# Toward a Federated Model for Human Context Recognition on Edge Devices

**Wandemberg Gibaut** * **Alexandre Osorio** * **Amparo Muñoz** *
**Sildolfo F. G. Neto** * **Daniel Miranda** * **Felipe Santos** *
**Michelle Scarassati** * **Fabio Grassiotto** *

* *Hub de Inteligência Artificial e Arquiteturas Cognitivas (H.IAAC), Eldorado Research Institute, Campinas- SP, Brazil (e-mails: wandemberg.gibaut@eldorado.org.br, alexandre.osorio@eldorado.org.br, fabio.grassiotto@eldorado.org.br).*

**Abstract:** Federated Learning is a promising technology to address crucial problems, such as those related to data privacy, involved in training Machine Learning (ML)/Deep Learning (DL) models in a distributed way. On the other hand, Human Activity Recognition (HAR) has recently gained more attention due to the evolution of the technologies involved, such as sensor availability, advances in ML/DL/Edge AI, and IoT. Due to computational resource constraints, techniques must be employed to reduce the effort required to train the model on the device. Meanwhile, there's the need to customize the ML model of each Federated Learning (FL) client with the specific data collected by that client. The present work explores the FL of an ML model for HAR in a set of twelve simulated FL clients, each with its own set of data from smartphone sensors. The FL loop starts from a global model that was previously trained in a centralized way, using a large dataset, different from the data used individually by each client during the FL. In this way, the FL constitutes a fine-tuning of the base model. The metrics collected are balanced accuracy and loss. Data is pulled from the ExtraSensory dataset, creating a benchmark for future applications across device farms and in-the-wild devices. The results show that our models achieve equivalent or better performance than most methods found in the literature, using a relatively simple Multilayer Perceptron (MLP) model. The proposed method can then reduce the time needed to retrain the model when data is acquired from the device's own sensors.

*Keywords:* Artificial Intelligence, Fuzzy and neural systems relevant to control and identification, Human Activity Recognition, Federated Learning.

## 1. INTRODUCTION

According to Cisco, in 2018 there were 8.8 billion mobile devices and connections globally which could grow to 13.1 billion by 2023 (Cisco, 2018). At the same time, the development and improvement of Machine Learning (ML) and Deep Learning (DL) techniques are revolutionizing the way we interact with devices and the world itself. This success depends on the availability of large-scale training infrastructures, such as large GPU clusters and the ever-increasing demand for large amounts of training data. The most common approach to data storage and model training is to use cloud servers running centralized machine learning methods. Although some edge nodes collaborate with the cloud performing distributed tasks, this protocol still has inherent challenges, one of the most relevant being the transmission over the network of all data collected from edge devices to train a model in a centralized server. This approach raises major concerns due to communication costs, as well as issues of reliability, privacy and data security and restrictions derived from administrative and government policies.

Due to the ever-growing availability of embedded sensors, as well as advancements in ML, AI and IoT, the research topic of Human Activity Recognition (HAR) is getting more attention. The main idea of HAR is to correctly identify activities such as walking, sleeping, running, etc, as well as placement (indoor or outdoor) by using a combination of data collected from wearable/smartphone sensors like accelerometers, microphones, and cameras. HAR techniques can be used for medical applications like disease diagnosis and elderly monitoring, smart home applications like profile creation by the daily activity recognition, and many others. A survey of HAR state-of-the-art is presented in Jobanputra et al. (2019).

Looking at data privacy and security challenges, Federated Learning (FL), introduced by McMahan et al. (2017), is a promising approach to address such problems. In recent years, FL has been seen as a paradigm that allows collaborative training of ML and DL models to preserve data privacy and security, according to Bonawitz et al. (2019), Li et al. (2020) and others. Only the model weights are shared between clients and the server. The data used for training remains locally stored in the client. There are two main types of communication in FL: central server orchestrated and fully-decentralized (Kairouz et al., 2021). The central-orchestrated type (Bonawitz et al., 2019), the most prevalent, has in its structure a federation of clients, which can be small mobile devices (smartphones) or even organizations (like, for example, hospitals). These clients train a ML model and share its weights with the FL server, which is responsible for aggregating the set of weights from each client into a single global model, which is then submitted to each client to start a new round. Li et al. (2020) and Zhao et al. (2018) demonstrate that with decentralized communication it is possible to update the global weights directly.

