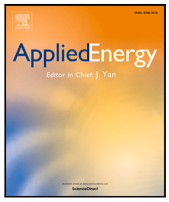
Applied Energy 337 (2023) 120860

Contents lists available at ScienceDirect

Applied Energy

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



Privacy-preserving knowledge sharing for few-shot building energy prediction: A federated learning approach

Lingfeng Tang, Haipeng Xie *, Xiaoyang Wang, Zhaohong Bie

State Key Laboratory of Electrical Insulation and Power Equipment, School of Electrical Engineering, Xi'an Jiaotong University, Xi'an, 710049, China

ARTICLE INFO

Dataset link: https://www.kaggle.com/datasets/claytonmiller/buildingdatagenomeproject2

Keywords:
Few-shot building energy prediction
Federated learning
Privacy protection
Knowledge sharing
Data heterogeneity

ABSTRACT

The data-driven method is a promising way to predict the energy consumption of buildings, however suffering from the data shortage problem in various scenarios. Even though transfer learning can improve the few-shot prediction performance by utilizing other buildings' data, the centralized approach poses potential privacy risks. To tackle this issue, the paper proposes a privacy-preserving knowledge sharing framework to facilitate the few-shot building energy prediction based on federated learning. First, a private data aggregation scheme is established to encrypt the sensitive data with shared random masks and guarantee the privacy of the data preprocessing and model optimization. Then, to alleviate the intrinsic data heterogeneity, a dynamical clustering federated learning algorithm is proposed to implement the intra-cluster and inter-cluster knowledge sharing along with the iterative clustering process for participating buildings. Finally, the network-based transfer learning approach is incorporated into the distributed framework to establish the customized model based on trained cluster models and further boost the prediction performance for each building. Extensive experiments on the Building Data Genome Project 2 (BDGP2) dataset indicate that the federated approach witnesses a desirable prediction performance while preserving the privacy of building occupants.

1. Introduction

Building energy consumption and associated carbon dioxide (CO_2) emissions have substantially increased due to the urbanization and population growth. Statistics indicate that the building sector is responsible for 36% of the world's total final energy consumption and 37% of energy-related CO_2 emissions in 2020 [1]. Thereby, the building energy management system (BEMS) plays a significant role in improving the building energy efficiency and reducing energy consumption, finally contributing to net-zero energy buildings with low carbon emissions.

As the fundamental basis of the BEMS, building energy prediction can provide control-oriented models for operation strategies, evaluate operation energy efficiency and improve the reliability of grid-building integrated systems [2]. With the mass deployment of smart meters in buildings and rapid development of machine learning methods, data-driven prediction models have drawn wide attention in academics and have been exploited extensively based on the abundant operational data to predict building energy consumption [3–5].

Although data-driven building energy prediction methods have witnessed unprecedented achievements, the prediction performance immensely relies on the quantity and the quality of available data [6,7]. However, numerous buildings lack reliable measurement systems to provide accurate operational data with appropriate temporal granularity. The operational data records may also be unavailable due to the devoid of data storage systems. Moreover, newly constructed buildings suffer from limited operation time and conditions, which is intrinsically impossible to accumulate sufficient operational data within a short period. These scenarios extensively limit the practicability and scalability of data-dependent building energy prediction methods. Meanwhile, current research mainly focuses on establishing customized energy prediction models of individual buildings with sufficient data [8–14]. These methods are not applicable to the few-shot building energy consumption problem, namely predicting building energy consumption for buildings with limited data.

Recent studies proposed a knowledge sharing framework for building energy prediction via transfer learning [15–18]. They aimed at fully utilizing the sufficient data of source buildings to train a shared model and then transferring it to facilitate the energy consumption prediction performance of the target building. However, the pretraining process of transfer learning refers to utilizing the data in the source domain to train a model [19], so it requires aggregating all data of buildings in the source domain. The centralized data aggregation and analysis disclose the occupancy information and inevitably lead to privacy issues among buildings. This method is also inconsistent

* Corresponding author.
E-mail address: haipengxie@xjtu.edu.cn (H. Xie).

https://doi.org/10.1016/j.apenergy.2023.120860
Received 1 July 2022; Received in revised form 1 December 2022; Accepted 14 February 2023
Available online 1 March 2023

with data protection regulations pursued by governments, such as the General Data Protection Regulation (GDPR) of the Europe Union since 2018 [20] and the California Consumer Privacy Act (CCPA) of the United States since 2020 [21]. The above privacy concerns immensely hinder the implementation of transfer learning on the building energy prediction modeling with limited data.

Federated learning has been proposed to train a global machine learning model collaboratively without exchanging the data [22], presenting a promising solution to address the aforementioned research gap. In the typical framework of federated learning, all clients exploit the local data to cooperatively approximate the globally optimal model under the coordination of the central server. The whole training process does not involve any raw data exchange. Thereby, federated learning, as a scalable distributed machine learning paradigm, can avoid transferring the privacy-sensitive data and utilize local computational and storage resources to collaboratively build the model. Additionally, federated learning is able to incorporate multi-party computation (MPC) techniques [23] and personalization techniques [24] to further enhance the privacy protection and mitigate performance degradation caused by the data heterogeneity, respectively.

Thereby, this paper proposes a privacy-preserving knowledge sharing framework based on federated learning to improve the prediction performance of buildings with limited data. The privacy of participating buildings is guaranteed along with the entire knowledge sharing process. Two related works have investigated the application potential of federated learning in the electric load forecasting field and broken the traditional system-level aggregated load forecasting scheme [25,26]. Compared to them, this paper aims to tackle the data shortage problem in the building energy prediction field via federated learning and establish a knowledge sharing framework in contrast to previous customized building energy consumption modeling methods. Moreover, this paper establishes a privacy protection scheme for sensitive data aggregation, and implements the dynamical clustering federated learning and personalized federated transfer learning towards the intrinsic data heterogeneity phenomenon. Experimental results on the benchmark dataset also validate the performance improvement of the proposed federated approach.

This paper establishes a privacy-preserving knowledge sharing framework based on federated learning to tackle the few-shot building energy prediction problem. On the basis of the distributed framework, the major contributions of this paper include: (1) Establishing a private data aggregation scheme to encrypt the sensitive information by adding agreed random masks and providing guarantees for the privacy of collaboratively extracting data statistical characteristics and optimizing federated cluster models. The privacy protection scheme can be implemented without establishing interconnected communication links among buildings or introducing an extra trusted third party. (2) Proposing a dynamical clustering federated learning algorithm to group participating buildings according to local data distributions and avoid the model performance degradation caused by apparent data discrepancies. Furthermore, transfer learning is integrated into the federated building energy prediction framework to establish the customized prediction model and boost the prediction performance for each participating and nonparticipating building.

The remainder of the paper is organized as follows. Section 2 investigates the related literature on data-driven building energy prediction and federated learning. Section 3 elaborates on the proposed privacy-preserving knowledge sharing method for building energy prediction. Section 4 describes the experiment setup and provides a detailed discussion of experiment results. Conclusions are given in Section 5.

2. Literature review

2.1. Data-driven methods for building energy prediction

Data-driven methods have aroused wide attention in building energy consumption prediction due to the highly expressive capacity for nonlinear relationships and the fast computational ability for real-time operational data. This approach is able to achieve a similar or even better prediction performance and avoid the time-consuming energy simulation based on physical laws.

Advanced machine learning models, such as eXtreme Gradient Boosting (XGBoost) [8] and random forest (RF) [9], have been applied to predict energy usage and validated by real or simulated datasets. Moreover, deep learning has been studied extensively in the building energy prediction field for its universal approximation power [10]. Long short term memory (LSTM) networks were applied to learn the temporal interdependence of the building energy consumption data [11]. On the basis of LSTM, attention mechanism was adopted to enhance the memory capability [12] and improve the model interpretability [13]. Furthermore, convolution neural networks (CNN) was combined with LSTM to extract the spatial and temporal features simultaneously [14]. Comprehensive and extensive experiments have been conducted in the above studies, and corresponding results have proven the effectiveness of machine learning methods in describing the complex nonlinear relationship between multiple influencing factors and building energy predictions.

However, these customized building prediction methods cannot be directly applied to limited data scenarios owing to the over-fitting problem. Transfer learning, as one practical approach to breaking restrictions of limited data, has been adopted to utilize the knowledge learned from other buildings to improve the robustness and generalization of prediction models [6]. The combined transfer learning and Adaboost algorithm were proposed to forecast the cooling load of buildings, and the prediction accuracy based on load data of one week increased by 10% compared with no knowledge transfer [7]. The combined LSTM and CNN in the transfer learning framework were adopted to extract features of energy usage data owned by 407 buildings, then network parameters were reused to enhance the prediction performance of target buildings with limited data [15]. Other advanced data-driven models, for instance, LSTM and domain adversarial neural networks [16], the sequence-to-sequence model [17] have also been incorporated into the transfer learning framework to share the profitable prior knowledge.

The knowledge sharing process based on transfer learning relies on aggregating the data of multiple buildings in the source domain and extracting the universal knowledge in a centralized manner. It is evident that the centralized method will lead to energy usage pattern leakage and impose significant privacy risks for building occupants. Except for the mainstream transfer learning-enabled approach illustrated above, existing studies that adopt only one building with sufficient data as the source domain can avoid privacy issues among different buildings [16]. However, this approach may suffer from data heterogeneity and result in model performance degradation when there exists a large domain shift between the source domain and the target domain [27]. Meanwhile, the building in the source domain is always reluctant to consume the computation and communication resources to train a pretraining model without any benefits such as the prediction performance improvement for itself. Moreover, transfer learning based on open datasets can be another approach with relatively fewer privacy issues. However, existing open datasets are concentrated in the United States and Europe, and the energy usage pattern may differ considerably in different climates or countries, which significantly impacts the scalability of this approach. Furthermore, some research proposed adopting the simulation dataset as the source domain [28]. Nevertheless, simulation procedures for individual buildings are time-consuming and cumbersome. Compared with above methods based on transfer learning, this paper proposes a privacy-preserving knowledge sharing framework based on federated learning to fully take advantage of existing building energy data resources and exploit the learned knowledge to improve the prediction performance under limited data scenarios.

2.2. Concepts and applications of federated learning

Federated learning, as a distributed machine learning paradigm, has recently gained significant popularity for its privacy-preserving property [22]. A typical federated learning framework comprises a central server and multiple local clients. In each communication round of federated learning, selected clients utilize the local data to train the broadcasted global model individually. Then the central server aggregates local models to obtain the updated global model. The training goal of federated learning is to approximate a globally optimal model for most local data without compromising privacy.

Federated learning has been widely applied in various fields [29], such as mobile devices, financial security and health care, since it was firstly proposed by Google in 2016 [30]. Combined federated learning and deep reinforcement learning were incorporated into mobile edge systems to mitigate the communication overhead [31] and optimize the computation offloading [32]. Federated meta-learning framework was proposed to establish the shared fraud detection model without exchanging sensitive transaction records [33]. Healthcare data monitoring was also implemented based on federated learning to detect skin diseases [34].

