

EMERGING TECH CONFERENCE – Edge Intelligence

Volume 02, 2023, Page 100 – 106

**Proceedings of Emerging Tech Conference:
Edge Intelligence 2023**

Federated transformers for non-intrusive load monitoring in heat pumps¹

Stylios Kanoutas, Athanasios Bachoumis, Michael Birbas, and Alexios Birbas
University of Patras, Department of Electrical and Computer Engineering. Patras, Greece
ece8346@upnet.gr, {abachoumis, mbirbas, [birbas](mailto:birbas}@ece.upatras.gr)}@ece.upatras.gr

Abstract

Edge intelligence, i.e., the execution of Machine Learning (ML) algorithms in computing resources at the edge, provides unprecedented benefits for applications in different verticals regarding data privacy, bandwidth, costs, and latency. Non-Intrusive Load Monitoring (NILM) is an application in the smart grid technology domain that could benefit from the advancements in edge intelligence to ensure consumer data privacy and decrease implementation costs. This paper proposes a federated learning-based transformer architecture for the NILM of energy-intensive residential devices, i.e., Heat Pumps (HP). We evaluate the architecture on an open-source dataset, showcasing that the performance does not deteriorate significantly compared to the centralized and is robust against the distribution shifts between the training and inference datasets and the increasing heterogeneity level between clients in the training data.

1 Introduction

Non-intrusive load monitoring (NILM) is a key technology that enables smart grids to monitor and disaggregate individual appliance energy consumption [1]. By providing granular energy consumption data, NILM can help ensure system stability, enable flexibility services, and support demand-side management, improving smart grids' overall efficiency and sustainability. NILM is an enabler for local aggregators to unlock the real-time demand-side flexibility potential of energy-intensive devices, by avoiding the capital expenses of installing new meters. Heat pumps (HP) are controllable loads that provide heating and cooling services, assisting the electricity grid's supply and demand balancing. Moreover, HP can be controlled through demand response programs to shift their electricity consumption to off-peak hours, reducing the strain on the grid during peak periods. However, some HP may connect to the main electrical meter of the building; thus, their consumption cannot be measured directly.

Privacy concerns are raised about end-consumer data in the digitalized era of power grids, where edge intelligence is an emerging topic, i.e., the deployment of ML algorithms and models on devices located at the edge, such as smart meters and gateways. To tackle this issue, a decentralized ML technique, named Federated learning (FL), enables edge devices to collaboratively train a shared model while keeping the training data locally. Lately, FL has been applied in smart grids [2]. Employing FL for NILM can address the privacy concerns associated with collecting and centralizing fine-grained energy consumption data, while

¹ This research has been partially financed by the EU Horizon 2020 project ENERMAN (GA: 958478).

not significantly decreasing the accuracy compared to centralized learning paradigms. Numerous NILM techniques have been developed, including probabilistic procedures (e.g., Hidden Markov Model), ML techniques such as support vector machines [1], and deep learning models [3, 4]. In addition, works have emerged employing FL to collaborative train of shared NILM models [5, 6].

In that direction, this paper contributes to the literature by applying for the first time an FL-based architecture, using Transformers as a model for the NILM of HP. It evaluates the architecture on an open-source dataset by comparing the accuracy versus other learning architectures, i.e., centralized and personalized learning. Finally, this work investigates the architecture's performance in unseen data while encountering for distribution shift between the training and test sets and heterogeneity among clients' data.

2 Problem statement and proposed solution

2.1. Problem Details

The primary goal of NILM algorithms is to deduce the consumption patterns of household appliances by extracting patterns from the aggregated power consumption data at a smart meter level. This procedure involves breaking down the overall power signal of a household into its constituent parts, namely the power signal of each appliance. This is expressed by:

$$P(t) = \sum_{i=1}^N p_i(t) + p_{noise}(t) \quad (1)$$

where N represents the total amount of appliances, and i denotes the index for the i_{th} appliance. At any given time t, the total power consumption $P(t)$ equals the combined power usage of all appliances N, represented by p_i . The $+ p_{noise}$ term denotes the noise. Another valuable purpose of the NILM is to classify the state of the i -th appliance except for the latter. We assume that the HP display two states (on/off), and for this categorization, a threshold is set to discriminate between those states. Additionally, we set a minimum on- and minimum off-time (3 minutes each) to avoid wrong predictions due to momentary inaccurate data from sensing.

