engineering proceedings

MDPI

Proceeding Paper

Federated Learning for Frequency-Modulated Continuous Wave Radar Gesture Recognition for Heterogeneous Clients †

Tobias Sukianto [1,2,*], Matthias Wagner [2], Sarah Seifi [1,3], Maximilian Strobel [1] and Cecilia Carbonelli [1]

[1] Infineon Technologies AG, 81726 Munich, Germany; Sarah.Seifi@infineon.com (S.S.); Maximilian.Strobel@infineon.com (M.S.); Cecilia.Carbonelli@infineon.com (C.C.)
[2] Institute for Signal Processing, Johannes Kepler University Linz, 4040 Linz, Austria; matthias.wagner@jku.at
[3] Chair for Design Automation, Technical University of Munich, 80333 Munich, Germany
* Correspondence: tobias.sukianto@infineon.com
† Presented at the 10th International Electronic Conference on Sensors and Applications (ECSA-10), 15–30 November 2023; Available online: https://ecsa-10.sciforum.net/.

Abstract: Federated learning (FL) is a field in distributed optimization. Therein, the collection of data and training of neural networks (NN) are decentralized, meaning that these tasks are carried out across multiple clients with limited communication and computation capabilities. In FL, the client NNs are first trained with locally available data. Next, they are aggregated to update a global NN. FL suffers from non-independent and identically distributed (iid) data and asynchronous communication between the server and the clients, which degrades the NN's overall performance. In this work, we investigate FL for a small-live-gesture-sensing NN, using a low-power 60 GHz frequency modulated continuous wave radar from Infineon Technologies. The challenges of data sparsity, i.e., only a fraction of a gesture recording corresponds to an executed gesture combined with non-iid data, pose issues during neural network training. It is shown that FL reaches an accuracy higher than 96.2% for an iid setting. However, an increasing level of non-iid data degrades the accuracy to 64.8%. To tackle the accuracy degradation, we propose to dynamically adapt the class weights during the training procedure based on each client's varying ratio of data sparsity. Moreover, regularization terms are included in the loss function to prevent client drift and overconfidence in the client's NN prediction. Finally, it is shown that the proposed modifications increase the NN's performance, such that an accuracy of 97% is obtained despite a high degree of non-iid data.

Keywords: radar sensors; gesture recognition; machine learning; federated learning; IoT

Citation: Sukianto, T.; Wagner, M.; Seifi, S.; Strobel, M.; Carbonelli, C. Federated Learning for Frequency-Modulated Continuous Wave Radar Gesture Recognition for Heterogeneous Clients. *Eng. Proc.* **2023**, *58*, 76. https://doi.org/10.3390/ecsa-10-16194

Academic Editor: Francisco Falcone

Published: 15 November 2023

1. Introduction

Gesture recognition revolutionizes human–machine interfaces by providing a contact-free and intuitive method of interaction. Unlike touch-based systems, gesture-controlled systems introduce a touchless approach that enhances hygiene and enables interaction without direct hand exposure. Consequently, gesture sensing is one of the leading solutions for effortlessly managing a wide range of consumer and IoT devices [1]. Gesture recognition using radar sensing is a prominent application, merging signal processing-based feature extraction with the classification capabilities of neural networks (NNs). Radar sensors are advantageous to vision-based sensors in terms of privacy preservation, monetary cost, and memory efficiency. Early studies by Lien et al. in 2016 on gesture recognition with radar sensors propose a feature extraction based on range-Doppler images [2], which relies on intricate two-dimensional (2D) data processing. Furthermore, the training of NNs requires a rich database with broad distributional coverage that is then transmitted to a central server. While radar gesture sensing protects user privacy, data collection often involves sensitive user data, mainly because the radar sensor is paired with cameras to obtain accurate ground truth labels.

