Federated Learning for Short-term Load Forecasting

Zhaorui Meng, Xiaozhu Xie, Yanqi Xie, Jinhua Sun

Abstract—**Deep learning architectures have exhibited robust performance in short-term load forecasting tasks, contingent upon access to substantial training datasets. However, the acquisition of such datasets presents significant challenges, as load operators may be reluctant to share their data due to privacy protection policies. Consequently, the accurate prediction of loads becomes increasingly complex. This study proposes a novel approach to short-term load forecasting, leveraging federated learning and local tuning techniques. The proposed two-stage methodology commences with the training of a generic model through decentralized learning and central aggregation, aimed at preserving data privacy. Subsequently, the resulting generic model undergoes local fine-tuning on each operator's specific load data for enhanced forecasting accuracy. Empirical evaluations conducted on multiple regional load datasets from the GEFCom2012 and AEMO datasets demonstrate the trained model's capacity to improve short-term load forecasting accuracy while maintaining the confidentiality of individual operators' data.**

Index Terms—Load forecasting, federated learning, local tuning, data privacy

I. Introduction

Short-term load forecasting (STLF) is an important task in energy systems that involves predicting the future electricity demand of a particular area over a short period, typically ranging from a few hours to a few days. Accurate STLF is crucial for efficient energy generation, transmission, and distribution planning [1].

However, obtaining accurate STLF predictions is challenging due to the complex and dynamic nature of the electricity demand. Machine learning algorithms have become a popular choice for STLF, as they can handle large volumes of data and learn complex relationships between various factors affecting load demand. Niu et al. [2] introduced a novel approach for STLF, which utilizes a CNN-BiGRU model enhanced by an attention mechanism. The proposed hybrid model exhibits superior accuracy when compared to the conventional Long Short-Term Memory (LSTM) model. Ma et al. [3] introduced a hybrid model for STLF, utilizing a multi-trait-driven approach in conjunction with secondary decomposition. The effectiveness of the model was evaluated through multi-step-ahead forecasting and demonstrated superior performance compared to all benchmark models. Similarly, Saeed et al. [4] proposed a hybrid CNN-LSTM model specifically designed for load forecasting in smart grids. The researchers conducted extensive experimentation on electricity load data, resulting in remarkable outcomes. The findings showcase the model's exceptional accuracy, achieving 98.3% accuracy and a 0.4560 MAPE error. Jeong et al. [5] introduced a novel day-ahead electric load forecasting model specifically designed for buildings. The proposed model, named the logistic mixture vector autoregressive model (LMVAR), exhibits superior performance when compared to benchmark methods. Liu et al. [6] introduced a pioneering model known as the Long-Term and Short-Term Time Series Network for load prediction. This model incorporates a convolutional layer to effectively capture the short-term attributes and interdependencies among variables. Syed et al. [7] introduced a hybrid STLF model utilizing clustering-based deep learning methodology. The primary objective of the study is to enhance scalability. The research endeavors to evaluate the influence of this approach on both training time and model accuracy. In a similar vein, Pham et al. [8] proposed a hybrid method for short-term load demand forecasting that amalgamates the Singular Spectrum Analysis (SSA) technique with LSTM techniques. Song et al. [9] introduced the VMD-Prophet-Seq2seq model for STLF. This model incorporates the Variational Mode Decomposition (VMD) method to decompose the original power load sequence into multiple integral mode functions (IMFs). Additionally, the Prophet method is utilized to decompose the original power load series into subsequences, facilitating multistep prediction on single-feature original data. Experimental findings support the superior performance of the proposed VMD-Prophet-Seq2seq model, demonstrated by a lower mean absolute percentage error (MAPE) compared to the suboptimal model.

Manuscript received July 3, 2023, revised October 22, 2024.

Zhaorui Meng is a senior engineer in School of Computer and Information Engineering, Xiamen University of Technology, Xiamen, Fujian,361024, China (Corresponding author; Email: mengzhaorui@ xmut.edu.cn).

Xiaozhu Xie is a lecturer in School of Computer and Information Engineering, Xiamen University of Technology, Xiamen, Fujian,361024, China (Email: xiaozhuxie@ xmut.edu.cn).

Yanqi Xie is a lecturer in School of Computer and Information Engineering, Xiamen University of Technology, Xiamen, Fujian,361024, China (Email: yqxie@ xmut.edu.cn).

Jinhua Sun is an associate professor in School of Computer and Information Engineering, Xiamen University of Technology, Xiamen, Fujian,361024, China (Email: jinhuasun@ xmut.edu.cn).