2.2. Federated Transformers

Our study uses a Federated Transformer for NILM application in HP. The employed model architecture is the ELECTRICity [7]. The model comprises two transformer layers with two attention heads and a hidden size of 64. In the encoding layer, a 1D convolutional layer is implemented for feature extraction, followed by a squared average pooling layer. For the decoding side, a de-convolutional layer is employed. Additionally, we apply a 10% dropout probability to all layers. In contrast with the original training process of ELECTRICity, we discarded the pre-training unsupervised stage and, consequently, the generator-discriminator implementation, as we tried to keep our algorithm as lightweight as possible. In the supervised training process, each model gets the aggregated data for the smart meter and the HP signal as an input. The loss function L is defined:

$$L = \frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)^2 D_{KL} \left(\text{softmax} \left(\frac{\hat{y}}{\tau} \right) \parallel \text{softmax} \left(\frac{y}{\tau} \right) \right) + \frac{1}{N} \sum_{i=1}^N \log(1 + \exp(\hat{s}_i s_i)) + \frac{\lambda}{N} \sum_{i \in \mathcal{O}} |\hat{y}_i - y_i| \quad (2)$$

where λ represents a hyperparameter that governs the influence of the absolute error derived from the collection \mathcal{O} of inaccurately predicted samples and time points when the appliance's state is active. During

the inference, the input of the federated model is the smart meter's data and the output is the prediction of the HP consumption. The loss function considers the appliance's true status and the predicted consumption signal's on-off status, denoted as s .