Federated learning (FL) offers a solution by shifting the training of a global NN to different clients, where all clients train a client NN with local data. These client NNs are then aggregated to learn a global NN, ensuring that the data remain local and are not transmitted to a central server. Although FL has gained significant traction over the past years, the used NN architectures are primarily large and require substantial computational power for training and inference. This poses a challenge in situations where the NNs should be implemented in computational and memory-efficient devices. In 2017, McMahan et al. [3] introduced the idea of FL and suggested an efficient way to train deep networks collaboratively. The challenge of heterogeneous FL, introduced by Zhao et al. [4], involves the training of NNs across different devices or servers with varying characteristics, such as data distributions and features. Challenges in FL when dealing with data heterogeneity, also known as non-independent and identically distributed (iid) data, were also addressed in [5–8]. They investigated how fluctuating data distributions in the clients affect the NN's convergence and proposed strategies to mitigate the impact of non-iid data. Communication efficiency is also a crucial research domain in FL, where the objective is to reach a high accuracy while minimizing the data exchange or the required communication rounds between the clients and the server. In [9,10], communication efficiency is enhanced by introducing various optimization techniques, and reducing the communication overhead while maintaining the NN's performance. Within the relatively unexplored domain of using FL for radar sensors, Savazzi et al. [11] investigated in 2021 a serverless FL approach, which addresses the task of tracking the position of individuals. In 2022, Yang et al. proposed an autoencoder-based technique to encode local gradients from client NNs into a lower-dimensional latent representation to decrease the transmission error within a three-class classification task across three clients [12]. However, these current state-of-the-art methods of FL in the context of radar do not account for the effects of data sparsity, imbalanced data distributions, or varying levels of non-iid data. Furthermore, they require significant changes in the NN architecture while utilizing computationally intensive 2D processing and large network architectures.

In this work, we apply FL on a small real-time gesture sensing NN designed for a low-power 60 GHz frequency modulated continuous wave (FMCW) radar sensor developed by Infineon Technologies and Google [1]. The NNs are designed to be efficient in terms of computational power and memory usage, aiming for minimal hardware requirements. The presented FL algorithms are evaluated on a diverse dataset, including approximately 26.000 gesture recordings. Our approach adopts the lightweight 1-dimensional (1D) radar processing algorithm from Strobel et al. [13], which requires fewer computational operations and smaller NNs than the 2D radar processing in [2]. Besides computationally efficient architectures, we address client heterogeneity and asynchronous client communication. To effectively overcome these training challenges, our main contribution involves dynamically adjusting the training process by assigning weights to the gesture recordings. These weights are based on the ratio of distinct gesture recordings and background within client's data, where all non-gesture recordings are considered as background. This strategy aims to counteract the accuracy degradation due to increasing levels of non-iid data and to decrease the number of communication rounds. An overview of our novel radar-based FL approach compared to prior work is highlighted in Table 1. The remainder of this paper is organized as follows. Section 2 discusses the radar processing setup and the neural network architectures and outlines the proposed contributions. The results are presented and discussed in Section 3. Section 4 concludes this paper.

Table 1. Comparison of radar-based FL approaches.

	Radar Processing	Parameters	Mitigates Non-Iid Data	Task
This work	1D	1.000	Yes	6-class classification
Savazzi et al. [11]	2D	3.000.000	No	Regression
Yang et al. [12]	2D	30.400	No	3-class Classification

2. Methods

In this section, the necessary radar processing steps, the sensor setup, and the utilized NN architectures are briefly discussed. Furthermore, the proposed modifications in the NN's learning method are introduced.