However, the standard federated learning approach only establishes a single model for all clients. The federated model without personalization suffers from model performance degradation due to the intrinsic data heterogeneity phenomenon of distributed clients. Therefore, personalized federated learning has been proposed to learn from the non-independent and identically distributed (non-IID) data and build a customized model for each client [24]. From the perspective of enhancing statistical homogeneity, data augmentation methods such as the generative autoencoder are adopted to generate samples and augment local datasets [35]. Moreover, reinforcement learning is applied to select participating clients and compose balanced datasets for federated learning [36]. From the perspective of model personalization, clustering algorithm, knowledge distillation and transfer learning have been applied to establish customized models in recent literature [37–39].

Although federated learning has witnessed noticeable progress in multiple industrial and commercial fields, it has rarely been applied in building energy management systems, especially in the building energy prediction. Thus this paper provides an exhaustive exploration of adopting the federated learning integrated with cryptography techniques and deep learning methods for the few-shot building energy prediction. More specifically, the private data aggregation scheme is established to guarantee the data privacy for the proposed framework. Furthermore, towards the inherently heterogeneous building energy consumption data, dynamical clustering federated learning algorithm is proposed to group buildings iteratively without compromising privacy and conduct the inter-cluster and intra-cluster knowledge sharing along with the clustering process. On the basis of established cluster models, transfer learning is integrated into this framework to implement personalized prediction for each building.

3. Methodology

3.1. Research outline

This study is conducted to establish a privacy-preserving knowledge sharing framework for one-hour ahead building energy consumption prediction with limited data based on federated learning.

The overall architecture of the federated knowledge sharing framework is depicted in Fig. 1. The urban energy system can undertake the role of the central server, and distributed buildings participate in the knowledge sharing framework to cooperatively improve the prediction performance without establishing interconnected communication links and exchanging the raw data. As shown in Fig. 1, a private data aggregation scheme is firstly proposed to compute the statistical characteristics of the distributed data owned by different buildings and guarantee the privacy of the model aggregation process. On the basis of the building energy prediction model, a dynamical clustering federated learning algorithm is proposed to group participating buildings iteratively and implement intra-cluster and inter-cluster knowledge sharing along with the grouping process. To eliminate data heterogeneity's impacts, transfer learning is eventually incorporated into the federated learning framework to establish the personalized prediction model for each building. A knowledge sharing mechanism for nonparticipating buildings is also provided in this Section to improve the scalability of the proposed federated learning framework.

The central server and all buildings have to follow the communication protocol and machine learning algorithms in the framework, however, they may also attempt to analyze the exchanged data and then obtain the private information of other participating individuals [40]. In such a semi-honest environment, our research is required to guarantee the honest-but-curious server and buildings learn no sensitive information along with the federated building energy prediction process.

Note that the proposed federated learning approach belongs to the standard offline training and online application machine learning paradigm. The time-consuming offline training has no impact on the efficiency of the online application process. Thereby, the training process has no strict requirements on computation time and communication time. Moreover, the central server, namely the urban energy system, and multiple participating buildings, all have enough resources to complete the data encryption, model training, and communication process.

3.2. Private data aggregation scheme

Data preprocessing is an indispensable step for obtaining an accurate and reliable building energy prediction model. The procedures of the data preprocessing process contain data cleaning, data transformation, and data normalization. The distributed buildings can reach a consensus on the outlier elimination, missing data interpolation, feature selection, and categorical variable transformation without leaking private information, and the detailed description will be presented in Section 4, Experiments and results. However, the data normalization in the data preprocessing stage aims at transforming input numerical variables with different magnitudes into a similar scale, requiring the statistical characteristics of all distributed training data. Meanwhile, the input normalization needs to be conducted on the union dataset to maintain the interoperability of prediction models among all buildings according to other studies [16,27,41]. As a result, the direct collection of each building's statistical characteristics will inevitably result in privacy issues. Moreover, the model aggregation or gradient aggregation is also an essential procedure in federated learning. But existing studies have also pointed out that the private training data can be reconstructed from the uploaded model or gradient of each client [42,43]. Therefore, a privacy protection scheme is needed to normalize numerical variables and aggregate uploaded models without compromising occupancy privacy. For the sake of clarity, data normalization is adopted as the instance to illustrate the proposed scheme in the following part. The private model aggregation process will be provided in Section 3.3.2, Dynamical clustering federated learning algorithm.

Given N participating buildings and their local training datasets $D_1, D_2, \ldots, D_N$, the z-score normalization is adopted to process numerical variables and compute the mean $\boldsymbol{\mu} = [\mu_1, \ldots, \mu_C]^{\mathrm{T}}$ and standard deviation $\boldsymbol{\sigma} = [\sigma_1, \ldots, \sigma_C]^{\mathrm{T}}$ with respect to C numerical features on the whole training dataset $D = D_1 \cup D_2 \cup \cdots \cup D_N$. Then, the mean and standard deviation are applied to normalize the test set to avoid future information leakage. For the ith numerical variable x_i, the z-score normalization operation can be reformulated as:

$$\hat{x}_i = \frac{x_i - \mu_i}{\sigma_i} \tag{1}$$

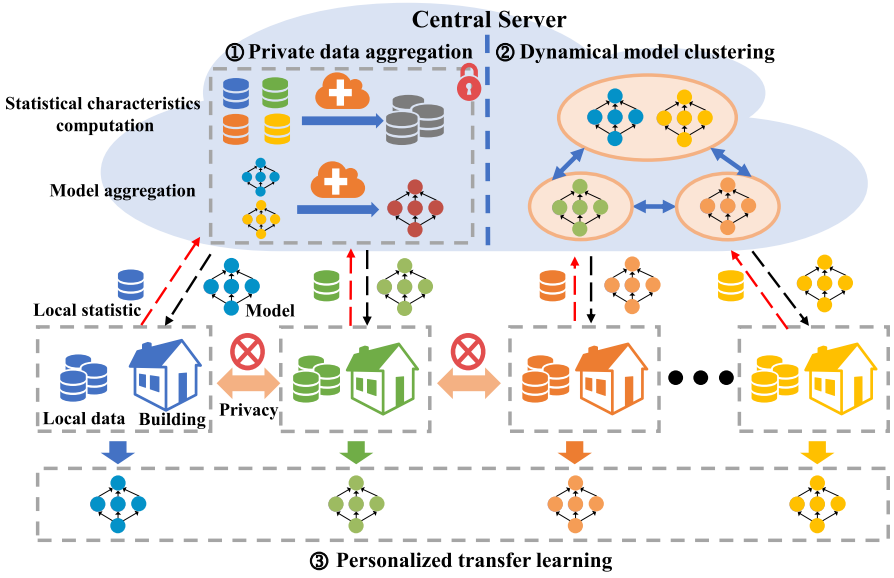


Fig. 1. Federated-learning-based building energy prediction architecture.

$$\mu_i = \frac{1}{M}\sum_{j=1}^{N}\sum_{x_i \in D_j} x_i \tag{2}$$

$$\sigma_i = \sqrt{\frac{1}{M-1}\left(\sum_{j=1}^{N}\sum_{x_i \in D_j}(x_i)^2 - M\mu_i^2\right)} \tag{3}$$

where $\hat{x}_i$ is the ith numerical variable after normalization, M is the total sum of samples for all local training datasets, namely $M = \sum_{j=1}^{N}|D_j|$, $|D_j|$ means the sample size of the jth local training dataset D_j. The sample size of each building is not sensitive information, thus the central server can obtain M by directly aggregating all sample sizes of local training datasets.

Each building is required to locally calculate a $2C$-dimensional vector $\boldsymbol{X}^i$ with respect to C numerical features on its own dataset D_i, and then upload it to the server. The vector $\boldsymbol{X}^i$ can be represented as:

$$\boldsymbol{X}^i = [\sum_{x_1 \in D_i} x_1, \ldots, \sum_{x_C \in D_i} x_C, \sum_{x_1 \in D_i}(x_1)^2, \ldots, \sum_{x_C \in D_i}(x_C)^2]^{\mathsf{T}} \tag{4}$$

However, the uploaded vector $\boldsymbol{X}^i$ reflects the energy consumption pattern for the ith building and may give rise to privacy issues. Thereby, the private data aggregation scheme is established based on the pair-wise masking [44] and elliptic curve Diffie–Hellman (ECDH) [45]. The scheme is able to guarantee the correctness of computations on the mean and standard deviation of the whole dataset on the premise of privacy protection. The overall architecture is provided in Fig. 2.

(1) *ECDH key agreement*

ECDH key agreement protocol allows buildings to generate shared secret keys over a non-secure communication channel, namely the honest-but-curious server in this paper. As is shown in Fig. 2, the domain parameters of ECDH, including the elliptic curve, the base point G, the prime p, the order n, and the cofactor h of the subgroup, are firstly broadcasted to all buildings by the server. On the basis of the same domain parameters, each building i selects an integer k_i from $[2, n-1]$ as the private key, then calculates and uploads the public key k_iG to the server. Subsequently, the server, as the intermediate communication node, collects all public keys and distributes them to all buildings. Each building i finally computes $N-1$ shared secret keys $\{j \in N, j \neq i | k_i k_j G\}$ for reaching an consensus on pair-wise random vectors.

The security of the ECDH key agreement protocol is guaranteed by the elliptic curve discrete logarithm problem (ECDLP). Although the server possesses the base point G and the public key k_iG, it is still impossible to solve the private key k_i in polynomial time [46].

(2) *Random vectors agreement*:

Each building i selects and encrypts $2C$-dimensional random vectors $\boldsymbol{R}^{i,j}$ $(j \in N, j > i)$ locally as $[\boldsymbol{R}^{i,j}]$ based on shared secret keys $k_i k_j G$ $(j \in N, j > i)$. After the server collects all pairs of encrypted random vectors, each building i will receive related encrypted random vectors and decrypt them as $\boldsymbol{R}^{i,j}$ $(j \in N)$ to mask the privacy-sensitive $\boldsymbol{X}^i$.

(3) *Pair-wise masking*:

Pair-wise masking method is applied to conceal the actual values of $\boldsymbol{X}^i$ by adding random vectors, and meanwhile guarantee the correctness of computations on the mean and standard deviation of the whole dataset.