Federated transformers for NILM

Kanoutas, Bachoumis, M. Birbas and A. Birbas

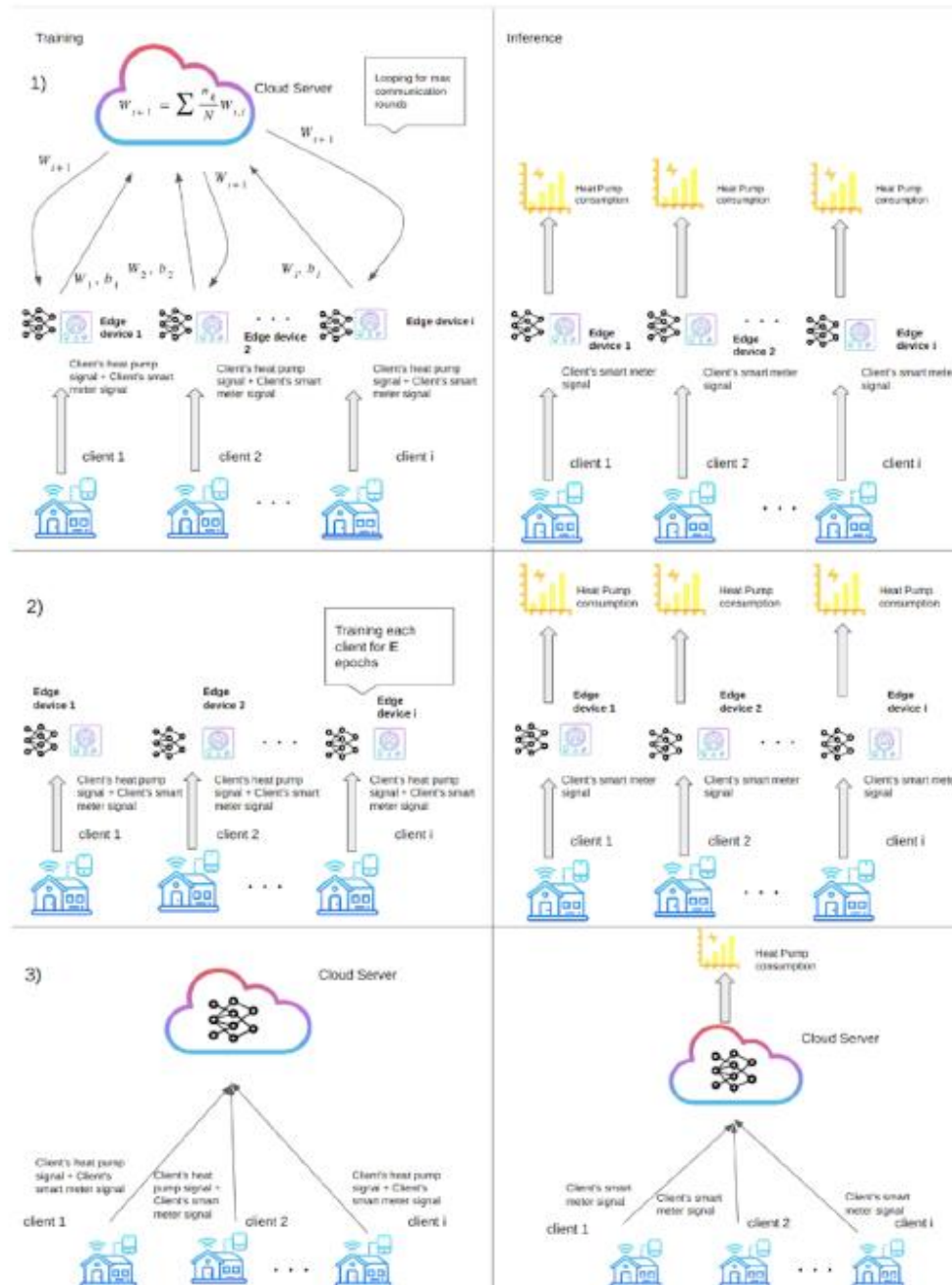


Figure 1: Different learning architectures for training and inference modes. Upper: FL-based, Mid: Personalized, and Lower: Centralized aggregated.

For the FL, we applied the FedAvg algorithm. Particularly, all clients are equipped with an edge device, e.g., a computationally capable smart meter. Each edge device trains a local model, with the specific household's data. We set the number of inner epochs to 1 and the max communication rounds to 49. We utilized the Adam optimizer with a batch size of 64 and set the λ and τ values equal to 1 and 0.1, respectively. The deployed architecture is illustrated in Fig. 1.

3 Algorithm evaluation case study

We utilized *WPuQ*, an open-source dataset, for training our model and evaluating its performance [8]. The dataset comprises electricity measurements of total household consumption and HP consumption from 38 single-family houses located in Northern Germany with resolution ranging from 10 seconds to 60 minutes. We used the 1-minute resolution of the active power data from 23 of the 38 houses. We excluded houses with limited data available from the analysis and those equipped with photovoltaic (PV) systems to use them further to validate the framework in unseen data. Data preprocessing takes place so that they comply with the following properties: Maximum limit (the upper limit that we set to protect our model from spikes that may distort it) is set to 12 kW and 9 kW for the aggregate signal and HP signal, respectively. The "on threshold", i.e., the value defining the on/off status, is set to 100 W. Finally, we also set the minimum off/on values equal to 3 min, i.e., the minimum time the HP signature signal should be greater or lower than the "on threshold" for the model to capture the on or the off status, respectively. The training set includes the data from 2018 and 2019, where we split it into 90% for the training and 10% for the validation set. The model was tested on out-of-sample data from the year 2020. In addition, the training set was split into windows of 480 samples with a stride of 120 samples.

We evaluate the performance of our architecture, by setting two learning architectures, i.e., personalized and centralized learning (mid and lower plots in Fig. 1). For each test case, we utilize 23 clients of the *WPuQ* dataset, either by aggregating the clients' HP data in the server (centralized case) or by developing personalized models for each client separately at the edge without any weight-sharing between the clients and the server (personalized case).

To assess model performance, we employ both regression and classification metrics. The regression metrics evaluate the model's ability to predict HP consumption, while the classification metrics evaluate the model's ability to predict the on-off status of the appliance. Regarding the former, we employ:

$$MRE = \frac{1}{\text{man}(Y)} \sum_{i=1}^N |\hat{y}_i - y_i| \quad (3)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |\hat{y}_i - y_i| \quad (4)$$

where y_i and \hat{y}_i denote the ground truth and the model prediction, respectively. For the latter, we use:

$$F1 = \frac{TP}{TP + \frac{1}{2}(FP + FN)} \quad (5)$$

where TP, TN, FP, FN denote True Positive, True Negative, False Positive, and False Negative, respectively. Besides those metrics, we also utilize two statistical approaches: (i) The Kolmogorov-Smirnov (KS) test to measure the level of non-independent and identically distributed (non-IID) data between all the client's data in a specific training scenario and (ii) the Kullback-Leibler (KL) divergence to measure the distribution shift between the training and test sets.