As discussed above, the application of machine learning techniques in the field of STLF has shown promising results. However, in real-world scenarios, extant methods are beset with certain issues. These issues are elucidated below. First, current machine learning-based load forecasting methods rely on having adequate training data sets with a significant number of samples. However, practitioners face challenges in meeting this condition during actual load forecasting work. Secondly, the potential to enhance model prediction capability exists by increasing the number of training samples through data sharing between regions. However, high data confidentiality requirements of individual power companies impose restrictions on data sharing for model training, and related privacy and confidentiality issues need to be addressed. Lastly, it is widely recognized that models trained via conventional machine learning approaches demonstrate inadequate generalizability, thereby requiring separate training on distinct datasets.

Consequently, there is a pressing demand for a method that permits collaborative construction of load forecasting models while ensuring the confidentiality of load and other data. To meet this demand, this paper proposes a federated learning approach.

Federated learning [10] represents a contemporary approach in the realm of machine learning, which facilitates collaborative training of a machine learning model among multiple entities, all while preserving the privacy of their respective data. Federated learning provides a promising solution to the challenges of STLF by allowing various stakeholders to contribute their data to a central machine learning model without sharing the data itself. This approach ensures data privacy and security while still enabling the model to learn from the collective data of all participants.

Yan et al. [11] presented an application scheme for federated learning in financial credit risk management. The authors conducted an empirical investigation, comparing and analyzing their proposed approach. The findings demonstrated a substantial 14% enhancement in the performance of the application framework and methodology, particularly for small banks with restricted data samples.

Another recent study by Perifanis et al. [12] aims to investigate the efficacy of federated learning applied to raw base station-aggregated Long-Term Evolution (LTE) data for time-series forecasting. Specifically, the authors assess the performance of five distinct neural network architectures in generating one-step predictions. The proposed models were trained within a federated setting on non-identically and non-independently distributed (non-iid) data. The findings of the study suggest that neural network architectures adapted to the federated setting demonstrate comparable prediction error to those trained in a centralized setting.

A recent study by Briggs et al. [13] proposed a federated learning-based model for STLF that utilized both local and global information. The authors reported that their model achieved improved prediction accuracy and outperformed traditional machine learning models.

A separate investigation conducted by Taik et al. [14] examines the effectiveness of incorporating edge computing and federated learning, a decentralized machine learning paradigm, to enhance the quantity and heterogeneity of data utilized in training deep learning models, while simultaneously safeguarding privacy. Notably, this research marks the inaugural utilization of federated learning in the domain of household load forecasting and presents promising outcomes.

To address the issues of inadequate data samples needed for constructing the load prediction model, and the high level of confidentiality of load-related data, which cannot be shared for training, this paper presents a federated learning approach. This approach proposes a short-term load forecasting model based on LSTM networks and incorporates local tuning of the federated learning-trained model. This method achieves collaborative training of the load prediction model among multiple load operators while ensuring data privacy, thereby enhancing the accuracy and generalization ability of the short-term load prediction model.

