grant 61833011, in part by the Natural Science Foundation of Jiangsu Province under Grant BK20191377, and in part by the 1311 Talent Project of Nanjing University of Posts and Telecommunications. Also, the authors would like to acknowledge Ms. Yufeng Zhang for her preparation for the original manuscript.

References

[1] NR Badurally Adam, MK Elahee, and MZ Dauhoo. Forecasting of peak electricity demand in mauritius using the non-homogeneous gompertz diffusion process. *Energy*, 36(12):6763–6769, 2011.

[2] Miguel Lopez, Carlos Sans, and Sergio Valero. Automatic classification of special days for short-term load forecasting. *Electric Power Systems Research*, 202:107533, 2022.

[3] Qiang Yang, Yang Liu, Tianjian Chen, and Yongxin Tong. Federated machine learning: Concept and applications. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 10(2):1–19, 2019.

[4] Haochen Hua, Yuchao Qin, Chuantong Hao, and Junwei Cao. Stochastic optimal control for energy internet: A bottom-up energy management approach. *IEEE Transactions on Industrial Informatics*, 15(3):1788–1797, 2019.

[5] Michel Bessani, Julio AD Massignan, Talysson MO Santos, João BA London Jr, and Carlos D Maciel. Multiple households very short-term load forecasting using bayesian networks. *Electric Power Systems Research*, 189:106733, 2020.

[6] Mohamed El-Hendawi and Zhanle Wang. An ensemble method of full wavelet packet transform and neural network for short term electrical load forecasting. *Electric Power Systems Research*, 182: 106265, 2020.

[7] Quande Qin, Zhaorong Huang, Zhihao Zhou, Yu Chen, and Weigang Zhao. Hodrick–prescott filter-based hybrid arima–slfns model with residual decomposition scheme for carbon price forecasting. *Applied Soft Computing*, 119:108560, 2022.

[8] Alex D Papalexopoulos and Timothy C Hesterberg. A regression-based approach to short-term system load forecasting. *IEEE Transactions on Power Systems*, 5(4):1535–1547, 1990.

[9] Zibo Dong, Dazhi Yang, Thomas Reindl, and Wilfred M Walsh. Short-term solar irradiance forecasting using exponential smoothing state space model. *Energy*, 55:1104–1113, 2013.

[10] Yuting Lu, Gaocai Wang, and Shuqiang Huang. A short-term load forecasting model based on mixup and transfer learning. *Electric Power Systems Research*, 207:107837, 2022.

[11] Zhenyu Zhao, Yao Zhang, Yujia Yang, and Shuguang Yuan. Load forecasting via grey model-least squares support vector machine model and spatial-temporal distribution of electric consumption intensity. *Energy*, page 124468, 2022.

[12] Hamed HH Aly. A hybrid optimized model of adaptive neuro-fuzzy inference system, recurrent kalman filter and neuro-wavelet for wind power forecasting driven by dfig. *Energy*, 239:122367, 2022.

[13] Yang Yang, Zijin Wang, Yuchao Gao, Jinran Wu, Shangrui Zhao, and Zhe Ding. An effective dimensionality reduction approach for short-term load forecasting. *Electric Power Systems Research*, 210:108150, 2022.

[14] Yang Yang, Hu Zhou, Jinran Wu, Chan-Juan Liu, and You-Gan Wang. A novel decompose-cluster-feedback algorithm for load forecasting with hierarchical structure. *International Journal of Electrical Power & Energy Systems*, 142:108249, 2022.

[15] Weimin Yue, Qingrong Liu, Yingjun Ruan, Fanyue Qian, and Hua Meng. A prediction approach with mode decomposition-recombination technique for short-term load forecasting. *Sustainable Cities and Society*, 85:104034, 2022.

[16] Long H Nguyen, Zhenhe Pan, Opeyemi Openiyi, Hashim Abu-gellban, Mahdi Moghadasi, and Fang Jin. Self-boosted time-series forecasting with multi-task and multi-view learning. *arXiv preprint arXiv:1909.08181*, 2019.

[17] Mohammad Navid Fekri, Katarina Grolinger, and Syed Mir. Distributed load forecasting using smart meter data: Federated learning with recurrent neural networks. *International Journal of Electrical Power & Energy Systems*, 137:107669, 2022.