On the basis of shared random vectors $\boldsymbol{R}^{i,j}(j \in N)$, $\boldsymbol{X}^i$ can be transformed into an masked form $\boldsymbol{Y}^i$ as:

$$\boldsymbol{Y}^i = \boldsymbol{X}^i + \sum_{j \in N, j > i} \boldsymbol{R}^{i,j} - \sum_{j \in N, j < i} \boldsymbol{R}^{j,i} \tag{5}$$

Each building i subsequently uploads the masked vector $\boldsymbol{Y}^i$ to the server without leaking the private information of $\boldsymbol{X}^i$. Meanwhile, the server is able to obtain the mean $\boldsymbol{\mu}$ and standard deviation $\boldsymbol{\sigma}$ of the whole dataset D by summing up $\boldsymbol{Y}^i$ $(i \in N)$ and operating Eq. (2) and (3), and then broadcasts $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ to all buildings for data normalization. The correctness of the aggregation operation to reveal the actual sum of $\boldsymbol{X}^i$, namely $\boldsymbol{X}$, can be verified by the following equation:

$$\boldsymbol{X} = \sum_{i=1}^{N}\boldsymbol{Y}^i = \sum_{i=1}^{N}\boldsymbol{X}^i + \sum_{i=1}^{N}\left(\sum_{j \in N, j > i}\boldsymbol{R}^{i,j} - \sum_{j \in N, j < i}\boldsymbol{R}^{j,i}\right) = \sum_{i=1}^{N}\boldsymbol{X}^i \tag{6}$$

As illustrated above, the private data aggregation scheme does not require interconnected communication links among buildings or an extra trusted third party. Thus the corresponding potential risks can be avoided in the scheme.

3.3. Federated-learning-based building energy prediction

Traditional customized building energy prediction methods are not applicable for buildings with limited data due to the over-fitting problem. Utilizing the prior knowledge extracted from other buildings to improve the prediction performance of the target building can be an efficient and practical approach in few-shot learning scenarios. However, local data distributions present apparent discrepancies for different buildings due to diverse occupancy schedules, and meanwhile, it is hard for machine learning models to process and analyze the data with independent but non-identical distribution, thus the data heterogeneity

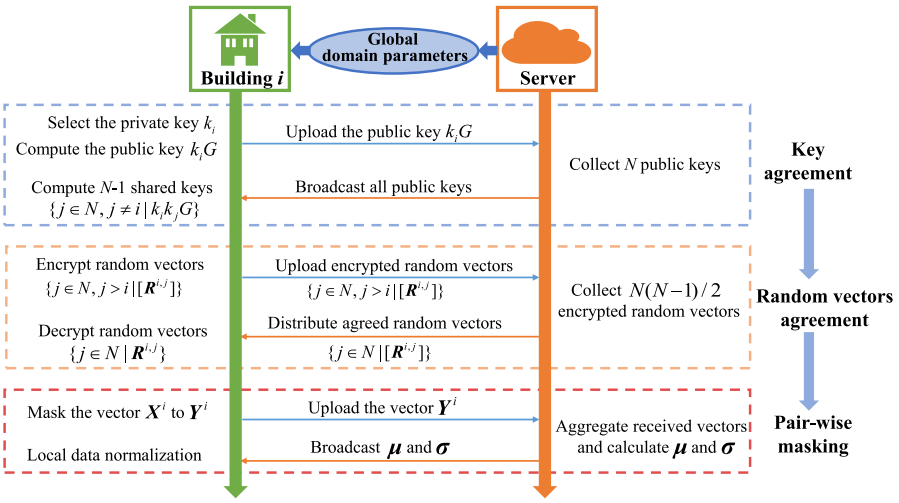


Fig. 2. Private data aggregation scheme.

phenomenon will hinder the knowledge sharing process for multiple buildings and result in the prediction performance degradation.

To tackle the data heterogeneity problem in this distributed framework, this paper proposes a dynamical clustering federated learning algorithm to obtain clusters composed of buildings with similar data distribution. Furthermore, a federated transfer learning algorithm is applied to establish the personalized prediction model for each building. The building energy prediction model architecture and these two algorithms are described below.

3.3.1. *Building energy prediction model architecture*

The research aims at predicting one-hour ahead energy consumption for each building, which is a typical time-series prediction problem. The building energy consumption prediction model is composed of two blocks. Firstly, LSTM has implemented excellent prediction performance in recent studies [11,27] for its recurrent structure and powerful memory cells to process the time series data with long-term dependencies. Thereby, LSTM is selected as the first block of the building energy prediction model to extract the temporal features of the input data. Sigmoid function and tanh function are selected as the activation functions of gated units and hidden neurons of LSTM, respectively [47]. In order to alleviate the over-fitting problem, Dropout, as a regularization technique, is adopted for each LSTM layer to drop out randomly selected LSTM units during the training process. The second block of the prediction model is the multi-layer perceptron (MLP). On the basis of the flatten layer for removing the redundant time dimension, MLP utilizes the output at the last time step of LSTM as the block input, and establishes the mapping relationship between the extracted temporal features and the building energy consumption. ReLU function is selected as the activation function of MLP for its ability to avoid the vanishing gradients problem. Other model architectures such as two LSTM blocks, two MLP blocks, or two gated recurrent unit (GRU) blocks are also applicable in the federated building energy prediction framework but are not the focus of this work. Finally, the mean-squared error loss function is utilized to measure prediction errors and optimize network parameters. Meanwhile, the L2 regularization term is added to the loss function to constrain parameters and prevent the over-fitting problem.

3.3.2. *Dynamical clustering federated learning algorithm*

In the typical federated learning framework, the central server firstly broadcasts the global prediction model to all participating buildings in each communication round. Then, distributed buildings utilize the local data to update individual model parameters. Finally, the server aggregates all uploaded local models and obtains the updated global prediction model. Although the typical federated learning has implemented the knowledge sharing across multiple buildings and guaranteed the privacy of the distributed training process, this global aggregation approach only establishes a single model for all buildings and does not take the data heterogeneity phenomenon into consideration. Thereby, the dynamical clustering federated learning algorithm depicted in Fig. 3 is proposed to identify cluster components based on the similarities among participating buildings. Meanwhile, it can optimize all federated learning models of clusters by intra-cluster and inter-cluster knowledge sharing along with the clustering process. The detailed description of the dynamical clustering federated learning algorithm is provided as follows.

In terms of the clustering process in the federated building energy prediction framework, the traditional centralized clustering approach based on energy consumption profiles needs to aggregate local data of all buildings [48], it violates privacy constraints and is not applicable to the privacy-preserving framework. Moreover, uploaded prediction model parameters in the training stage of federated learning can be utilized to cluster distributed buildings without compromising privacy [37]. However, prediction models with different parameters may have similar mapping relationships between input variables and prediction results. The phenomenon, namely permutation invariance [49], is essentially caused by the non-convex optimization characteristic of the parameter update process for deep learning. In order to group buildings with similar local data distributions into the same cluster, the validation loss is finally adopted as the clustering basis to select the most appropriate cluster model with respect to the local data distribution of each building and estimate cluster components. Moreover, iterative clustering is applied to avoid inappropriate grouping results with only one clustering iteration [50].

On the basis of the above considerations, the central server firstly generates initial model parameters θ_k^0, $k = 1, 2, \ldots, K$ for K clusters. Then, the cluster composition identification and cluster model optimization are conducted in each global communication round under the privacy constraints. Note that each building receives and updates K models owned by K clusters, but only uploads the models that participate in the knowledge sharing process.

(1) *Cluster composition identification*

In the tth global communication round, the server firstly distributes K model parameters to each participating building. Then, each building i conducts the local updates for K received cluster models $\theta_{i,k}^t$ individually. Specifically, in the eth local epoch, the local dataset D_i is firstly split into batches of size B. Then, the mean-squared error loss function F and the Adam algorithm are applied to update the model parameters $\theta_{i,k}^t$. The model parameters update is conducted for each

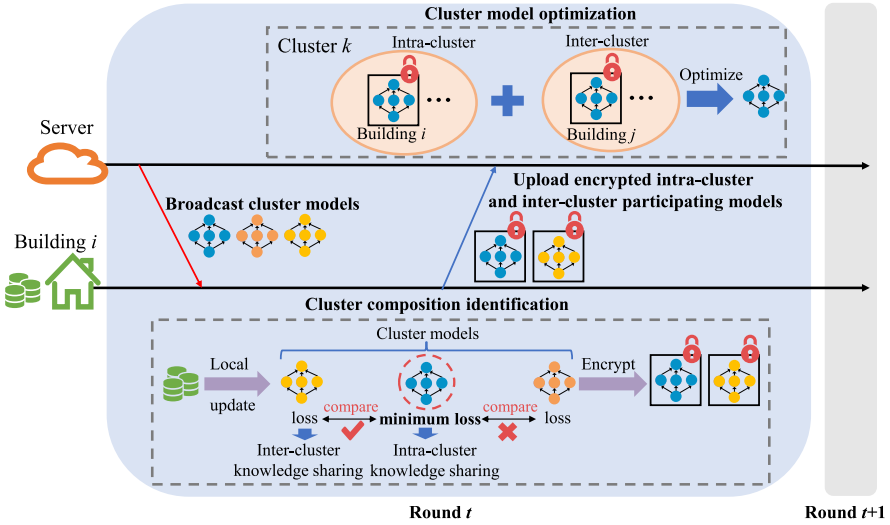


Fig. 3. Dynamical clustering federated learning algorithm.

batch of the local dataset in each local epoch, and the qth update process for parameters $\theta_{i,k}^{t,q}$ can be calculated as:

$$m_q = (\beta_1 m_q + (1-\beta_1)\nabla_{\theta_{i,k}^{t,q}} F_q)/(1-\beta_1) \tag{7}$$

$$v_q = (\beta_2 v_q + (1-\beta_2)(\nabla_{\theta_{i,k}^{t,q}} F_q)^2)/(1-\beta_2) \tag{8}$$

$$\theta_{i,k}^{t,q+1} = \theta_{i,k}^{t,q} - \alpha m_q/(\sqrt{v_q}+\varepsilon) \tag{9}$$

where F_q is the loss value of the qth update, and $\nabla_{\theta_{i,k}^{t,q}} F_q$ means the gradient of the loss function with respect to model parameters $\theta_{i,k}^{t,q}$. m_q and v_q are biased-corrected first moment estimate and biased-corrected second moment estimate, respectively. α, β_1 and β_2 are hyperparameters of the Adam algorithm.

After E local epochs, each building i computes the validation loss value of each cluster model, respectively. The loss list containing these values for building i in the tth global communication round is denoted as L_i^t. Then, the building i is assigned to the cluster with the minimum validation loss of L_i^t. Thus the components of each cluster are identified, which are also the participants of the intra-cluster knowledge sharing. The building set for cluster k is denoted as S_k.

In addition to the intra-cluster knowledge sharing, some buildings belonging to other clusters are also selected to optimize the prediction model of the cluster k and implement the inter-cluster knowledge sharing. The inter-cluster knowledge sharing process can fully utilize more local data to participate in the optimization process of cluster models, and meanwhile, the inter-cluster threshold is able to judge whether the inter-cluster sharing should be conducted along with the iterative grouping process. More specifically, it is assumed that building i obtains the minimum loss for the k'-th cluster model. Suppose the loss value of the kth cluster model for building i is lower than the product of the minimum loss and the inter-cluster threshold γ ($\gamma \geq 1$). In that case, building i with the corresponding cluster model is added to the inter-cluster building set S_k' for cluster k. Meanwhile, the ratio of the minimum loss to the loss value of the kth cluster model is computed as the decay factor $\eta_{i,k}^t$ ($1/\gamma \leq \eta_{i,k}^t \leq 1$). The decay factor is adopted as the weight coefficient of inter-cluster knowledge sharing in the following cluster model optimization process.