4 Results analysis

4.1. Test cases comparison

In our study, we evaluate the federated implementation by comparing the results to 2 other test cases, the personalized and the centralized ones. For the training and inference stage of the test cases, we utilize 23 clients from the WPUQ dataset. On those 23 clients, the mean metrics comparison is displayed in Fig. 2. Regarding accuracy, F1-score, and MAE approaches, the results of the FL are close to those of the centralized architecture. This indicates that our proposed privacy-preserving method has no significant deficit in performance compared with the centralized and personalized architectures, with the centralized one having the lowest MAE. The personalized case performs slightly better than the federated one due to the individualized training conducted for each model, specifically tailored to each edge device.

We also apply the federated global model to unseen clients (clients without historical data). Unlike the training clients, the data from the unseen clients include the PV production signal [PV capacity: 0.75 kW (SFH26) and 4 kW (SFH33)], which is aggregated with the consumption signal. The test results [MAE & MRE & ACCURACY & F1-SCORE || SFH26: 143.66 & 0.93 & 0.93 & 0.71 || SFH33: 185.45 & 0.9 & 0.83 & 0.62] reveal that our FL model exhibits satisfactory performance, especially in accuracy when applied to unseen clients. We observe a decrease in efficiency, specifically for the SFH33 client, which contains a larger PV. This suggests that there could be a decrease in the performance as the PV size increases. However, further research is required to draw a secure conclusion.

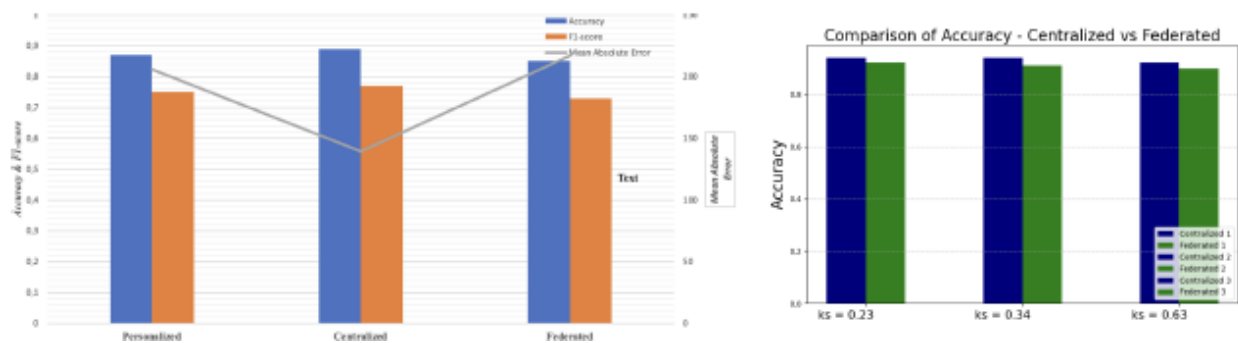


Figure 2: Left: Comparison for the different learning architectures, Right: Impact of non-IID data on testing accuracy between centralized model and federated model.

4.2. Impact of distribution shift on algorithm performance

Fig. 3 presents the relationship between each client's individual inference results and the distribution shift level between its training and test sets. It demonstrates that the model's architecture exhibits no significant decline in the performance associated with an increase in distribution shift until the KL-divergence surpasses a threshold of 4. At this point, there is a notable drop in the model efficiency. Therefore, it can be deduced that the federated transformer architecture is fairly robust to distribution shifts.

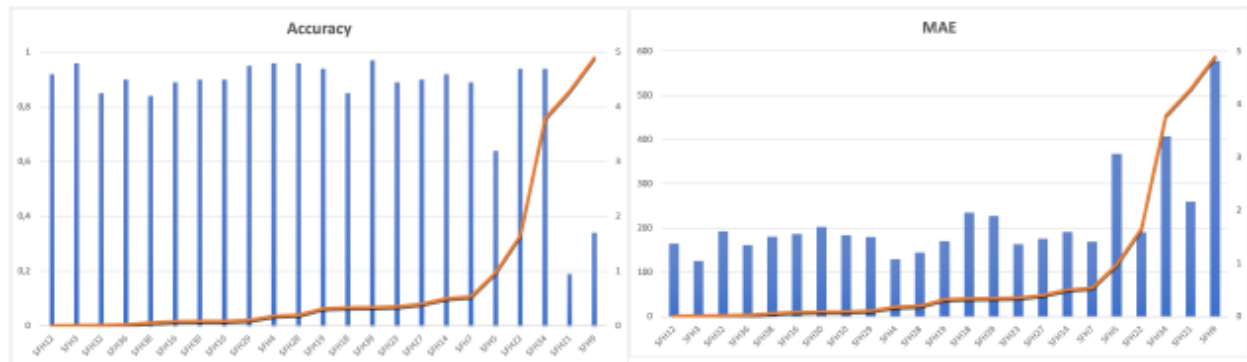


Figure 3: Impact of distribution shift level (KL-divergence with gold line) on federated clients testing accuracy (left plot) and MAE (right plot).