II. FEDERATED LEARNING-BASED SHORT-TERM LOAD FORECASTING MODEL

2.1 Method overview

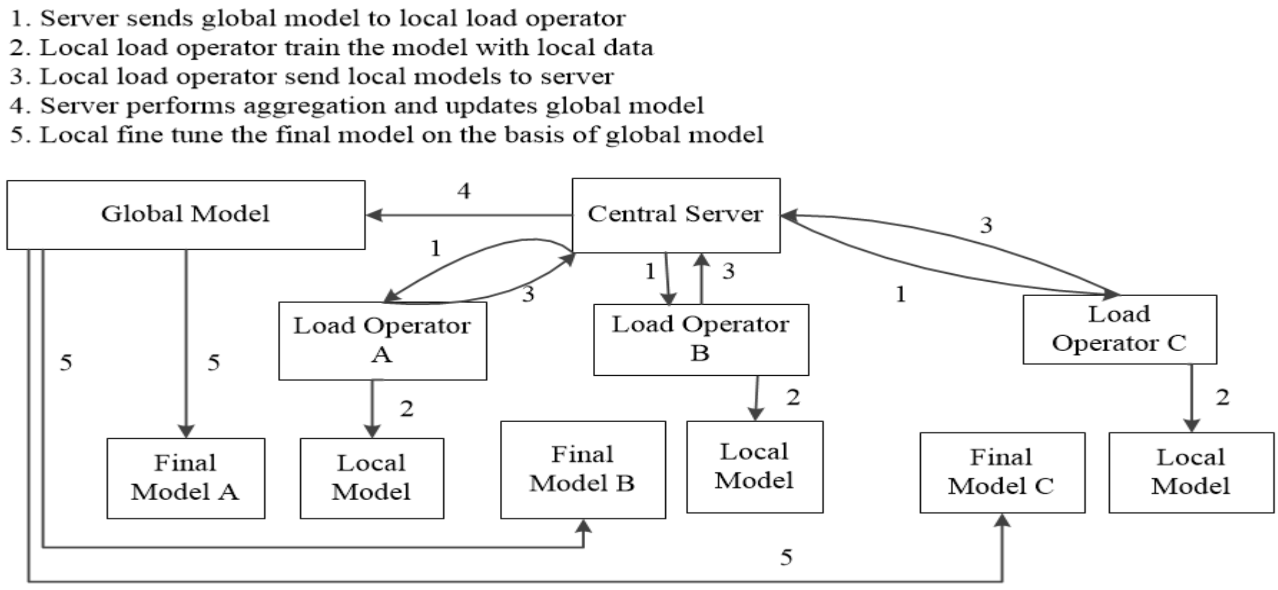


Fig. 1. LSTM model federal learning framework diagram

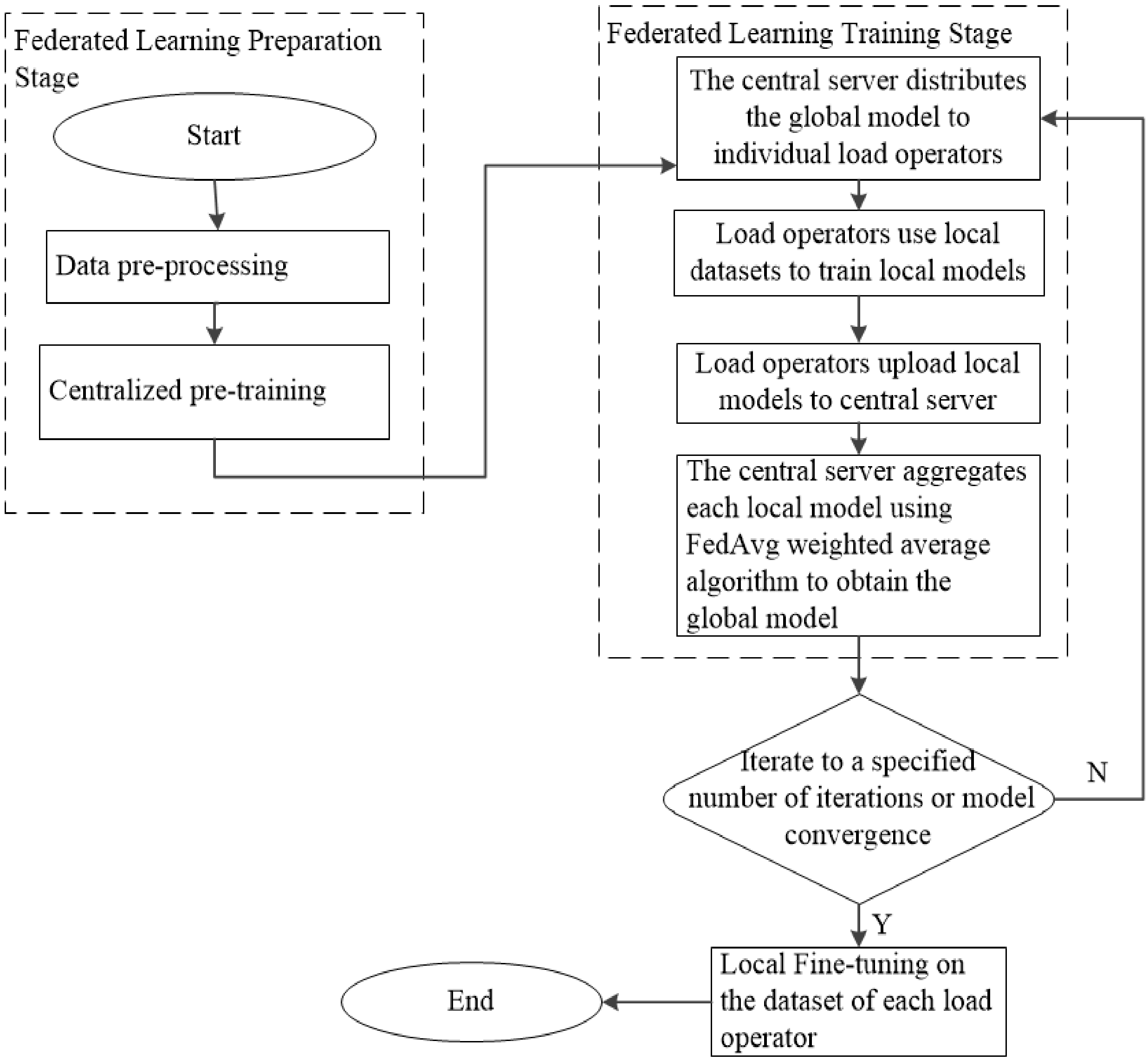


Fig. 2. LSTM model federal learning flow chart

The present study introduces a LSTM system for STLF that incorporates both federal learning and local tuning. This system is structured into two primary modules: the load operator and the central server. The load operator is mainly responsible for training the LSTM model with local data sets and then sending it to the designated central server. The central server continuously receives the local LSTM models from the load operators and then aggregates them using the specified aggregation algorithm to obtain a global model, which is then broadcasted to the load forecasting operators participating in the model training. After the federation learning is completed, a subsequent phase of local tuning is carried out on the individual load forecasting operators' local data. This process aims to derive a specific forecasting model for each load operator. The federated learning architecture design used in this paper is shown in Fig. 1.

The central server initializes a global model and distributes it to each load operator in the federated learning system. It collects the parameters of each load operator's local LSTM model and uses the FedAvg federated aggregation algorithm to compute the next round of global shared LSTM model parameters. The computed global LSTM model parameters are then shared with the load operators participating in the model training, completing a round of federated learning. The central server evaluates the model's performance based on performance requirements or the number of iterations. Finally, it obtains a global LSTM model with good performance. Each load operator fine-tunes their local model based on this global model to obtain the final training model. The specific workflow diagram is shown in Fig. 2, which mainly includes the following processes: (1) First, the data pre-processing work such as outlier processing and data normalization is performed on each load operator's data; (2) A global LSTM model is initialized by the central server using the existing public data, and then broadcasted to each load operator in the federated