(2) *Cluster model optimization*

After identifying participants of the intra-cluster and inter-cluster knowledge sharing, the federated learning algorithm is adopted to optimize each cluster model in a distributed manner, and meanwhile, the private data aggregation scheme illustrated in Section 3.2 is also applied to guarantee the privacy of the model aggregation.

Firstly, each building i is required to upload the participating information on the intra-cluster and inter-cluster knowledge sharing. Then, the central server can compute the total number of participating samples for each cluster, namely $\sum_{j\in S_k \cup S_k'} |D_j|$ for cluster k, $k = 1, 2, \ldots, K$. After receiving the broadcasted cluster sample sizes from the server, each building i modifies the original model parameters $\theta_{i,k}^t$ into $\theta_{i,k,\text{intra}}^t$ if it belongs to the cluster k and participates in the corresponding intra-cluster knowledge sharing, and modifies the original model parameters $\theta_{i,k'}^t$ into $\theta_{i,k',\text{inter}}^t$ if it participates in the inter-cluster knowledge sharing for cluster k'. The modified forms can be computed as:

$$\theta_{i,k,\text{intra}}^t = \frac{|D_i|}{\sum_{j\in S_k \cup S_k'} |D_j|}\theta_{i,k}^t \tag{10}$$

$$\theta_{i,k',\text{inter}}^t = \frac{|D_i|}{\sum_{j\in S_{k'} \cup S_{k'}'} |D_j|}\eta_{i,k'}^t \theta_{i,k'}^t \tag{11}$$

where the weights for building i are the proportion of the sample size $|D_i|$ in the total number of participating samples of cluster k and cluster k', respectively. Meanwhile, the decay factor $\eta_{i,k'}^t$ of building i is added to the weight of Eq. (11) to constrain the inter-cluster knowledge sharing process.

In order to avoid possible privacy disclosure caused by directly uploaded model parameters [42,43], all buildings conduct the random vectors agreement to obtain the agreed model parameters for participating clusters. After uploading all encrypted model parameters, the server conducts the model aggregation to optimize the federated cluster models. Specifically, the model optimization for the cluster k in the tth global communication round can be calculated as:

$$\theta_k^{t+1} = \sum_{i\in S_k} (\theta_{i,k,\text{intra}}^t + \omega_{i,k}^t) + \sum_{i\in S_k'} (\theta_{i,k,\text{inter}}^t + \omega_{i,k}^t) \tag{12}$$

where $\omega_{i,k}^t$ is the shared model parameters for building i to encrypt the model that takes part in the knowledge sharing of cluster k. The Eq. (12) averts the potential privacy risks brought by the original model parameters. Meanwhile, the corresponding result is also consistent with the output of the directly model aggregation.

At the end of the last global communication round, each building will receive the cluster model from the server based on the cluster membership. The cluster model is the foundation of establishing the personalized prediction model for each participating or nonparticipating building. In order to further illustrate the effectiveness of the method, theoretical proofs provided in the appendix part indicate there exists an gap between the optimal solution of the clustering federated learning method and the standard federated learning in the data heterogeneity settings.

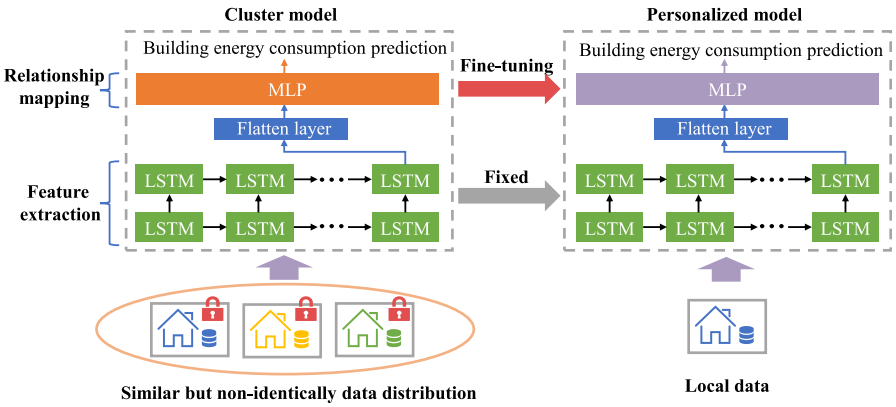


Fig. 4. Personalized federated transfer learning algorithm.

3.3.3. Personalized federated transfer learning

Although the proposed dynamical clustering federated learning algorithm has divided buildings with similar local data distributions into the same group, buildings in each cluster may still have minor but non-negligible differences among their data distributions. Federated learning aiming at establishing a single public prediction model for buildings within a cluster cannot supply customized prediction models to avoid the impacts of these differences according to the similar theoretical proofs in the appendix part. Therefore, on the basis of the model architecture, transfer learning is integrated into the federated building energy prediction framework to establish the personalized prediction model for each building, which is depicted in Fig. 4.

After the dynamic clustering federated learning process, the server obtains final cluster models and distributes them to corresponding buildings belonging to each cluster. In the federated building energy prediction framework, the data owned by different buildings cannot be directly aggregated considering privacy constraints. In other words, the data in the source domain cannot be aggregated directly and utilized to train a prediction model in a centralized manner, which presents a significant difference from existing transfer learning. Thereby, the cluster model that has been trained by buildings with similar local data distributions in the distributed setting is utilized to implement the personalized federated transfer learning due to the prior knowledge implicitly contained in its model parameters.

On the basis of the trained cluster model, a network-based transfer learning strategy is applied to establish the personalized prediction model for each building. Prior studies [51] suggest that similar tasks may share the same representation model, and shallow layers of deep neural networks tend to extract common features of the original data. Therefore, the parameters of the LSTM block for temporal feature extraction are fixed to retain the shared knowledge from other buildings. Meanwhile, the parameters of the MLP block are fine-tuned based on the local data to obtain a building-specific model block and establish a more accurate mapping relationship. In such a case, only part of the model parameters needed to be optimized, and the corresponding requirements for the local data amount and the computing power are not high.

3.4. Knowledge sharing mechanism for nonparticipating buildings

In addition to the participating buildings, a knowledge sharing mechanism for nonparticipating buildings is also developed to improve the scalability of the proposed federated learning framework. Specifically, these nonparticipating buildings do not need to upload any information. Meanwhile, they can still obtain applicable weight initializations provided by the central server and then build customized prediction models. The procedures are described below.

The central server firstly supplies all trained cluster models to nonparticipating buildings, and this process does not involve any raw data exchange. Then, cluster models are independently performed on the local data of each nonparticipating building to obtain the cluster membership. The model with minimum loss refers to the cluster to which the nonparticipating building belongs. Eventually, the personalized federated transfer learning algorithm described in Section 3.3.3 can be directly adopted to establish personalized building energy consumption prediction models for nonparticipating buildings.

4. Experiments and results

In order to comprehensively verify the effectiveness of the proposed privacy-preserving knowledge sharing framework for the few-shot building energy prediction, detailed experiment setups are firstly provided. Then basic performance evaluation of the proposed federated building energy prediction framework is conducted, and the corresponding result discussions are illustrated. The impacts of uneven local datasets on the federated learning approach are analyzed in the final part of this Section.

4.1. Experiment setup

4.1.1. Data description and preprocessing

The building dataset from the Building Data Genome Project 2 (BDGP2) [52] was selected to evaluate the prediction performance of the proposed federated learning approach. The open dataset is mainly composed of the historical energy consumption data of 1636 nonresidential buildings located in multiple cities in 2016 and 2017, and the data sampling interval is one hour. Meanwhile, the dataset also provides the basic information on these buildings and corresponding weather conditions.

In this study, buildings located in Florida, USA, were selected as the research targets. In terms of data cleaning, outliers such as the negative or extremely high energy consumption data were firstly eliminated. Then, buildings with more than continuous 48-hour missing data were also removed from the following experiments. Missing values such as the weather data were finally filled by the linear interpolation method. Finally, 30 buildings were selected to conduct experiments and verify the effectiveness of the proposed method. The primary usage types of selected buildings are office, classroom, and research.

The input variables of the prediction model are listed in Table 1, mainly consisting of weather conditions, calendar attributes, occupancy schedules, building features, and the historical building energy demand. In terms of numerical variables, the proposed private data aggregation scheme was applied to obtain the mean and the standard deviation of the union dataset, and then the z-score normalization technique was adopted to transform them into the normalized format. In terms of categorical variables, recurrent variables such as the calendar attributes were transformed by sine and cosine functions to retain their

Table 1
Input variables of the building energy prediction model.

Variable	Type	Unit/Range	Domain
Air temperature	Numerical	°C	Weather conditions
Dew point temperature	Numerical	°C	
Atmospheric pressure	Numerical	hPa	
Wind speed	Numerical	m/s	
Hour of the day	Categorical	0~23	Calendar attributes
Day of the week	Categorical	1~7	
Day of the month	Categorical	1~31	
Period of the day	Categorical	1~3	Occupancy schedules
Weekend indicator	Categorical	0~1	
Floor area of the building	Numerical	m^2	Building features
Primary usage type of the building	Categorical	1~3	
Historical energy consumption	Numerical	kW	Energy demand

recursive features. Other categorical variables, such as the primary usage type of the building and the period of the day, were processed by the one-hot encoding technique. Note that the period of the day is a categorical variable to indicate the occupancy schedules, where 1 denotes the time period [9PM, 7AM), 2 denotes the time period [7AM, 5PM), and 3 denotes the time period [5PM, 9PM).

In terms of the model output, the building energy consumption of the upcoming hour is selected as the target prediction value. Moreover, target predictions were transformed by the function $\ln(x+1)$ to reduce the skewness and avoid the performance degradation caused by the fluctuation of building energy consumption curves. In terms of the model input, the window length is set as 24 h in Table 1 due to obvious daily patterns of building energy consumption curves and the one-hour ahead energy consumption prediction task. Moreover, building features are constant input variables for each building. In summary, the research goal of this study is to predict one-hour ahead building energy consumption based on the data of historical 24 h.

4.1.2. Evaluation metrics

The coefficient of variation (CV), the mean absolute error (MAE), the mean absolute percentage error (MAPE) and the root mean square error (RMSE) are adopted to evaluate the performance of the building energy prediction model. These four metrics can be computed as:

$$CV = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \Big/ \frac{\sum_{i=1}^{n} y_i}{n} \tag{13}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i| \tag{14}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\hat{y}_i - y_i}{y_i}\right| \tag{15}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \tag{16}$$

where n is the number of samples in the test dataset, $\hat{y}_i$ and y_i are the predicted and the actual building energy consumption, respectively.