According to IBGE (2019), about 81% of Brazilians over ten years old own a smartphone, making Brazil the fifth country in the world in terms of number of users. In the banking technology sector, investment in Brazilian fintech companies totaled around 9 billion reais in 2020, representing an increase of 86% over the previous year, according to Fintechnews (2021). In 2020, half of the banking transactions were carried out through mobile devices, according to FEBRABAN (2021). Furthermore, the Brazilian penetration level in social media is approximately 68%, based on the studies of Statista (2022).

With the pervasiveness of edge devices in modern society, such as smartphones and wearable devices, the growth of private data originating from sources distributed through wearable and inertial sensors has increased, according to Cisco (2018). These data allow better and more discreet recognition of human activity and the state of rest, sleep, and stress. Combining these two factors (private sensor data and artificial intelligence techniques) is gaining significant interest in general customer products and systems in an industrial context. The present work focuses on technologies and techniques for HAR from mobile sensors (accelerometers, gyroscopes, and others) data and their applications in edge devices such as smartphones.

The present work explores Federated Learning through the resources and structure of the Flower framework. The objective is to evaluate the training of a federated model based on a centralized one in an experiment to recognize human activities with public data from the ExtraSensory dataset by Vaizman et al. (2017). The Extrasensory dataset aims to validate the recognition activities in-the-wild, getting closer to practical applications that work in a real-life environment.

The related works are in Section 2. The theory and tools are in Section 3. The methodology and experiments are in Section 4, and, finally, the final considerations are presented in Section 5.

## 2. RELATED WORKS

In the past few years, relevant advancements in the area of Human Activity Recognition took place. The use of Multilayer Perceptrons (MLP) and other structures to recognize activities from mobile sensors is a common approach in publications. Mantyjarvi et al. (2001) used waist accelerometers to recognize a limited set of body movements. The authors combined wavelet transform with principal component analysis and independent component analysis for feature generation. At the same time, for the classification task, a MLP classifier and the best classification results for recognition of different human motions were 83-90%. Kwapisz et al. (2011) used a built-in accelerometer in a smartphone with fixed placement (front pant pocket) to recognize six body states. They also compared different models besides MLP, including logistic regression and decision tree. The model achieves an accuracy of 98% in some activities but underperforms in others. The model is trained on some features generated from descriptive statistics, the interval in milliseconds between peak values in sinusoidal waves associated with activities. Guiry et al. (2014); Shoaib et al. (2015) showed the advantage of fusing sensors from smartphones and smartwatches to improve the detection of some activities, including those associated with bad habits, like smoking. In Guiry et al. (2014), five algorithms were evaluated, C4.5, CART, Naive Bayes, MLP, and Support Vector Machines. The authors reported 100% accuracy for all instances. In Shoaib et al. (2015), only three algorithms were evaluated, Support Vector Machine, k-Nearest Neighbors, and Decision Tree. The evaluation is for each sensor, set of sensors, and, finally, a comparison is made between the approaches.

Kerr et al. (2016) showed that data collected under heavily controlled conditions may result in poor generalization to real-life situations. Natarajan et al. (2016) addressed problems that arise when training classifiers with data collected in laboratory and validating with data collected in the field, like class distribution and sensor feature distribution. Also, they found it difficult to have reliable ground-truth labels when collecting in-the-wild data. Ermes et al. (2008) used a system where a participant could use a personal digital assistant (PDA) to self-report activities and select a physical activity, a location, and an indication of eating vs non-eating. They addressed that modeling some structures depends on researchers' assumptions, which may not hold in real-life situations. Choudhury et al. (2008) addressed a system for practical context recognition, which is basically about being unobtrusive and lightweight, allowing more natural behavior. Khan et al. (2014) provided a smartphone and collected data of participants in their natural environments for a month.

Nishio and Yonetani (2019) implemented federated averaging into practical mobile edge computing (MEC) frameworks. They used a MEC framework operator to manage the resources of heterogeneous clients. Wang et al. (2019) performed FL on MEC systems with limited resources. They addressed the problem of how to manage limited computing and communication resources at the edge efficiently. They implemented many ML algorithms like linear regression, SVM, and CNN using federated averaging. He et al. (2020) also considered the limited computing resources of edge devices, with the proposal called FedGKT, where each device trains only a small part of an entire ResNet to reduce computation overhead.