learning system; (3) In each training round, the participating load operators use the local data on the device to assess the model's performance. (4) The central server aggregates the local models from different load operators and updates the global shared model; (5) The above steps are repeated until the model accuracy reaches the required standard or the specified number of iterations. (6) Perform local tuning on the private data of each load operator to obtain the final training model. Unlike the traditional centralized training method, the federated learning training method only uploads the parameters of the training model instead of uploading the load data. This method can effectively protect the data privacy of the participants and effectively avoid the privacy security issue, while the availability of the data is not reduced because no desensitization operation is performed on the training data.

2.2 LSTM

LSTM is a type of recurrent neural network (RNN) specifically designed to learn long-term dependencies in sequential data. LSTM networks have memory cells that can retain information over extended periods, distinguishing them from traditional RNNs.

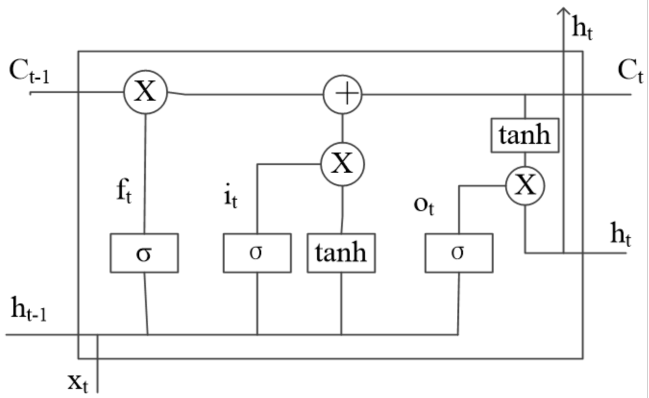


Fig. 3. LSTM network basic unit

In the architecture of an LSTM network, the memory cells establish connections with three essential gates, namely the input gate, output gate, and forget gate. These gates play a crucial role in regulating the information flow within the network by enabling selective retention or omission of information, guided by the present input and the internal state of the memory cells. Consequently, LSTM networks possess the capacity to adeptly handle and interpret intricate sequential data, attaining a notable level of precision. Fig. 3 illustrates the fundamental unit of the LSTM network.

The equations governing the input, forget, and output gates are as follows:

$$i_t = \sigma(W_i \times [h_{t-1}, x_t] + b_i) \quad (1)$$

$$f_t = \sigma\left(W_f \times [h_{t-1}, x_t] + b_f\right) \quad (2)$$

$$o_t = \sigma(W_o \times [h_{t-1}, x_t] + b_o \quad (3)$$

where σ denotes the sigmoid function, x_t is the current input, h_{t-1} is the previous hidden state, W and b are weight matrices and biases, and i_t, f_t , and o_t are the activations of the input, forget, and output gates, respectively.