2.1. Radar Setup and Processing

The radar sensor emits a chirp signal, which ranges from 58.5 GHz to 62.5 GHz. This chirp signal is reflected by all objects in front of the sensor such that the reflected signal is received by the receive antenna. The received signal is downconverted with the transmit signal, yielding the intermediate frequency (IF) signal which is digitized with a sampling rate of 2 MHz into 64 so-called fast time samples which translates to a range resolution of 0.0375 m and a maximum detectable range of 1.2 m. Transmitting multiple (32) chirps allows to store the corresponding IF signals in a 32×64 matrix as illustrated in Figure 1, left. Note that the row index along the chirps is referred to as slow time. Computing a discrete Fourier transform (DFT) along fast time yields the range profile, wherein each peak at a range bin corresponds to an object at a certain radial distance. This is illustrated by Figure 1, center, where the range bins related to the hand and the body are highlighted in black. Finally, computing the so-called Doppler DFT along slow time at the hand's range bin yields another peak (Figure 1, right). The position of this peak relates to the radial velocity of the hand and the magnitude will be referred to as amplitude. As suggested by [13] only the hand's range bin is used for further processing. Hence, this is referred to as 1D radar processing. In addition to radial distance and radial velocity, the angle in azimuth and elevation of an object may be estimated by using three receive antennas arranged in an L-shape [14]. Consequently, five input features, i.e., radial distance, radial velocity, azimuth angle, elevation angle and the amplitude averaged over all receive antennas, are extracted from the radar data. Note that the radar data required to compute all five features will be referred to as a radar frame throughout this work. Within the classification task, we consider six gestures, i.e., swipe left, swipe right, swipe up, swipe down and push, and no gesture which is referred to as background.

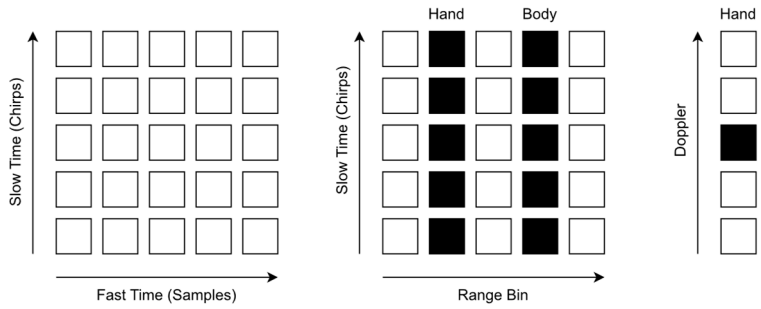

Figure 1. Illustration of the reduced 1D radar processing algorithm. Raw radar involves illustrated data for one antenna (**left**). The hand and the body are resolved with a discrete Fourier transform along the fast time samples (**center**). The discrete Fourier transform is applied to the detected hand range bin to resolve its velocity (**right**).

The network consists of a long short-term memory (LSTM)-based architecture to capture the time dependencies for the gesture recordings that are each comprised of a sequence of radar frames. In our model, the inputs are sequences of the five extracted features, while the output consists of gesture predictions. The start until the end of each gesture is labeled as the executed gesture to allow a real-time recognition of the gestures, while the remainder of the gesture recording is labeled as background. The LSTM layer is initialized with 16 hidden units, followed by a dense layer with six output neurons and a softmax activation representing the five gestures and the background.

2.2. Learning Methods

In each communication round t of the FL method, d NN weights $\mathbf{w}_t \, \epsilon \, \mathbb{R}^d$ are transmitted from the server to a selected group of $K \, \epsilon \, \mathbb{N}$ clients with $n_i \, \epsilon \, \mathbb{N}$ data samples. These clients collectively possess $n = \sum_{i=1}^{K} n_i$ data samples, allowing them to engage in localized learning using their local data samples while referencing the server's weights for their individual NNs. The resulting client weights $\mathbf{w}_{t+1}^{i}$ are then sent to the server after local training, which aggregates them into an updated set of global weights,

$$\mathbf{w}_{t+1} = \sum_{i=1}^{K} \frac{n_i}{n} \mathbf{w}_{t+1}^{i}. \tag{1}$$