4.3. Impact of non-IID data on algorithm's performance

Furthermore, we estimate the impact of non-IID data on our model's performance. Non-IID data refers to situations where the training and test data distributions significantly differ among clients or households. The FL sensitivity in scenarios involving imbalanced or non-IID data is well-known. Fortunately, this limitation is effectively addressed by integrating the FL algorithm with the transformer architecture, thereby mitigating the effects of non-IID data [9]. Fig. 2 depicts the effect of different levels of non-IID data on the FL algorithm's accuracy. We tested the FL transformer architecture in 3 different scenarios with different non-IID data levels, i.e., 0.23 (first pair of bars), 0.34 (second pair of bars), and 0.63 (third pair of bars). In each of those scenarios, we have at least 8 clients. The results showcase the robustness of our model in handling non-IID data, as evidenced by the absence of any accuracy degradation while increasing the non-IID data level when compared with the accuracy of the centralized architecture.

5 Conclusions

The privacy-by-design principle is the backbone of any solution that ingests the raw consumer's data. In this work, we showcased the application of federated transformers for the NILM of HP, exhibiting performance similar to centralized and personalized cases. We demonstrated the advantages of transformer architecture in dealing with heterogeneous data and its robustness toward distribution shifts. Further experiments will be conducted to validate our method in different datasets while making adaptations in the transformer to increase accuracy.

6 References

- [1] Weicong Kong, Zhao Yang Dong, David J Hill, Jin Ma, JH Zhao, and FJ Luo. A hierarchical hidden markov model framework for home appliance modeling. *IEEE Transactions on Smart Grid*, 9(4):3079–3090, 2016.
- [2] Dimitris Mourtzis, John Angelopoulos, and Nikos Panopoulos. Integration of federated learning to smart grid for efficient and secure energy distribution. In *Proceedings of the Changeable, Agile, Reconfigurable and Virtual Production Conference and the World Mass Customization & Personalization Conference*, pages 477–486. Springer, 2023.

- [3] Weicong Kong, Zhao Yang Dong, Bo Wang, Junhua Zhao, and Jie Huang. A practical solution for non-intrusive type ii load monitoring based on deep learning and post-processing. *IEEE Transactions on Smart Grid*, 11(1):148–160, 2019.
- [4] Akriti Verma, Adnan Anwar, MA Mahmud, Mohiuddin Ahmed, and Abbas Kouzani. A comprehensive review on the nilm algorithms for energy disaggregation. *arXiv preprint arXiv:2102.12578*, 2021.
- [5] Shuang Dai, Fanlin Meng, Qian Wang, and Xizhong Chen. Federatednilm: A distributed and privacy-preserving framework for non-intrusive load monitoring based on federated deep learning. In *2023 International Joint Conference on Neural Networks (IJCNN)*, pages 01–08. IEEE, 2023.
- [6] Yu Zhang, Guoming Tang, Qianyi Huang, Yi Wang, Kui Wu, Keping Yu, and Xun Shao. Fednilm: Applying federated learning to nilm applications at the edge. *IEEE Transactions on Green Communications and Networking*, 2022.
- [7] Stavros Sykiotis, Maria Kaselimi, Anastasios Doulamis, and Nikolaos Doulamis. Electricity: An efficient transformer for non-intrusive load monitoring. *Sensors*, 22(8):2926, 2022.
- [8] Marlon Schlemminger, Tobias Ohrdes, Elisabeth Schneider, and Michael Knoop. Dataset on electrical single-family house and heat pump load profiles in germany. *Scientific Data*, 9(1):56, 2022.
- [9] Liangqiong Qu, Yuyin Zhou, Paul Pu Liang, Yingda Xia, Feifei Wang, Ehsan Adeli, Li Fei-Fei, and Daniel Rubin. Rethinking architecture design for tackling data heterogeneity in federated learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10061–10071, 2022.