[18] Arooj Arif, Nadeem Javaid, Mubbashra Anwar, Afrah Naeem, Hira Gul, and Sahiba Fareed. Electricity load and price forecasting using machine learning algorithms in smart grid: A survey. In *Workshops of the International Conference on Advanced Information Networking and Applications*, pages 471–483. Springer, 2020.

[19] Regulation (eu) 2016/679 of the european parliament and of the council of 27 april 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing directive 95/46/ec (general data protection regulation) (text with eea relevance), May 2016.

[20] Nastaran Gholizadeh and Petr Musilek. Federated learning with hyperparameter-based clustering for electrical load forecasting. *Internet of Things*, 17:100470, 2022.

[21] Kai Hu, Yaogen Li, Min Xia, Jiasheng Wu, Meixia Lu, Shuai Zhang, and Liguo Weng. Federated learning: a distributed shared machine learning method. *Complexity*, 2021, 2021.

[22] Nastaran Gholizadeh and Petr Musilek. Distributed learning applications in power systems: A review of methods, gaps, and challenges. *Energies*, 14(12), 2021.

[23] Seifeddine Messaoud, Abbas Bradai, Syed Hashim Raza Bukhari, Pham Tran Anh Quang, Olfa Ben Ahmed, and Mohamed Atri. A survey on machine learning in internet of things: algorithms, strategies, and applications. *Internet of Things*, 12:100314, 2020.

[24] Zichen Ma, Zihan Lu, Yu Lu, Wenye Li, Jinfeng Yi, and Shuguang Cui. Federated two-stage learning with sign-based voting. *arXiv preprint arXiv:2112.05687*, 2021.

[25] Yang Liu, Zhuo Ma, Zheng Yan, Zhuzhu Wang, Ximeng Liu, and Jianfeng Ma. Privacy-preserving federated k-means for proactive caching in next generation cellular networks. *Information Sciences*, 521:14–31, 2020.

[26] Hemant H Kumar, Karthik V R, and Mydhili K Nair. Federated k-means clustering: A novel edge ai based approach for privacy preservation. In *2020 IEEE International Conference on Cloud Computing in Emerging Markets (CCEM)*, pages 52–56. IEEE, 2020.

[27] Yi Wang, Mengshuo Jia, Ning Gao, Leandro Von Krannichfeldt, Mingyang Sun, and Gabriela Hug. Federated clustering for electricity consumption pattern extraction. *IEEE Transactions on Smart Grid*, 13:2425–2439, 2022.

[28] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794, 2016.

[29] Kewei Cheng, Tao Fan, Yilun Jin, Yang Liu, Tianjian Chen, Dimitrios Papadopoulos, and Qiang Yang. Secureboost: A lossless federated learning framework. *IEEE Intelligent Systems*, 36:87–98, 2021.

[30] Jian Ma, Jian Zheng, Yifan Yang, Chunting Kang, Yuxin Wang, Yang Wang, and Yutong Li. Research on analysis of power and water consumption data for group tenant identification based on federated learning. In *International Conference on Computational Modeling, Simulation, and Data Analysis (CMSDA 2021)*, volume 12160, pages 548–552. SPIE, 2022.

[31] Yuxin Liang, Zhiyong Liu, Yong Song, Aidong Yang, Xiaozhou Ye, and Ye Ouyang. A methodology of trusted data sharing across telecom and finance sector under china's data security policy. In *2021 IEEE International Conference on Big Data (Big Data)*, pages 5406–5412. IEEE, 2021.

[32] Konstantin Dragomiretskiy and Dominique Zosso. Variational mode decomposition. *IEEE Transactions on Signal Processing*, 62(3):531–544, 2013.

[33] Gang Liang and Sudarshan S Chawathe. Privacy-preserving inter-database operations. In *International Conference on Intelligence and Security Informatics*, pages 66–82. Springer, 2004.

[34] Dongchuan Yang, Ju-e Guo, Shaolong Sun, Jing Han, and Shouyang Wang. An interval decomposition-ensemble approach with data-characteristic-driven reconstruction for short-term load forecasting. *Applied Energy*, 306:117992, 2022.