The server and clients repeat this procedure through multiple communication rounds to fit the global NN to the client data without exchanging training data between clients and server. Given that the gesture execution constitutes only a fraction of each recording, we are confronted with an imbalanced dataset, wherein the majority of ground truth labels correspond to background. Hence, we propose to adapt the loss function with respect to the ratio of background and gesture samples for each recording with length $F \, \epsilon \, \mathbb{N}$, and for all $C \, \epsilon \, \mathbb{N}$ classes during the training procedure. Specifically, the loss function may be written as

$$L_{\mathrm{S}} = -\frac{1}{F} \sum_{i=1}^{F} \sum_{c=1}^{C} \left(\frac{F}{\sum_{j=1}^{F} y_{j,0}} y_{i,c} \log(\hat{y}_{i,c}) + y_{i,0} \log(\hat{y}_{i,0}) \right), \tag{2}$$

where $y_{i,c}$ and $\hat{y}_{i,c}$ is the actual and predicted probability of the i-th frame to be the gesture c, where the index 0 corresponds to the background class. Furthermore, a constraint proposed in [9] prevents non-iid and asynchronous clients' weights from drifting too strongly compared to the server's weights $\mathbf{w}_t$ and serves as the regularization term

$$L_{\mathrm{cons}} = ||\mathbf{w}_t - \mathbf{w}_{t+1}^{i}||^2. \tag{3}$$

A different loss function, known as the confidence constraint, is utilized when multiple client NNs encounter varying label distributions and undergo different numbers of local training epochs. In such scenarios, these client NNs are sensitive to overfitting to their respective heterogeneous distributions, resulting in overly confident predictions on their individual local datasets. To address this issue, we utilize the constraint from [15],

$$L_{\mathrm{C}} = \sum_{c=0}^{C} \log(\hat{y}_{i,c}), \tag{4}$$

enhancing the generalization capabilities. The final loss function is defined as

$$L = L_{\mathrm{S}} + \lambda L_{\mathrm{C}} + L_{\mathrm{cons}}, \tag{5}$$

with $\lambda \, \epsilon \, [0,1]$ and $\epsilon \, [0,1]$ as weighting coefficients. For the baseline approach, (2) is replaced in (5) with the cross-entropy loss.

3. Results and Discussion

The first study includes varying degrees of non-iid data in each client. The data are split in an iid partition, where the data are shuffled, and an equal number of gesture recordings is assigned to each client. In the non-iid configuration, each client is assigned an equal number of gesture recordings. However, the client's dataset only contains a subset of classes ranging from 1 to 4. The batch size, referring to the number of gesture recordings utilized in one iteration of the training, is fixed at 32. The stochastic gradient descent optimizer is used with a learning rate of 0.0001, and the algorithms are evaluated in 800 communication rounds. The total number of clients is 100, and the number of selected clients in each communication round is 10. Furthermore, a model is trained where all the data are centralized based on Strobel et al. [13] to compare FL with classical ML. Table 2 illustrates the test accuracy. In the synchronous client setting, all clients complete the same number of local epochs (Table 2, columns 2–3), whereas in the asynchronous client setting, the number of local epochs can vary across clients (Table 2, columns 4–5). A fixed number of five local epochs is used for all clients in the synchronous client scenario, while in the asynchronous client scenario, the number of local epochs is randomized for each client within the range of 1 to 20 epochs. The results in Table 2 reveal that, as the degree of non-iid data increases, there is a noticeable drop in accuracy for the baseline approaches. The proposed approaches prevent the effects of non-iid-related accuracy degradation for synchronous and asynchronous clients. For instance, higher than 97% accuracy is achieved in the one label per client setting, which is comparable to the iid case with an accuracy higher than 98%. One should note, also for the baseline approaches, that randomizing local epochs within the clients does not affect the accuracy drastically. The underlying reason could be ascribed to the regularization terms in (3) and (4), which might effectively mitigate the effects of asynchronous clients or that the randomization of the local epochs effectively yields more training iterations. As can be seen in Figure 2, by integrating the proposed approach, not only the necessary communication rounds are reduced, but also superior performance is achieved even in scenarios with increased levels of non-iid data. Furthermore, classical ML outperforms FL scenarios by a small margin. Nevertheless, FL has the advantage of not requiring to aggregate the client data in a centralized fashion.