2.3 Federated aggregation

Federated Averaging (FedAvg) is a popular federated aggregation algorithm that addresses this challenge by computing the weighted average of the local model updates from the clients.

The concept of FedAvg involves updating a global model iteratively by consolidating local model updates from a subset of clients. Each client autonomously trains its local model using its distinctive dataset and subsequently transmits the model updates to the server. The server aggregates these updates using a weighted average mechanism, with weights determined proportionally to the number of data samples available at each client. The resulting weighted average updates the global model, which is then communicated back to the clients for the next round of training.

Mathematically, the FedAvg algorithm can be expressed as follows:

Step 1. Initialize the global model parameters.

Step 2. For each round t=1, 2, ..., T:

a. Select a random subset of clients from the client pool.

b. Each client k in performs K local iterations of stochastic gradient descent to update its local model parameters.

c. Each client k sends its updated model parameters to the server.

d. The server calculates the weighted average of the model parameters received from all clients to derive the updated global model parameters.

e. These updated global model parameters are then disseminated by the server to all clients.

Step 3. Return the final global model parameters.

2.4 Local fine-tuning

Once a global model has been trained using federated learning, the next step is to deploy the model on the clients and use it to make predictions on new data. However, it is often the case that the global model may not perform well on the local data of each client, due to differences in data distributions or other factors.

To address this issue, local tuning can be applied to fine-tune the global model on each client's local data. This involves allowing each client to perform multiple rounds of local updates using its own data, similar to the process during the global model training phase.

Specifically, each client performs K rounds of local updates using the global model as a starting point. During each round, the client uses its own data to update the model parameters using a local optimization algorithm such as stochastic gradient descent (SGD) or Adam.

Local tuning can improve the performance of the global model on each client's local data by allowing the model to adapt to the specific features or patterns of that data. This is particularly useful in scenarios where the clients have diverse data distributions or where the clients are heterogeneous in terms of their computation or storage capabilities.

III. Experiments

3.1. Data description and error evaluation

This study undertook simulation experiments utilizing the GEFCom2012 dataset [15], comprising load and temperature data sourced from 20 unique regions within the United States. Both load and temperature data were captured on an hourly basis. For the purposes of this experiment, a sub-sample of data from five distinct regions was chosen. The training dataset was composed of observations ranging from January 1st, 2007 to March 31st, 2007, while the testing dataset encompassed a period from April 1st, 2007 to April 7th, 2007.

This paper introduces a federal learning-based LSTM STLF system, which necessitates constructing a distributed environment to facilitate the training process of the model. A comparative analysis is undertaken to assess the performance of the aforementioned system in relation to the centralized LSTM-based STLF system. To conduct this experiment, two machines with identical configurations were prepared. Each machine was equipped with an Intel Xeon platinum 8124 CPU, an RTX 3080 GPU, and 128GB of RAM. One of these machines was selected as the central server, while the other was used to generate multiple processes that simulated load operators situated in diverse regions.

Two commonly utilized metrics for evaluating the accuracy of load forecasting models are Root Mean Square Error (RMSE) and Mean Absolute Error (MAE). RMSE is more sensitive to outliers and assigns greater penalties to larger errors compared to MAE. The definitions of RMSE and MAE are as follows:

$$RMSE = \sqrt{\frac{1}{N}\sum_{I=1}^{N}(P_i - A_i)^2} \quad (4)$$

$$MAE = \frac{1}{N}\sum_{I=1}^{N}|P_i - A_i|^2 \quad (5)$$

In the above equation, N is the prediction interval, P_i and A_i are the ith predicted and actual values, respectively.

3.2. Method comparison

Table 1. RMSE performance comparison on GEFCom2012 dataset. Best performances are highlighted with boldface font.

Zone No.	LSTM	Federated Learning	Federated Learning + Local Fine Tuning
1	847.5	6523.35	**679.54**
11	4813.92	5561.15	**3311.24**
12	5292.81	8257.64	**4735.57**
15	4927.88	5422.22	**4709.01**
20	4513.55	5202.88	**4499.24**

Table 2. MAE performance comparison on GEFCom2012 dataset. Best performances are highlighted with boldface font.