4.2. Basic performance evaluation of building energy prediction

In this subsection, 30 buildings possessing the same amount of local data were adopted to evaluate the basic performance of the proposed federated learning method. These buildings are composed of 26 participating buildings and 4 nonparticipating buildings. For each building, 720 samples (30 days) are utilized as the training set to implement the knowledge sharing and establish the personalized prediction model on the premise of privacy protection, 240 samples (10 days) are used as the validation set to tune hyperparameters, and 480 samples (20 days) are adopted as the final test set. The local datasets are all selected from the same time period, namely 2016-07-15 to 2016-09-12.

Firstly, the model architecture provided in Section 3.3.1 was adopted to implement the building energy consumption prediction via the proposed approach, and meanwhile, Adam algorithm was applied to optimize model parameters in the training process. Moreover, the grid search method based on the validation loss is adopted to tune hyperparameters and improve the prediction performance of the proposed approach. The detailed grid search ranges are presented in Table 2. Meanwhile, values for some hyperparameters with constantly poor performance were removed in the grid search process to improve efficiency. Finally, Table 3 provides the hyperparameter configuration of the federated building energy prediction on the even datasets. Note that the early stopping technique is adopted to determine the actual communication rounds of dynamical clustering federated learning and the actual finetuning epoch of the federated transfer learning based on the minimum loss value of the validation set.

The research focus of this paper is to improve the energy consumption prediction performance for buildings with limited data. Therefore, the test set is only set to 20 days (480 samples) due to data scarcity and the generalization ability of prediction models in different seasons cannot be verified. Nevertheless, this paper aims at the short-term building energy consumption prediction and the prediction models can be applicable and effective in a short period for buildings. Meanwhile, the proposed approach can be conducted again based on the current data resources of all buildings to update models when the prediction error is too large.

4.2.1. Experiment results of participating buildings

(1) *Convergence process*

Based on the federated building energy prediction framework and local datasets after preprocessing, the dynamical clustering federated learning algorithm was firstly applied to iteratively cluster buildings in terms of their local data distributions and conduct the intra-cluster and the inter-cluster knowledge sharing. Note that the grid search result of the inter-cluster threshold has illustrated the necessity of inter-cluster knowledge sharing along with the clustering process. Fig. 5 fully presents the convergence processes of different clusters. In each figure, points refer to the training loss values of participating buildings in this communication round, and meanwhile, the curve represents the average training loss of all buildings belonging to this cluster. The inserted histogram shows the building number changes in this cluster during the dynamical grouping process. As shown in Fig. 5, the average training loss of each cluster is basically monotonous decreasing, and the training loss of each building descends rapidly at the first several epochs, indicating the fast convergence rate of the proposed dynamical clustering federated learning algorithm. Moreover, the building number of each cluster is gradually stable along with the iterative clustering process in Fig. 5 and all cluster components remain unchanged in the last 3 training epochs according to experimental results.

In order to further eliminate the impacts of the data heterogeneity phenomenon, the personalized federated transfer learning algorithm

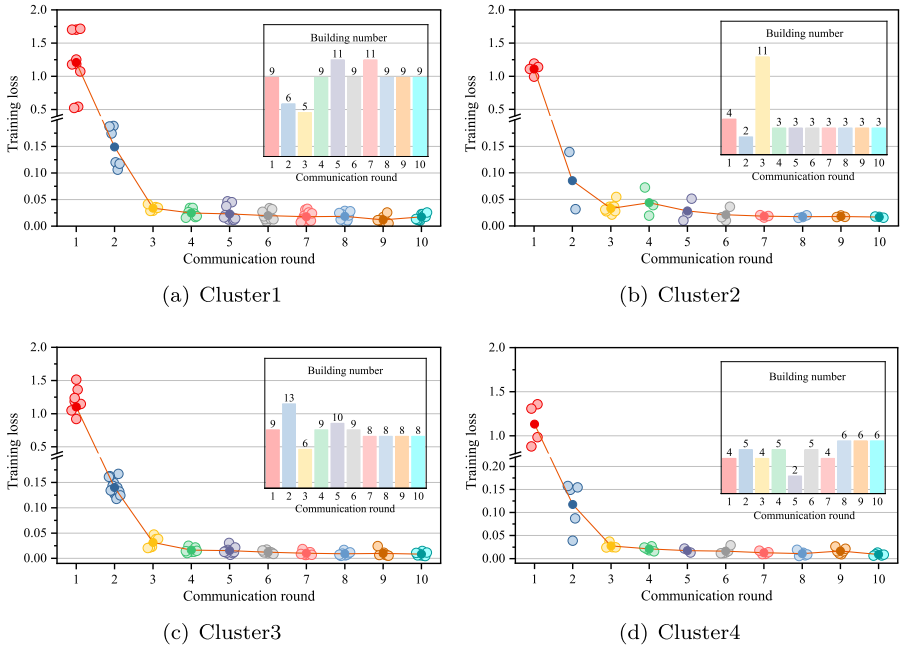


Fig. 5. Convergence of dynamical clustering federated learning algorithm.

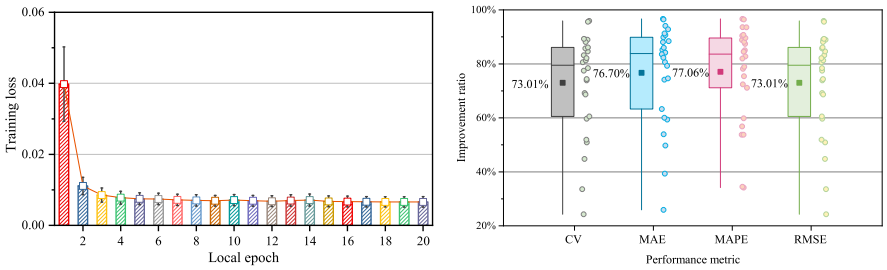
Table 2
Search ranges of hyperparameters for federated building energy prediction approach.

Aspects	Hyperparameters	Grid search values
Model architecture	LSTM layers	2, 3
	MLP layers	1, 2, 3
	LSTM units in each layer	32, 64, 128, 256, 512
	MLP units in each layer	32, 64, 128, 256, 512
Optimizer and regularization	Learning rate	0.001, 0.01, 0.1
	Dropout probability	0.5, 0.6
	L2 regularization	0.0001, 0.0005
Dynamical clustering federated learning	Local epochs	4, 5, 6, 10, 15
	Local batch	32, 64
	Cluster number	3, 4, 5, 6
	Inter-cluster threshold	1.0, 1.1, 1.2
Personalized federated transfer learning	Finetuning batch	32, 64

Table 3
Hyperparameters of federated building energy prediction on even datasets.

Aspects	Hyperparameters	Values
Model architecture	LSTM layer, unit	2, (256, 256)
	MLP layer, unit	2, (256, 1)
Optimizer and regularization	Adam parameters (α, β_1, β_2)	(0.001, 0.9, 0.999)
	Dropout probability	0.5
	L2 regularization	0.0001
Dynamical clustering federated learning	Communication rounds	10 (early stopping)
	Local epochs	5
	Local batch	32
	Cluster number	4
	Inter-cluster threshold	1.2
Personalized federated transfer learning	Finetuning epoch	20 (early stopping)
	Finetuning batch	32

was conducted to realize the customized building energy consumption prediction for each building based on trained cluster models. Fig. 6(a) presents the training process of the personalized federated transfer learning algorithm, the bars with different colors mean the average training loss in each local epoch, and the black lines show the minimum and the maximum local training loss values among all buildings. As observed in Fig. 6(a), the average training loss decreases rapidly in the first few local epochs and then gradually stabilizes at a low loss level. To verify the necessity of establishing personalized building energy consumption prediction models, the prediction performance of customized models and cluster models without personalization were compared. For all participating buildings, Fig. 6(b) depicts the distributions of performance improvement ratios in terms of four evaluation metrics. The individual data point denotes the improvement ratio of each building, and the box limits present the range of the central 50% of all data points. It can be seen that average values of evaluation



(a) Convergence of the training loss (b) Performance improvement ratio

Fig. 6. Convergence and performance improvement of personalized federated transfer learning algorithm.

Table 4
Performance metrics of five different approaches for building energy prediction.

Method	CV	MAE	MAPE	RMSE	Privacy guarantee
Proposed method	0.0694	7.4089	0.0557	9.7071	Yes
Local prediction [5]	0.1121	13.0002	0.0869	16.3856	Yes
Centralized prediction [16]	0.0994	12.0934	0.0816	14.9845	No
Centralized transfer learning 1 [15]	0.0680	7.1066	0.0540	9.2655	No
Centralized transfer learning 2 [15]	0.0685	7.4387	0.0546	9.7146	No

metrics of all buildings are improved by at least 70% on the test set, strongly indicating the effectiveness of the proposed personalized federated transfer learning algorithm.

(2) *Performance comparison*

To verify the effectiveness of the proposed federated learning method, this paper comprehensively compared it with the other four approaches, including the local prediction, the centralized prediction, the centralized transfer learning 1 and the centralized transfer learning 2. The local prediction means adopting the local data to directly train a customized model for each building without any knowledge sharing. On the contrary, the centralized prediction refers to aggregating all local building energy consumption data and establishing a single prediction model for all buildings. Similarly, the centralized transfer learning 1 and 2 also need to aggregate all local data owned by different buildings. Nevertheless, in order to obtain building-specific models, centralized transfer learning 1 and 2 utilize the local data to fine-tune the parameters of the MLP block and all model parameters, respectively. These two approaches are the mainstream methods to predict building energy consumption with limited data in the current studies [6,11,18,41]. Note that all centralized methods adopt the same model architecture illustrated in Table 3, and the local prediction approach utilizes a two-layer LSTM with 32 neurons in each layer as the prediction model to avoid the over-fitting phenomenon caused by the limited available data.

Table 4 provides the performance evaluation results of the proposed approach and the other four contrast methods. All metrics except for the privacy guarantee evaluate the average prediction performance of all participating buildings. Fig. 7 presents the performance metrics of each participating building based on the above five methods. As shown in Table 4 and Fig. 7, the proposed federated learning method witnessed the comparable prediction performance with centralized transfer learning 1 and 2, and meanwhile, it performed better than the local prediction approach on almost every building. Moreover, it implemented privacy protection for the sensitive building energy consumption data owned by different buildings. For the centralized transfer learning 1 and 2, although they implement satisfactory prediction performance for participating buildings, they depend on aggregating the local building energy consumption data to pretrain a transferable prediction model in a centralized manner, which inevitably leaks the occupancy information and may lead to privacy issues. In terms of the local prediction and the centralized prediction, they both gained relatively poor prediction performance on all participating buildings. For the poor performance causes, local prediction suffers from the limited amount of available training data, leading to the inaccurate mapping relationship of the prediction model. Centralized prediction only establishes a single model for all participating buildings and cannot take the data heterogeneity phenomenon into consideration compared with centralized transfer learning 1 and 2. Meanwhile, it also makes compromises on the building occupancy privacy and has potential privacy risks.