Table 2. Gesture accuracy for varying levels of non-iid data and asynchronous clients.

Labels Per Client	This Work: 5 Local Epochs (Synchronous)	Baseline: 5 Local Epochs (Synchronous)	This Work: 1 to 20 Local Epochs (Asynchronous)	Baseline: 1 to 20 Local Epochs (Asynchronous)	Baseline: Traditional Learning
5 (iid)	98.2%	96.2%	98.4%	96.0%	98.8%
4	98.0%	86.2%	98.4%	91.9%	-
3	97.7%	83.2%	97.8%	90.0%	-
2	97.4%	79.0%	97.4%	88.1%	-
1	97.0%	64.8%	96.1%	78.5%	-

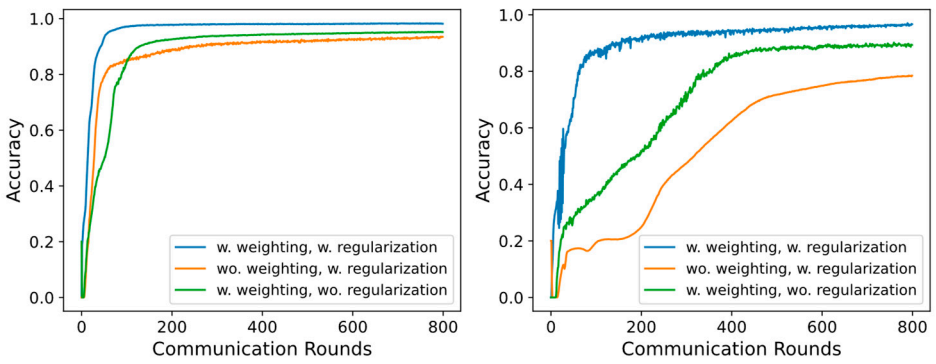

Figure 2. FL accuracy for five labels for each client (iid, **left**) and one label for each client (non-iid, **right**). The number of local epochs is randomized in each client between 1 and 20.

The second study illustrates the effects of the regularization terms denoted in (3) and (4). As may be seen in Figure 2, it is evident that the inclusion of those regularization terms in the loss function enhances the training of the NN, considering an increasing level of non-iid data. This results in a substantial reduction of required communication rounds. Asynchronously aggregating the clients' NNs into the global NN negatively impacts the accuracy of both the iid and the non-iid data partition. It should further be noted that combining the proposed loss function (2) and the regularization terms (3) and (4) is beneficial to achieve high accuracy and reduce the communication rounds. When only applying the weighted loss, significantly more communication rounds are needed to achieve comparable accuracy. Moreover, the negative impact of asynchronous clients increases with rising degrees of non-iid data, while synchronous clients achieve high performance without requiring the regularization term. This approach might be limited by the data quality available to the clients. In this work, it is assumed that the gestures are executed correctly. Falsely executed gestures could, therefore, degrade the performance of this approach. Therefore, future research should also address the open topic of varying data quality in the clients and could weight each client based on this.

4. Conclusions

In this work, we utilized FL in the scope of varying levels of non-iid data and client asynchronicity for low-power and small NN architectures within FMCW radar gesture sensing. We introduce a modified loss function to mitigate accuracy degradation caused by varying levels of non-iid data and client asynchronicity. We showed how an increasing degree of non-iid data decreases the NN's accuracy. By introducing a new loss function that incorporates the varying degrees of label sparsity in the training procedure, the gesture accuracy is increased by up to 33%. Furthermore, we identified adapting the class weights as a crucial component in the training procedure to maintain high accuracy and low communication overhead.

Author Contributions: Conceptualization, methodology and software, T.S.; validation and formal analysis, T.S., M.W., S.S., M.S. and C.C.; writing, T.S., M.W., S.S., M.S. and C.C.; supervision, C.C. All authors have read and agreed to the published version of the manuscript.