Zone No.	LSTM	Federated Learning	Federated Learning + Local Fine Tuning
1	660.35	6386.86	**503.67**
11	3891.69	4529.28	**2433.83**
12	4022.41	6682.17	**3738.4**
15	3934.31	3901.80	**3411.93**
20	3322.94	3889.25	**3226.32**

To evaluate the forecasting performance of the model introduced in this paper, load prediction was conducted on the experimental dataset utilizing the proposed model, as well as two alternative models: the LSTM-based federated learning model and the LSTM neural network model. The obtained results are presented in Table 1 and Table 2, respectively. Based on the experimental outcomes, it is apparent that the model presented in this paper attains superior predictive performance across all five regions, thereby demonstrating a high level of accuracy and robustness. In the comparative analysis between the LSTM-based federated learning model and the LSTM neural network, the findings demonstrate that the former exhibits inferior predictive performance in all five regions. This outcome remains consistent despite the utilization of a federated learning approach to augment the experimental dataset. In comparison to the LSTM neural network, the proposed model exhibits a marked improvement in optimal RMSE and MAE, with enhancements of 30% and 40%, respectively. This observation implies that the model can better augment predicted regional load samples, while concurrently safeguarding the privacy of data from distinct regions. As a result, the model can more effectively capture the patterns of load data, leading to an enhanced predictive accuracy. To demonstrate the forecasting efficacy of the model introduced in this research, a comparative analysis is presented in Fig. 4, illustrating the predicted values generated by the proposed model alongside the corresponding actual values.

3.3. Experiments on AEMO dataset

To better validate the applicability of the method proposed in this paper, we conducted short-term load forecasting experiments using the Australian AEMO electricity dataset. The AEMO dataset [16] includes electricity load data from five states in Australia: New South Wales (NSW), Queensland (QLD), Victoria (VIC), South Australia (SA), and Tasmania (TAS). This dataset is a freely available public dataset that has been widely used in many studies. The following experiments use data from March 1, 2024, to April 5, 2024. Specifically, data from March 1, 2024, to March 31, 2024, is used for training, and the prediction data is from April 1, 2024, to April 5, 2024. The data sampling interval is 5 minutes. Therefore, there are 288 observations per day. Each training set contains 8,928 observation points, and each prediction set contains 1,440 observation points.

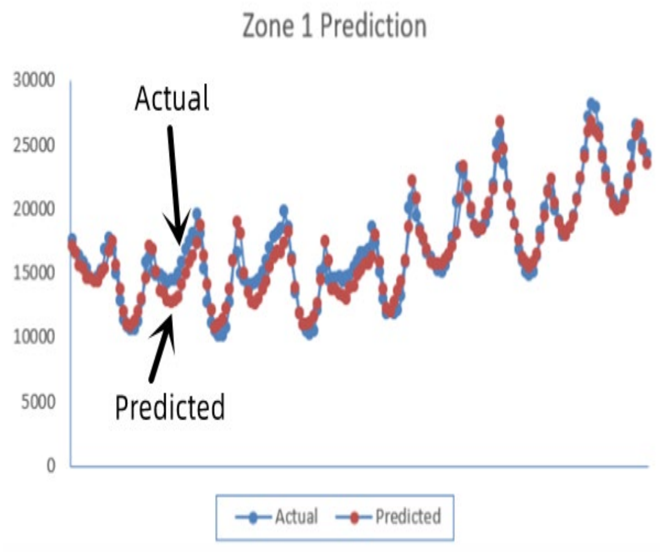
Tables 3 and 4 respectively present the RMSE and MAE results of three short-term load forecasting methods on the AEMO dataset. Based on the data in Tables 3 and 4, the same conclusion can be drawn from the experiments conducted on GEFCom2012 dataset, that the method proposed in this paper performs better than the other two methods in the short-term load forecasting results for all five states.

Table 3. RMSE performance comparison on AEMO dataset. Best performances are highlighted with boldface font.

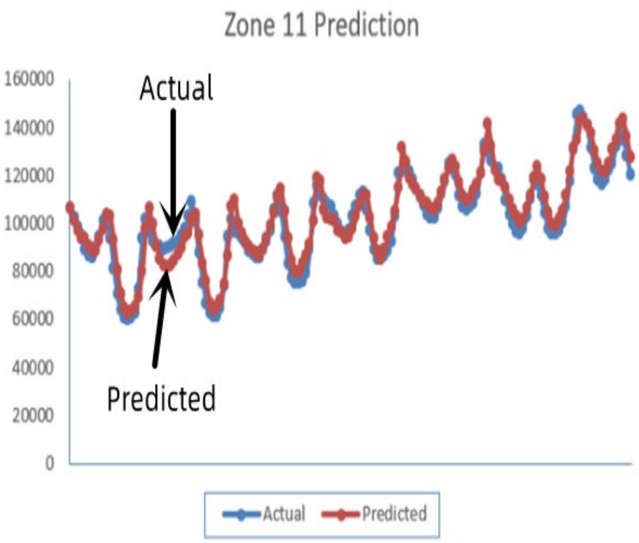
State	LSTM	Federated Learning	Federated Learning + Local Fine Tuning
NSW	975.88	1065.31	**876.43**
QLD	643.75	875.19	**523.85**
VIC	875.43	908.67	**805.66**
SA	654.86	813.26	**589.32**
TAS	705.38	896.31	**650.23**

Table 4. MAE performance comparison on AEMO dataset. Best performances are highlighted with boldface font.