Table 5
Performance comparison with local prediction for nonparticipating buildings.

Approach	Building	CV	MAE	MAPE	RMSE
Proposed method	Building1	0.1261	2.3782	0.0992	3.2175
	Building2	0.1774	10.8012	0.1577	15.4955
	Building3	0.0848	3.2479	0.0631	4.3016
	Building4	0.0723	2.9347	0.0586	3.6693
Local prediction [5]	Building1	0.2306	4.4883	0.1893	5.8805
	Building2	0.1927	12.2608	0.1712	16.8350
	Building3	0.1358	5.0422	0.0919	6.8872
	Building4	0.0782	3.1576	0.0636	3.9709

As observed in Fig. 7, all buildings except for building 2 obtained better prediction performance via the proposed approach compared with the local prediction approach. For building 2, the evaluation metric CV of the proposed approach and the local prediction approach are 0.0594 and 0.0580, respectively. The performance gap between the two approaches is only 0.0014, demonstrating that the local data distribution of building 2 may significantly differ from other buildings and the transferred knowledge even have few negative impacts on its performance improvement. Therefore, building 2 obtains the worse prediction performance via the proposed approach, but the performance degradation is small enough to be ignored.

4.2.2. *Experiment results of nonparticipating buildings*

The scalable federated building energy prediction framework can implement the knowledge sharing process for nonparticipating buildings by providing weight initializations and establishing personalized prediction models. Fig. 8 shows the prediction and actual energy consumption curves of the randomly selected 4 nonparticipating buildings. Table 5 provides prediction performance evaluation metrics of each building based on the proposed method learning and local prediction, respectively.

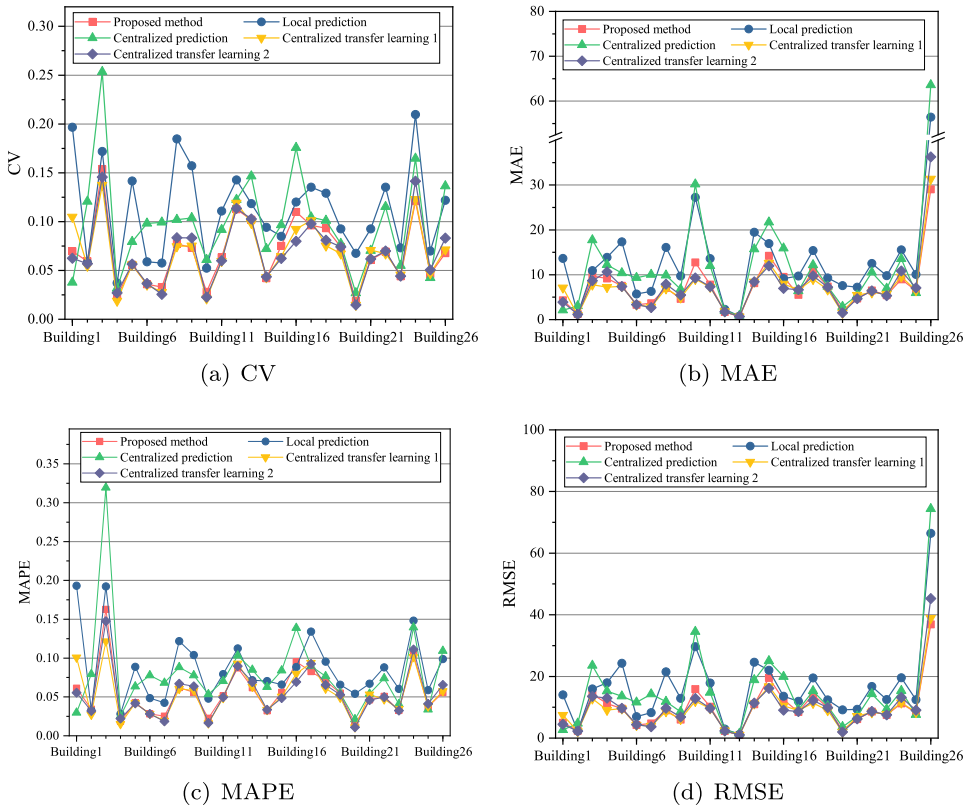


Fig. 7. Performance comparison of five different approaches for each building.

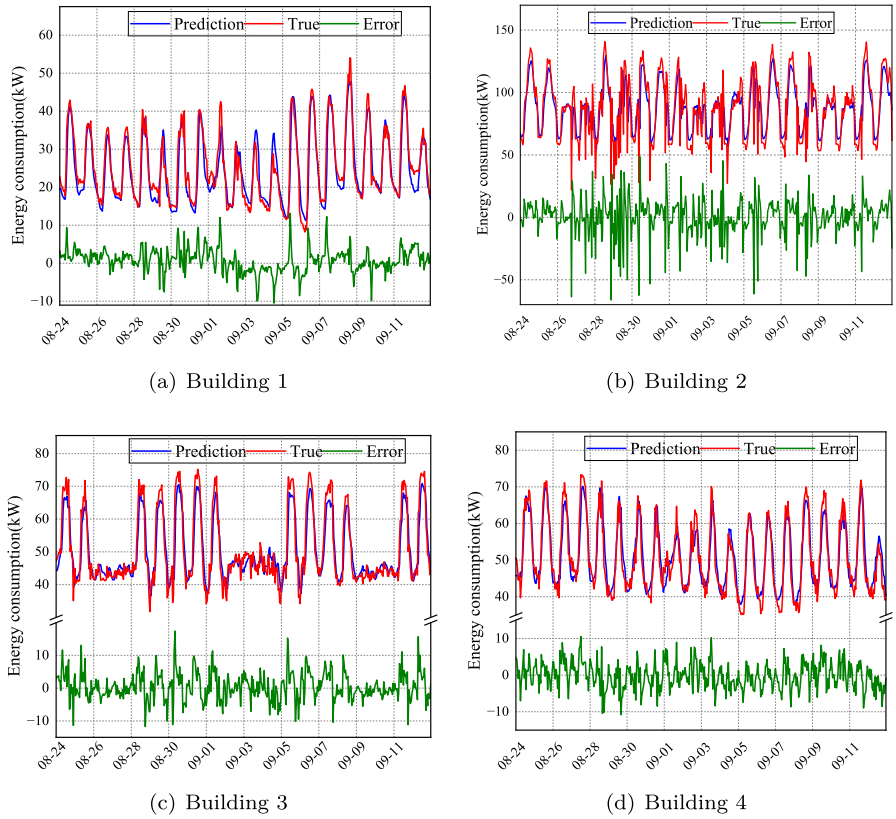


Fig. 8. The prediction and actual energy consumption curves of nonparticipating buildings.

Table 6
Local data amounts of different buildings.

Groups	Training set	Validation set	Test set
Building 1–5	480 (20 days)	240 (10 days)	480 (20 days)
Building 6–10	600 (25 days)	240 (10 days)	480 (20 days)
Building 11–15	720 (30 days)	240 (10 days)	480 (20 days)
Building 16–20	840 (35 days)	240 (10 days)	480 (20 days)
Building 21–26	960 (40 days)	240 (10 days)	480 (20 days)

As shown in Fig. 8, the prediction curve presents a similar energy consumption pattern as the actual curve for each building. Moreover, Table 5 indicates that the prediction performance of building 1, 2, and 3 has witnessed the relatively high improvement based on provided weight initializations. It also reveals the effectiveness of the proposed federated building energy prediction framework. However, the prediction accuracy improvement of building 4 is relatively low, possibly because the data distribution of building 4 presents relatively large differences from the data distribution of each cluster. Thereby, the trained cluster models are not quite appropriate weight initializations for building 4. A practical solution for it is to establish an effective incentive mechanism further and attract more buildings to participate in the federated building energy prediction framework.

4.3. Performance evaluation of building energy prediction with uneven datasets

For the proposed federated building energy prediction framework, the participating buildings may possess different amounts of local datasets. Meanwhile, the uneven data distribution phenomenon may degrade the prediction performance of aggregated cluster models and the following personalized models [53,54]. Therefore, 26 buildings possessing different amounts of local datasets were adopted to evaluate the prediction performance under the impacts of uneven local data amounts. The local datasets are randomly selected from 2016-07-15 to 2016-10-15 and the specific amounts of local data samples owned by different buildings are presented in Table 6. Table 7 shows the corresponding hyperparameter configuration of the federated building energy prediction on uneven datasets.

Although the total number of selected samples with uneven distribution equals that of even datasets in the previous experiment, buildings possess different amounts of local datasets, and the corresponding participating local samples are obviously different. Therefore, it is unreasonable to compare the evaluation metrics with those of the previous experiment. Existing methods consisting of the local prediction, the centralized prediction, the centralized transfer learning 1 and 2 are adopted again to validate the effectiveness of the proposed method. Table 8 provides average metric values of all participating buildings obtained by five different approaches, and Fig. 9 depicts specific CV metric values of all buildings with different amounts of local samples.

As observed in Table 8 and Fig. 9, the proposed federated learning approach still witnessed comparable performance to the centralized transfer learning 1 and 2 on uneven datasets. Meanwhile, the centralized prediction and the local prediction methods still received relatively poor performance. However, as observed in Fig. 9, building 2 obtained worse prediction performance by the proposed approach again compared with the local prediction approach, and the CV gap of these two approaches is 0.007. Therefore, more buildings are desirable to participate in the framework and train prediction models cooperatively for the mitigation of data heterogeneity. In summary, the proposed federated learning approach has taken the uneven data distribution phenomenon into consideration by weighted model aggregation, and this experiment validates that uneven datasets have limited effects on the prediction performance. Furthermore, the results also indicate that the proposed approach still has comparable performance to existing methods on local datasets belonging to different time periods.

5. Conclusions

To improve the few-shot building energy prediction performance and tackle the privacy issues of existing methods, this paper proposes a privacy-preserving knowledge sharing framework based on federated learning to implement the cooperation among multiple buildings with insufficient data while preserving the privacy. Extensive experiments are conducted on the benchmark dataset BDGP2 to evaluate the one-hour ahead building energy predictions in few-shot learning scenarios. In terms of participating buildings, the results of the convergence process analysis and performance comparison fully show that the federated approach outperforms the other three conventional approaches in both prediction accuracy and privacy protection under intrinsic data heterogeneity. In terms of nonparticipating buildings, the experiment results also validate the scalability of the federated approach. Moreover, the uneven local datasets owned by different buildings have limited influence on the prediction performance. To summarize, the proposed federated learning approach witnesses desirable prediction performance while preserving privacy during the knowledge sharing process of multiple buildings.