Vaizman et al. (2017) presented a work in which completely unconstrained, self-reported context data is identified through a series of individual Logistic Regression classifiers. Also, they introduced the ExtraSensory dataset, a rich, publicly available dataset about in-the-wild collected user data for HAR. In Vaizman et al. (2018b) the authors presented an unified neural network model, modelling the context identification as a multi-label classification problem. They also modified the objective function to better fit unconstrained data. The app used to collect the data was presented in Vaizman et al. (2018a) and is also available as an open-source software.

ISSN: 2525-8311

DOI: 10.20906/CBA2022/3669

Besides the original paper (Vaizman et al., 2017), other works used the ExtraSensory as the main dataset, e.g. Vaizman et al. (2018b) where the same research group presented a unified neural network model to solve a multi-label problem, and Fazli et al. (2020), where the authors presented an hierarchical classification with a Deep Neural Network to classify the six main labels.

Another approach for unconstrained activity identification was shown in Fazli et al. (2020), where the authors presented an hierarchical division with the use of Deep Neural Networks to identify six activities related with body states, like standing, running or lying down.

### *2.1 Datasets*

ExtraSensory was built from a collection of 60 participants. These participants performed a range of physical activities (e.g., walking, running), daily activities (e.g., sleeping, watching TV), and in different locations (e.g., school, home, work). The data was collected from sensors like accelerometer, gyroscope, magnetometer, clock accelerometer, location, and audio. The complete dataset has more than 300,000 minutes of collection. The advantages of this dataset are: collections in realistic settings (every day, without restriction of device position, natural behavior), annotations that provide a more realistic view of a person's life, and complex activities performed such as washing dishes, drinking alcohol, riding the bus, eating and watching TV). ExtraSensory is the dataset used to train and evaluate the model developed in this work.

The ExtraSensory dataset, originally presented in Vaizman et al. (2017), has as a differential the way in which the data was collected. As explained before in section 2, many of the datasets built for similar purposes deal with the clutter of the human activity data by imposing constraints on the use of the devices to capture the data. The ExtraSensory dataset, on the other hand, is built by capturing this data "in the wild", which means that each volunteer subject was free to define the way they use the devices. This feature makes ExtraSensory a good choice if the objective is to recognize the subject's actions in the most natural situations possible, although it is a very challenging dataset to deal with.

Other datasets have already been released to experiment and evaluate human activity recognition using smartphones. Garcia-Gonzalez et al. (2020) built HAR as a dataset of phone sensors. The HAR has a set of six activities, divided among these stationary ones, and was built in a controlled environment. Another dataset is SHL3, built by Gjoreski et al. (2018), which uses sensor data to identify human activities during transport. The DU-MD mobility dataset, built by Saha et al. (2018), is a dataset built to assess daily activities.

Finally, ExtraSensory may be the closest to the expected environment in real life among all these datasets. In addition, analyzing the data makes it possible to perceive the number of challenges expected in human activity recognition through smartphone sensors. For these reasons, the group chose ExtraSensory as a data source to evaluate the distributed learning model.

## 3. THEORY AND TOOLS

### *3.1 Federated Learning*

The term *Federated Learning* emerged in 2017, when McMahan et al. (2017) presented the concept alongside with the first algorithm of model aggregation, the *Federated Averaging*. The authors justify the name by stating that the training process is made by a "loose federation of participating devices" and the main motivations are the privacy and security risks of sharing a user's raw data and the communication constraints of mobile devices with centralized servers. The Federated Optimization deals with non-identically distributed (non-IID), very unbalanced data, like datasets built from specific users sensing. That is, a single user's data is not representative of an entire population and, from all possible labels in a supervised learning context, some labels could be extremely rare or even nonexistent. Under such condition, some users will have a lot more data than others, representing a heavier usage of some service.

The *Federated Averaging* is a simple yet powerful algorithm that plays a central role in the Federated Learning approach. The main principle is to aggregate the weights of the individual models by, iteratively, taking rounds of local training, local model sharing with the server, updating the aggregated model weights by taking the average of each participating model's weights and sending back this model to a new round of local training. The algorithm 1 in McMahan et al. (2017) shows the pseudocode of this method.