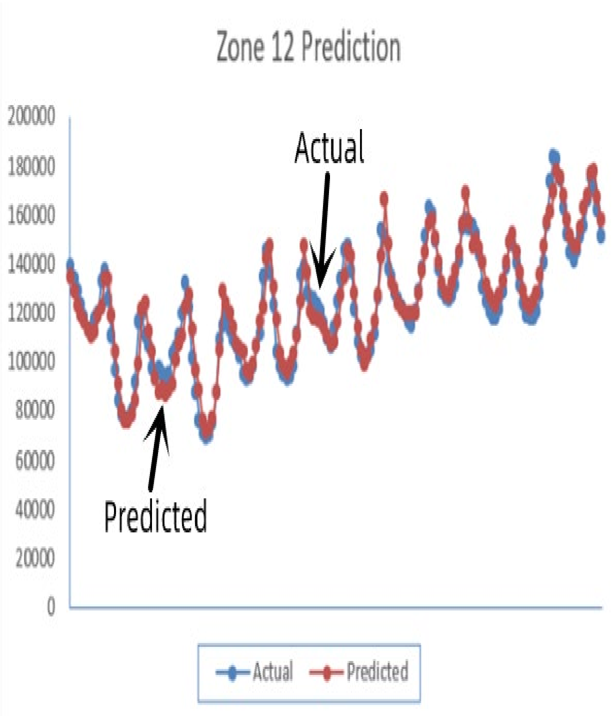
State	LSTM	Federated Learning	Federated Learning + Local Fine Tuning
NSW	753.41	821.65	**689.12**
QLD	506.58	649.12	**479.32**
VIC	965.32	1085.98	**856.73**
SA	723.58	905.39	**681.25**
TAS	786.32	975.32	**703.83**



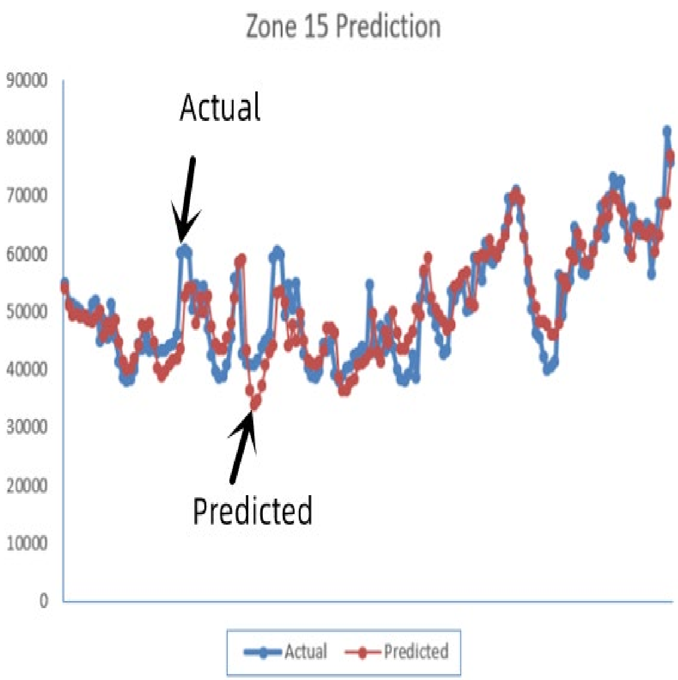
(a) Zone 1



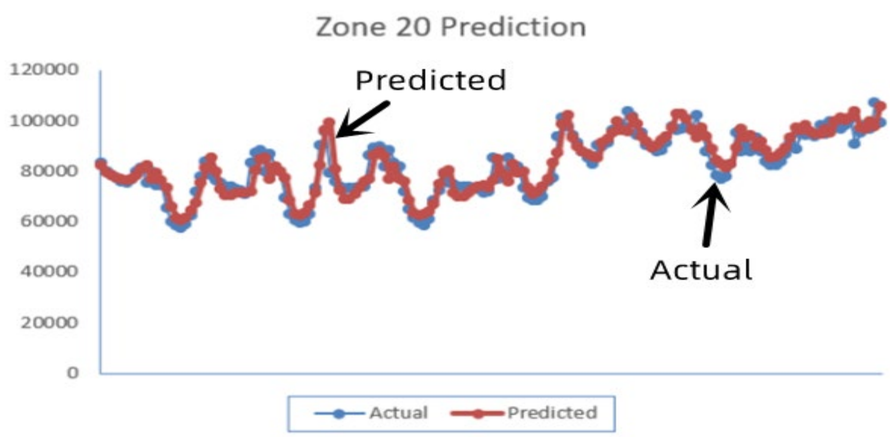
(b) Zone 11



(c) Zone 12



(d) Zone 15



(e) Zone 20

Fig. 4. Load forecasting results on GEFCom2012 dataset using the model in this paper