This paper proposes a novel federated approach to tackle the few-shot building energy prediction problem in a distributed manner. Meanwhile, the federated building energy prediction method can be extended into a probabilistic framework by integrating the quantile regression method to support the uncertainty-aware energy management of buildings. Future work will focus on quantifying the value of datasets and designing a practicable incentive mechanism to motivate more buildings to participate in the privacy-preserving knowledge sharing framework and thoroughly investigate the potential of improving the multi-horizon prediction performance by this federated approach.

CRediT authorship contribution statement

Lingfeng Tang: Conceptualization, Methodology, Writing – original draft. **Haipeng Xie:** Validation, Writing – review & editing. **Xiaoyang Wang:** Investigation, Visualization, Software. **Zhaohong Bie:** Supervision.

Data availability

Building Data Genome Project 2 (BDGP2) dataset that supports this article is available at https://www.kaggle.com/datasets/claytonmiller/buildingdatagenomeproject2.

Acknowledgments

This work was supported by the National Key Research and Development Program of China (2021YFB2401300).

Appendix

Proposition. *The optimal solution of the clustering federated method is better than that of the standard federated learning method in statistical heterogeneity settings.*

Proof. The optimization objective of the standard horizontal federated learning is to minimize the error function F, namely the sum of the expected loss values over local data distributions. The corresponding mathematic form can be presented as:

$$\min_{\theta} F(\theta) = \min_{\theta} \left\{ \frac{1}{N} \sum_{i=1}^{N} \mathbb{E}_{(\boldsymbol{x},y)\sim D_i}[l(f(\boldsymbol{x};\theta), y)] \right\} \tag{17}$$

where N is the number of participating buildings, D_i denotes the local data distribution of the ith building, θ denotes the federated

Table 7
Hyperparameters of federated building energy prediction on uneven datasets.

Aspects	Hyperparameters	Values
Model architecture	LSTM layer, unit	2, (256, 256)
	MLP layer, unit	2, (256, 1)
Optimizer and regularization	Adam parameters (α, β_1, β_2)	(0.001, 0.9, 0.999)
	Dropout probability	0.5
	L2 regularization	0.0001
Dynamical clustering federated learning	Communication rounds	10 (early stopping)
	Local epochs	4
	Local batch	32
	Cluster number	4
	Inter-cluster threshold	1.1
Personalized federated transfer learning	Finetuning epoch	30 (early stopping)
	Finetuning batch	32

Table 8
Performance evaluation of five different approaches on uneven local datasets.

Method	CV	MAE	MAPE	RMSE	Privacy guarantee
Proposed method	0.0720	7.8210	0.0612	10.0917	Yes
Local prediction [5]	0.1074	11.9788	0.0873	15.2631	Yes
Centralized prediction [16]	0.0941	11.2437	0.0820	14.0580	No
Centralized transfer learning 1 [15]	0.0735	7.9041	0.0632	10.2070	No
Centralized transfer learning 2 [15]	0.0700	7.8086	0.0567	10.1107	No

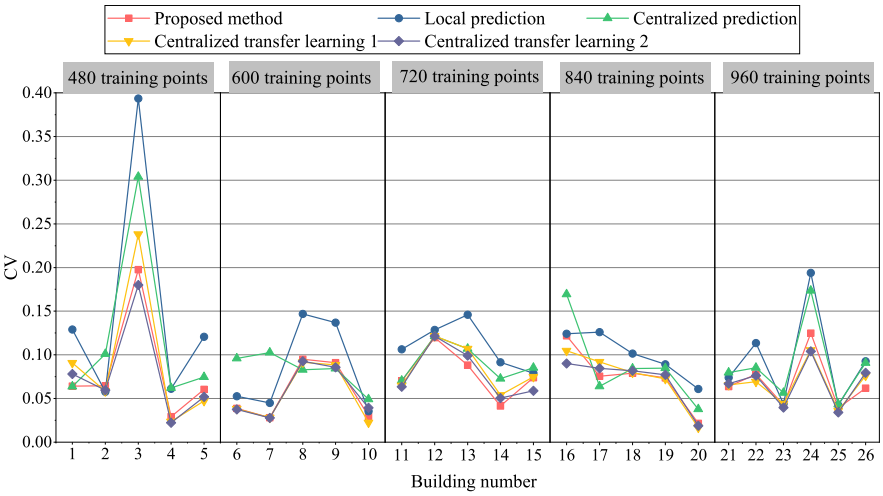


Fig. 9. CV metric comparison with other approaches for each building.

learning model, l and f are the loss function and the prediction model, respectively.

Compared with the single shared prediction model in the standard federated learning, the clustering federated learning method groups participating buildings into K clusters, and buildings within the same cluster adopt the shared prediction model θ_k $(k = 1, 2, \ldots, K)$. For the ith building, the prediction model with minimum loss is selected from K cluster models in the clustering federated learning method. The corresponding form can be presented as:

$$k = \arg \min_{k=1,\ldots,K} \mathbb{E}_{(\boldsymbol{x},y)\sim D_i}[l(f(\boldsymbol{x};\theta_k), y)] \tag{18}$$

Therefore, buildings in the same cluster have relatively similar local data distributions, namely $D_i \approx D_k$ $(i \in S_k)$, and S_k denotes the buildings belonging to the cluster k. The corresponding optimization objective of the clustering federated learning method can be formed as:

$$\begin{aligned} \min_{\theta_1,\ldots,\theta_K} F(\theta_1, \ldots, \theta_K) &= \min_{\theta_1,\ldots,\theta_K} \left\{ \frac{1}{N} \sum_{i=1}^{N} \mathbb{E}_{(\boldsymbol{x},y)\sim D_i}[l(f(\boldsymbol{x};\theta_i), y)] \right\} \\ &= \min_{\theta_1,\ldots,\theta_K} \left\{ \frac{1}{N} \sum_{k=1}^{K} \mathbb{E}_{(\boldsymbol{x},y)\sim D_k}[l(f(\boldsymbol{x};\theta_k), y)] \right\} \end{aligned} \tag{19}$$

In order to demonstrate the gap more clearly, the prediction model f is set as a linear regression function with zero bias, namely $f(x) = Ax$, and the loss function l is set as the standard mean squared loss function. Meanwhile, the local data distributions of buildings may be not identical. Thereby, the optimal model is denoted as A_i^* for the ith building. To sum up, the optimization objective of the standard federated learning and the clustering federated learning method can be reformulated as the following equations, respectively.

$$\min_{\boldsymbol{A}} F(\boldsymbol{A}) = \min_{\boldsymbol{A}} \left\{ \frac{1}{2N} \sum_{i=1}^{N} \|\boldsymbol{A}\boldsymbol{x}_i - \boldsymbol{A}_i^* \boldsymbol{x}_i\|_2^2 \right\} \tag{20}$$

$$\min_{\boldsymbol{A}_1 \sim \boldsymbol{A}_K} F(\boldsymbol{A}_1,\ldots,\boldsymbol{A}_K) = \min_{\boldsymbol{A}_1 \sim \boldsymbol{A}_K} \left\{ \frac{1}{2N} \sum_{i=1}^{N} \|\boldsymbol{A}_k \boldsymbol{x}_i - \boldsymbol{A}_i^* \boldsymbol{x}_i\|_2^2 \right\} \tag{21}$$

From Eqs. (20) and (21), it can be clearly observed that the optimization objective of the standard federated learning is to minimize the average loss of samples over local data distributions by a single model. On the contrary, the clustering federated learning method groups buildings with similar data distributions into the same cluster, namely $A_i^* \approx A_k^* (i \in S_k)$ for cluster k. Meanwhile, the optimization objective of the clustering federated learning method is to minimize the average loss of K clusters by K models.

Therefore, there exists a gap between the optimal solution of the clustering federated learning method and the standard federated learning in the data heterogeneity settings, namely:

$$\min_{\boldsymbol{A}} F(\boldsymbol{A}) \geq \min_{\boldsymbol{A}_1,\ldots,\boldsymbol{A}_K} F(\boldsymbol{A}_1,\ldots,\boldsymbol{A}_K) \tag{22}$$

Remark. The optimization process of deep learning is non-convex. Federated learning without a clustering process may not be able to provide appropriate initial parameters for transfer learning in the statistical heterogeneity scenarios. Moreover, the clustering federated learning method with only one cluster is exactly the standard federated learning. Therefore, the clustering method with the number of clusters as a hyperparameter is necessary to be adopted to obtain a relatively better weight initialization.

References

[1] The global alliance for buildings and construction (GlobalABC). 2021 Global Status Report for Buildings and Construction, 2021, https://globalabc.org/resources/publications/2021-global-status-report-buildings-and-construction.

[2] Zhang L, Wen J, Li Y, Chen J, Ye Y, Fu Y, et al. A review of machine learning in building load prediction. Appl Energy 2021;285:116452. http://dx.doi.org/10.1016/j.apenergy.2021.116452.

[3] Wei Y, Zhang X, Shi Y, Xia L, Pan S, Wu J, et al. A review of data-driven approaches for prediction and classification of building energy consumption. Renew Sustain Energy Rev 2018;82:1027–47. http://dx.doi.org/10.1016/j.rser.2017.09.108.

[4] Amasyali K, El-Gohary NM. A review of data-driven building energy consumption prediction studies. Renew Sustain Energy Rev 2018;81:1192–205. http://dx.doi.org/10.1016/j.rser.2017.04.095.

[5] Deb C, Zhang F, Yang J, Lee SE, Shah KW. A review on time series forecasting techniques for building energy consumption. Renew Sustain Energy Rev 2017;74:902–24. http://dx.doi.org/10.1016/j.rser.2017.02.085.

[6] Pinto G, Wang Z, Roy A, Hong T, Capozzoli A. Transfer learning for smart buildings: A critical review of algorithms, applications, and future perspectives. Adv Appl Energy 2022;5:100084. http://dx.doi.org/10.1016/j.adapen.2022.100084.

[7] Qian F, Gao W, Yang Y, Yu D. Potential analysis of the transfer learning model in short and medium-term forecasting of building HVAC energy consumption. Energy 2020;193:116724. http://dx.doi.org/10.1016/j.energy.2019.116724.

[8] Feng Y, Duan Q, Chen X, Yakkali SS, Wang J. Space cooling energy usage prediction based on utility data for residential buildings using machine learning methods. Appl Energy 2021;291:116814. http://dx.doi.org/10.1016/j.apenergy.2021.116814.

[9] Wang Z, Wang Y, Zeng R, Srinivasan RS, Ahrentzen S. Random forest based hourly building energy prediction. Energy Build 2018;171:11–25. http://dx.doi.org/10.1016/j.enbuild.2018.04.008.

[10] Cybenko G. Approximation by superpositions of a sigmoidal function. Math Control Signals Systems 1989;2(4):303–14. http://dx.doi.org/10.1007/BF02551274.

[11] Somu N, M R GR, Ramamritham K. A hybrid model for building energy consumption forecasting using long short term memory networks. Appl Energy 2020;261:114131. http://dx.doi.org/10.1016/j.apenergy.2019.114131.