Funding: This work received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Trotta, S.; Weber, D.; Jungmaier, R.W.; Baheti, A.; Lien, J.; Noppeney, D.; Tabesh, M.; Rumpler, C.; Aichner, M.; Albel, S.; et al. 2.3 SOLI: A tiny device for a new human machine interface. In Proceedings of the 2021 IEEE International Solid-State Circuits Conference (ISSCC), San Francisco, CA, USA, 13–22 February 2021.
2. Lien, J.; Gillian, N.; Karagozler, M.E.; Amihood, P.; Schwesig, C.; Olson, E.; Raja, H.; Poupyrev, I. Soli: Ubiquitous gesture sensing with millimeter wave radar. *ACM Trans. Graph. (TOG)* **2016**, *35*, 1–19. [CrossRef]
3. McMahan, B.; Moore, E.; Ramage, D.; Hampson, S.; Arcas, B.A. Communication-efficient learning of deep networks from decentralized data. In Proceedings of the Artificial Intelligence and Statistics, Ft. Lauderdale, FL, USA, 20–22 April 2017.
4. Zhao, Y.; Li, M.; Lai, L.; Suda, N.; Civin, D.; Chandra, V. Federated learning with non-iid data. *arXiv* **2018**, arXiv:1806.00582. [CrossRef]
5. Diao, E.; Ding, J.; Tarokh, V. Heterofl: Computation and communication efficient federated learning for heterogeneous clients. *arXiv* **2020**, arXiv:2010.01264.
6. Li, X.; Huang, K.; Yang, W.; Wang, S.; Zhang, Z. On the convergence of fedavg on non-iid data. *arXiv* **2019**, arXiv:1907.02189.
7. Rodio, A.; Faticanti, F.; Marfoq, O.; Neglia, G.; Leonardi, E. Federated Learning under Heterogeneous and Correlated Client Availability. *arXiv* **2023**, arXiv:2301.04632.

8. Li, T.; Sahu, A.K.; Zaheer, M.; Sanjabi, M.; Talwalkar, A.; Smith, V. Federated optimization in heterogeneous networks. *Proc. Mach. Learn. Syst.* **2020**, *2*, 429–450.
9. Sattler, F.; Wiedemann, S.; Müller, K.R.; Samek, W. Robust and communication-efficient federated learning from non-iid data. *IEEE Trans. Neural Netw. Learn. Syst.* **2019**, *31*, 3400–3413. [CrossRef] [PubMed]
10. Lohana, A.; Rupani, A.; Rai, S.; Kumar, A. Efficient privacy-aware federated learning by elimination of downstream redundancy. *IEEE Des. Test* **2021**, *3*, 73–81. [CrossRef]
11. Savazzi, S.; Kianoush, S.; Rampa, V.; Bennis, M. A framework for energy and carbon footprint analysis of distributed and federated edge learning. In Proceedings of the 2021 IEEE 32nd Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), Helsinki, Finland, 13–16 September 2021.
12. Yang, Y.; Hong, Y.G.; Park, J. Federated learning over wireless backhaul for distributed micro-Doppler radars: Deep learning aided gradient estimation. *IET Radar Sonar Navig.* **2022**, *16*, 885–895. [CrossRef]
13. Strobel, M.; Schoenefeld, S.; Daugalas, J. Gesture recognition for fmcw radar on the edge. *arXiv* **2023**, arXiv:2310.08876.
14. Gerstmair, M.; Melzer, A.; Onic, A.; Huemer, M. On the safe road toward autonomous driving: Phase noise monitoring in radar sensors for functional safety compliance. *IEEE Signal Process. Mag.* **2019**, *36*, 60–70. [CrossRef]
15. Zhang, B.B.; Zhang, D.; Li, Y.; Hu, Y.; Chen, Y. Unsupervised domain adaptation for device-free gesture recognition. *arXiv* **2021**, arXiv:2111.10602.