### *3.2 Flower*

First presented in Beutel et al. (2020), Flower is a Federated Learning framework that is proposed to support experimentation with both algorithmic and system-related challenges. It offers a stable, language and ML framework-agnostic implementation of core components of a Federated Learning System, dealing with the heterogeneity of the Edge Devices ecosystem, as shown in Mathur et al. (2021). In the present work, Flower was chosen as the Federated Learning framework because it shows good behavior in both Linux x86 systems with Tensorflow (Abadi et al., 2016) and Android ARM systems with Tensorflow Lite. Although the experiments were executed in Python, in threads running on a Linux PC, one of the possible future works is to replicate the experiment in Android devices.

## 4. METHODOLOGY AND EXPERIMENTS

The main goal of the present work is to test Federated Learning techniques in a task that is both complex and privacy-sensitive - HAR. As the objective is to move towards the development of a model that should achieve a good training performance on edge devices, the choice was to start with a reduced scope. In this sense, the developed classifier considered only the six main labels of the Extrasensory dataset: 'standing', 'sitting', 'lying down', 'running', 'walking' and 'bicycling'. Those labels are mutually exclusive, which means that a subject cannot choose both 'running' and 'sitting' in the same sample, for example. Taking this approach at an early stage of

ISSN: 2525-8311

DOI: 10.20906/CBA2022/3669

development would simplify the problem. The data were split into training (and validation) and testing sets in an 80-20 ratio, so that the ratio between the six labels was kept in the split, as the classes in the dataset are highly unbalanced.

To test the generalizability of the model, a cross-validation technique was used, in which the dataset was partitioned among the various FL clients. Thus, the data from the 60 subjects in the dataset were split into 5 folds, each containing data from a different set of 12 subjects/clients, while the data from the remaining 48 subjects were used to train a base model. Thus, the FL process is a fine-tuning of the base model, as the weights of the global model are initialized with the values resulted from the centralized pre-training, using data from a comprehensive set of 48 individuals, different from the data used at each client of the FL. In this way, it is expected to reduce the number of epochs needed to customize the model in federated mode. Such a result will be of great importance when the experiment is carried out on edge Android devices, with limited CPU and memory capabilities.

### *4.1 Centralized Training*

The first part of the training consists of a traditional ML training process with the combined data of 48 users. The used model was a feed-forward MLP (Multilayer Perceptron) Neural Network, in which a Random Search hyperparameter optimization was performed. The tested values of each hyperparameter were:

- number of hidden layers: 1 or 2
- number of neurons on each hidden layer: 4 to 64, with a step size of 4
- learning rate: 1e-1, 1e-2, 1e-3 or 1e-4

The metric to optimize was the *Balanced Accuracy*, better explained bellow. The model used *Rectified Linear Unit* (*relu*) as the activation function in all neurons of the hidden layers, as well as a *sigmoid* activation function in the output layer. The loss function choosen is the *Binary Crossentropy*, and the *Adaptive Moment Estimation* (*Adam*) is the optimizer.

As the authors explain (Vaizman et al., 2018b, 2017), the *Balanced Accuracy* is a good choice when dealing with a high imbalance between the classes of the dataset, avoiding false conclusions caused by underrepresented classes. The metric can be defined as:

$$BalancedAccuracy = 0.5 * (specificity + sensitivity)$$

where

$$specificity = tn/(tn + fp)$$
$$sensitivity = tp/(tp + fn)$$

, and **tp**, **tn**, **fp** and **fn** stand for True Positive, True Negative, False Positive and False Negative, respectively.

The optimal model found by the hyperparameters optimization are show in Table 1. Notice that the table shows one hidden layer (dense_38) and the output layer (dense_39). The learning rate found alongside with this topology was *0.01*.

Table 1. Optimal model found by hyperparameters optimization

| Layer (type) | Output Shape | Param # |
|---|---|---|
| dense_38 (Dense) | (None, 44) | 9988 |
| dense_39 (Dense) | (None, 6) | 270 |
| Total params | | 10,258 |
| Trainable params | | 10,258 |
| Non-trainable params | | 0 |

### *4.2 Federated Training*

Starting with the previously 5 trained models, the clients were implemented using the Flower FL framework in individual threads of a Linux system, as mentioned before. Here, each client performed training and test rounds with their own datasets and shared the model weights with the FL server (also running locally) between those rounds. Figure 1 illustrates the process.