[12] Ding Z, Chen W, Hu T, Xu X. Evolutionary double attention-based long short-term memory model for building energy prediction: Case study of a green building. Appl Energy 2021;288:116660. http://dx.doi.org/10.1016/j.apenergy.2021.116660.

[13] Li A, Xiao F, Zhang C, Fan C. Attention-based interpretable neural network for building cooling load prediction. Appl Energy 2021;299:117238. http://dx.doi.org/10.1016/j.apenergy.2021.117238.

[14] Kim T-Y, Cho S-B. Predicting residential energy consumption using CNN-LSTM neural networks. Energy 2019;182:72–81. http://dx.doi.org/10.1016/j.energy.2019.05.230.

[15] Fan C, Sun Y, Xiao F, Ma J, Lee D, Wang J, et al. Statistical investigations of transfer learning-based methodology for short-term building energy predictions. Appl Energy 2020;262:114499. http://dx.doi.org/10.1016/j.apenergy.2020.114499.

[16] Fang X, Gong G, Li G, Chun L, Li W, Peng P. A hybrid deep transfer learning strategy for short term cross-building energy prediction. Energy 2021;215:119208. http://dx.doi.org/10.1016/j.energy.2020.119208.

[17] Gao Y, Ruan Y, Fang C, Yin S. Deep learning and transfer learning models of energy consumption forecasting for a building with poor information data. Energy Build 2020;223:110156. http://dx.doi.org/10.1016/j.enbuild.2020.110156.

[18] Somu N, Sriram A, Kowli A, Ramamritham K. A hybrid deep transfer learning strategy for thermal comfort prediction in buildings. Build Environ 2021;204:108133. http://dx.doi.org/10.1016/j.buildenv.2021.108133.

[19] Zhuang F, Qi Z, Duan K, Xi D, Zhu Y, Zhu H, et al. A comprehensive survey on transfer learning. Proc IEEE 2021;109(1):43–76. http://dx.doi.org/10.1109/jproc.2020.3004555.

[20] Voigt P, von dem Bussche A. The EU general data protection regulation (GDPR): A practical guide. 1st ed.. Cham: Springer; 2017, http://dx.doi.org/10.1007/978-3-319-57959-7.

[21] State of California department of justice. 2021, California Consumer Privacy Act (CCPA), https://www.oag.ca.gov/privacy/ccpa.

[22] Kairouz P, McMahan HB, Avent B, Bellet A, Bennis M, Bhagoji AN, et al. Advances and open problems in federated learning. Found Trends® Mach Learn 2021;14(1–2):1–210. http://dx.doi.org/10.1561/2200000083.

[23] Mothukuri V, Parizi RM, Pouriyeh S, Huang Y, Dehghantanha A, Srivastava G. A survey on security and privacy of federated learning. Future Gener Comput Syst 2021;115:619–40. http://dx.doi.org/10.1016/j.future.2020.10.007.

[24] Tan AZ, Yu H, Cui L, Yang Q. Toward personalized federated learning. IEEE Trans Neural Netw Learn Syst 2022;1–17. http://dx.doi.org/10.1109/tnnls.2022.3160699.

[25] Taïk A, Cherkaoui S. Electrical load forecasting using edge computing and federated learning. In: 2020 IEEE international conference on communications. Dublin, Ireland; 2020, p. 1–6. http://dx.doi.org/10.1109/ICC40277.2020.9148937.

[26] Wang Y, Gao N, Hug G. Personalized federated learning for individual consumer load forecasting. CSEE J Power Energy Syst 2022;1–5. http://dx.doi.org/10.17775/CSEEJPES.2021.07350.

[27] Fang X, Gong G, Li G, Chun L, Peng P, Li W. A general multi-source ensemble transfer learning framework integrate of LSTM-DANN and similarity metric for building energy prediction. Energy Build 2021;252:111435. http://dx.doi.org/10.1016/j.enbuild.2021.111435.

[28] Ahn Y, Kim BS. Prediction of building power consumption using transfer learning-based reference building and simulation dataset. Energy Build 2022;258:111717. http://dx.doi.org/10.1016/j.enbuild.2021.111717.

[29] Li L, Fan Y, Tse M, Lin K-Y. A review of applications in federated learning. Comput Ind Eng 2020;149:106854. http://dx.doi.org/10.1016/j.cie.2020.106854.

[30] McMahan B, Moore E, Ramage D, Hampson S, Arcas BAy. Communication-efficient learning of deep networks from decentralized data. In: Proceedings of the 20th international conference on artificial intelligence and statistics, Proceedings of machine learning research, Fort Lauderdale, Florida, USA. 2016, p. 1273–82.

[31] Wang X, Han Y, Wang C, Zhao Q, Chen X, Chen M. In-edge AI: Intelligentizing mobile edge computing, caching and communication by federated learning. IEEE Netw 2019;33(5):156–65. http://dx.doi.org/10.1109/mnet.2019.1800286.

[32] Li J, Yang Z, Wang X, Xia Y, Ni S. Task offloading mechanism based on federated reinforcement learning in mobile edge computing. Digit Commun Netw 2022. http://dx.doi.org/10.1016/j.dcan.2022.04.006.

[33] Zheng W, Yan L, Gou C, Wang F-Y. Federated meta-learning for fraudulent credit card detection. In: Proceedings of the twenty-ninth international joint conference on artificial intelligence (IJCAI), International joint conferences on artificial intelligence organization, Vienna, Austria. 2020, p. 4654–60. http://dx.doi.org/10.24963/ijcai.2020/642.

[34] Elayan H, Aloqaily M, Guizani M. Sustainability of healthcare data analysis iot-based systems using deep federated learning. IEEE Internet Things J 2022;9(10):7338–46. http://dx.doi.org/10.1109/jiot.2021.3103635.

[35] Wu Q, Chen X, Zhou Z, Zhang J. Fedhome: Cloud-edge based personalized federated learning for in-home health monitoring. IEEE Trans Mob Comput 2022;21(8):2818–32. http://dx.doi.org/10.1109/tmc.2020.3045266.

[36] Wang H, Kaplan Z, Niu D, Li B. Optimizing federated learning on non-iid data with reinforcement learning. In: 2020 IEEE conference on computer communications, Virtual. 2020, http://dx.doi.org/10.1109/infocom41043.2020.9155494.

[37] Sattler F, Muller K-R, Samek W. Clustered federated learning: Model-agnostic distributed multitask optimization under privacy constraints. IEEE Trans Neural Netw Learn Syst 2021;32(8):3710–22. http://dx.doi.org/10.1109/tnnls.2020.3015958.

[38] He Y, Chen Y, Yang X, Yu H, Huang YH, Gu Y. Learning critically: Selective self-distillation in federated learning on non-iid data. IEEE Trans Big Data 2022;1–12. http://dx.doi.org/10.1109/TBDATA.2022.3189703.

[39] Yang H, He H, Zhang W, Cao X. Fedsteg: A federated transfer learning framework for secure image steganalysis. IEEE Trans Netw Sci Eng 2021;8(2):1084–94. http://dx.doi.org/10.1109/tnse.2020.2996612.

[40] Bhargava R, Clifton C. When is a semi-honest secure multiparty computation valuable? Springer International Publishing; 2019, p. 45–64. http://dx.doi.org/10.1007/978-3-030-32430-8_4.

[41] Park H, Park DY, Noh B, Chang S. Stacking deep transfer learning for short-term cross building energy prediction with different seasonality and occupant schedule. Build Environ 2022;109060. http://dx.doi.org/10.1016/j.buildenv.2022.109060.

[42] Zhu L, Liu Z, Han S. Deep leakage from gradients. Vol. 32. Cham, Switzerland: Springer International Publishing; 2020, http://dx.doi.org/10.1007/978-3-030-63076-8_2.

[43] Wen M, Xie R, Lu K, Wang L, Zhang K. FedDetect: A novel privacy-preserving federated learning framework for energy theft detection in smart grid. IEEE Internet Things J 2022;9(8):6069–80. http://dx.doi.org/10.1109/jiot.2021.3110784.

[44] Goryczka S, Xiong L. A comprehensive comparison of multiparty secure additions with differential privacy. IEEE Trans Dependable Secure Comput 2017;14(5):463–77. http://dx.doi.org/10.1109/TDSC.2015.2484326.

[45] Diffie W, Hellman M. New directions in cryptography. IEEE Trans Inform Theory 1976;22(6):644–54. http://dx.doi.org/10.1109/TIT.1976.1055638.

[46] Galbraith SD, Gaudry P. Recent progress on the elliptic curve discrete logarithm problem. Des Codes Cryptogr 2016;78(1):51–72. http://dx.doi.org/10.1007/s10623-015-0146-7.

[47] Greff K, Srivastava RK, Koutnik J, Steunebrink BR, Schmidhuber J. LSTM: A search space odyssey. IEEE Trans Neural Netw Learn Syst 2017;28(10):2222–32. http://dx.doi.org/10.1109/tnnls.2016.2582924.

[48] Xiang Y, Hong J, Yang Z, Wang Y, Huang Y, Zhang X, et al. Slope-based shape cluster method for smart metering load profiles. IEEE Trans Smart Grid 2020;11(2):1809–11. http://dx.doi.org/10.1109/TSG.2020.2965801.

[49] Stanley KO, Miikkulainen R. Evolving neural networks through augmenting topologies. Evol Comput 2002;10(2):99–127. http://dx.doi.org/10.1162/106365602320169811.

[50] Ghosh A, Chung J, Yin D, Ramchandran K. An efficient framework for clustered federated learning. IEEE Trans Inform Theory 2022;1. http://dx.doi.org/10.1109/TIT.2022.3192506.

[51] Guo Y, Shi H, Kumar A, Grauman K, Rosing T, Feris R. Spottune: Transfer learning through adaptive fine-tuning. In: 2019 IEEE/CVF conference on computer vision and pattern recognition. California, USA: IEEE; 2019, p. 4800–9. http://dx.doi.org/10.1109/cvpr.2019.00494.

[52] Miller C, Kathirgamanathan A, Picchetti B, Arjunan P, Park JY, Nagy Z, et al. The building data genome project 2, energy meter data from the ashrae great energy predictor III competition. Sci Data 2020;7(1):368. http://dx.doi.org/10.1038/s41597-020-00712-x.

[53] Wang Y, Bennani IL, Liu X, Sun M, Zhou Y. Electricity consumer characteristics identification: A federated learning approach. IEEE Trans Smart Grid 2021;12(4):3637–47. http://dx.doi.org/10.1109/tsg.2021.3066577.

[54] Wang Y, Jia M, Gao N, Von Krannichfeldt L, Sun M, Hug G. Federated clustering for electricity consumption pattern extraction. IEEE Trans Smart Grid 2022;13(3):2425–39. http://dx.doi.org/10.1109/tsg.2022.3146489.