IV. CONCLUSION

This paper presents a novel approach to address the various challenges encountered in short-term load forecasting (STLF). Specifically, a two-stage method leveraging federated learning and local tuning is proposed to tackle issues including the limited availability of sufficient data samples and features critical for developing accurate load forecasting models, the restricted generalization capabilities of current models, and the need to maintain privacy and confidentiality of load data, which constrains sharing for training. The proposed method was simulated and evaluated using multiple regional load datasets from the GEFCom2012 and AEMO datasets, demonstrating efficacy in improving model predictive accuracy. Consequently, the optimized federated learning algorithm shows promise for further exploration to enhance load forecasting performance while preserving data security.

REFERENCES

[1] Al Mamun, Abdullah, et al. "A comprehensive review of the load forecasting techniques using single and hybrid predictive models," IEEE Access, vol. 8, pp134911-134939, 2020.
[2] Niu, D. et al. "Short-term multi-energy load forecasting for integrated energy systems based on CNN-BiGRU optimized by attention mechanism," Applied Energy, vol. 313, pp118801-,2022.
[3] Ma, Y., L. Yu, and G. Zhang. "A Hybrid Short-Term Load Forecasting Model Based on a Multi-Trait-Driven Methodology and Secondary Decomposition," Energies, vol. 15, no. 16:5875, 2022.
[4] Saeed, F., A. Paul, and H. Seo. "A Hybrid Channel-Communication-Enabled CNN-LSTM Model for Electricity Load Forecasting," Energies, vol. 15, no. 6:2263, 2022.
[5] Jeong, D., C. Park, and Y M. Ko. "Short-term electric load forecasting for buildings using logistic mixture vector autoregressive model with curve registration," Applied Energy, vol. 282, pp116249-,2021.
[6] Liu, R., et al. "Short-term load forecasting based on LSTNet in power system," International Transactions on Electrical Energy Systems, vol. 31, no. 12,2021.
[7] Syed, D., et al. "Deep Learning-Based Short-Term Load Forecasting Approach in Smart Grid with Clustering and Consumption Pattern Recognition," IEEE Access, vol. 9, pp54992-55008, 2021.
[8] Pham, M. H., M. N. Nguyen, and Y. K. Wu. "A Novel Short-Term Load Forecasting Method by Combining the Deep Learning with Singular Spectrum Analysis," IEEE Access, vol. 9, pp73736-73746, 2021.
[9] Xingchen Song, Tongxi Wang, Ziming Xu, and Zhan Shi, "Multistep Forecasting Method of Short-term Power Load Based on VMD-Prophet-Seq2seq," IAENG International Journal of Computer Science, vol. 49, no.3, pp720-727, 2022.
[10] McMahan, Brendan, et al. "Communication-efficient learning of deep networks from decentralized data." Artificial intelligence and statistics, PMLR, pp1273-1282,2017.
[11] Yan Li, and Guihua Wen, "Research and Practice of Financial Credit Risk Management Based on Federated Learning," Engineering Letters, vol. 31, no.1, pp271-278, 2023
[12] Perifanis, Vasileios, et al. "Federated Learning for 5G Base Station Traffic Forecasting," arXiv preprint arXiv:2211.15220,2022
[13] Briggs, Christopher, Zhong Fan, and Peter Andras. "Federated Learning for Short-Term Residential Load Forecasting," IEEE Open Access Journal of Power and Energy, vol. 9, pp573-583, 2022.
[14] Taik, A., and S. Cherkaoui. "Electrical Load Forecasting Using Edge Computing and Federated Learning," ICC 2020 - 2020 IEEE International Conference on Communications (ICC), pp1-6, 2020.
[15] Hong, T., P. Pinson, and S. Fan. "Global Energy Forecasting Competition 2012," International Journal of Forecasting, vol. 30, no. 2, pp357-363,2014.
[16] Australian Energy Market Operator (AEMO). Available online: www.aemo.com.au (accessed on July 2024).