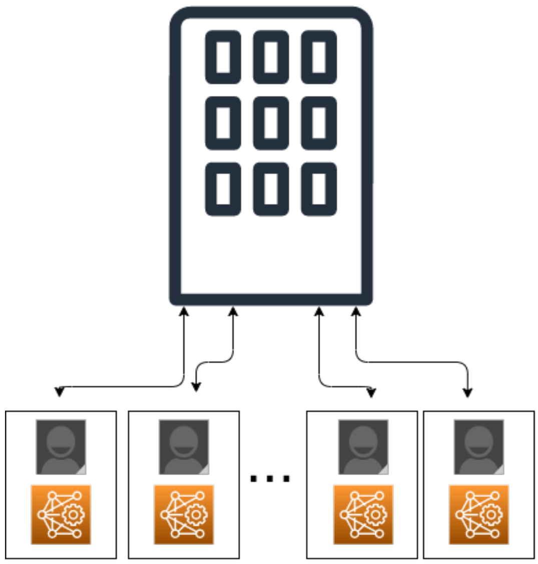

Figure 1. Federated Learning scheme illustration. Each individual client trains its local model with its own data. The weights are then shared with the FL server, which aggregates them with the *Federated Averaging* algorithm and the intermediate global model is then passed to all clients. The process continues iteratively, until a predetermined number of rounds has passed.

Flower allows to tune some strategy's parameters, like the fraction of clients that are chosen each round and the minimal available clients to start the learning process. The *FederatedAveraging* Strategy was configured as follows:

- *fraction fit*: 1.0
- *fraction evaluation*: 1.0
- *minimum fit clients*: 6
- *minimum evaluation clients*: 12
- *minimum available clients*: 12
- *batch size*: 50
- *local epochs*: 3
- *number of rounds*: 40

Five rounds of the FL training was run - one for each folder in the cross-validation process mentioned above - on 12

ISSN: 2525-8311

DOI: 10.20906/CBA2022/3669

threads running each flower client. The clients loaded the base model for each fold and train locally using their own data.

## 5. RESULTS AND CONCLUSIONS

Here are the discussions of the results. First, there was a low correlation between the number of samples available and the Balanced Accuracy for each client (approximately 0.282 using Pearson's correlation). Second, as can be seen in Table 2, there was a small difference in the metrics between the best and the worst fold (around 3.4% in Balanced Accuracy), lying between 0.826 and 0.853.

Table 2. Folds' statistics outlining a small difference in the metrics between the best and the worst fold

| fold | mean loss | mean n samples | mean balanced accuracy |
|---|---|---|---|
| 0 | 0.170 | 1110.454 | 0.853 |
| 1 | 0.198 | 1321.818 | 0.838 |
| 2 | 0.208 | 1279.727 | 0.825 |
| 3 | 0.162 | 1333.454 | 0.840 |
| 4 | 0.200 | 1277.636 | 0.826 |

Figures 2 and 3 show respectively a swarmplot and a boxplot of the Balanced Accuracy for each fold. The distribution shows that, despite most of the clients' BA being between 0.8 and 0.88 (the first and third quartile of all folds are in this range), the individual result can be as low as 0.75 or as high as 0.925. We argue that this phenomenon is close related with the **diversity** of data: clients that have higher diversity in the six main activities will produce data better suitable for the classification task. Bicycling and Running, for example, are activities that not everyone does.

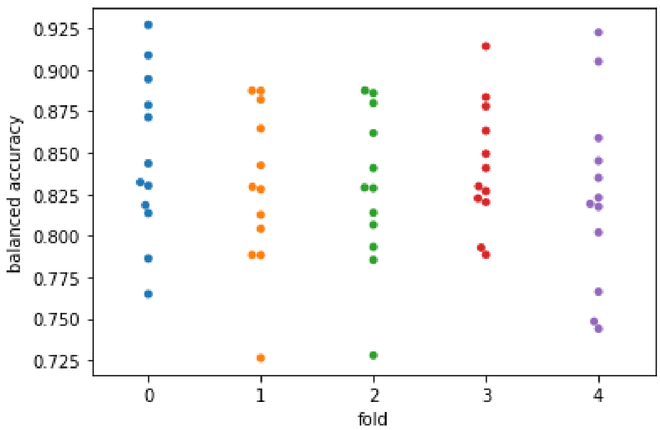


Figure 2. Swarm plot for each fold in the Federated Learning process. Note that despite most of the clients' BA being between 0.8 and 0.88 (the first and third quartile of all folds are in this range), the individual result can be as low as 0.72 or as high as 0.92.

Figure 4 shows the distribution of the Balanced Accuracy considering all 5 folds. The shape loosely resembles a gaussian distribution, but most of the curve is actually worst than the base models, which have a mean Balanced Accuracy of 0.854. We again argue that these discrepancies are due the diversity of data between the clients.

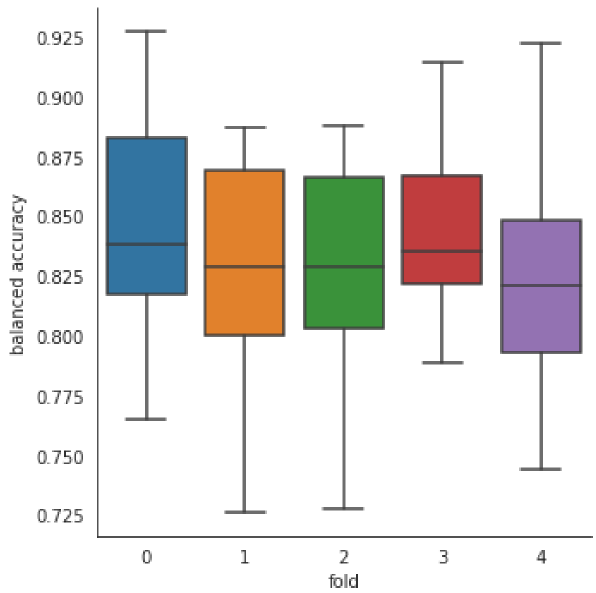


Figure 3. Box plot for each fold in the Federated Learning process. Note that the dispersion of the clients' BAs are relatively stable with close first and third quartile range, but the tails of the distributions may present very different values.

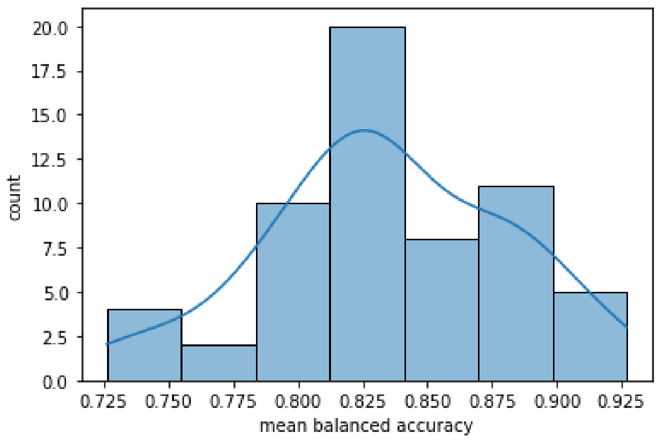


Figure 4. Histogram of Balanced Accuracy results considering all folds. The blue curve is a kernel density estimate (kde) that helps visualizing the data distribution.

Table 3 shows a comparison between the models of the present work and models found in publications using the same dataset. Numbers in parenthesis represent the number of neurons in hidden layers, that is, (8) would represent one hidden layer with 8 neurons while (16, 16) represents two hidden layers with 16 neurons each. Notice that the first two, from (Vaizman et al., 2018b) are for 51 labels, not only the main six activities. The models of the present work (the last two) achieved equivalent or better performance than most methods. Comparing with HHAR-Net, they are built also with simpler topology, being easier to embed in edge devices.

The experiments show that the approach of fine-tuning the centralized model in a federated manner can lead to good metrics for the HAR problem, with the advantage of having lighter models that could be trained on the edge devices, with a limited hardware. However, we know that some simplifications were made and an experiment with real devices would bring new challenges.

Table 3. Comparison between the models of the present work and models found in publications using the same dataset.

| Classifier | Balanced Accuracy |
|---|---|
| MLP(8) (Vaizman et al., 2018b) | 0.772 (51 labels) |
| MLP(16,16) (Vaizman et al., 2018b) | 0.773 (51 labels) |
| Decision Tree (Fazli et al., 2020) | 0.759 |
| k-NN (Fazli et al., 2020) | 0.788 |
| SVM (Fazli et al., 2020) | 0.792 |
| Random Forest (Fazli et al., 2020) | 0.709 |
| MLP (Fazli et al., 2020) | 0.814 |
| Flat DNN (Fazli et al., 2020) | 0.841 |
| HHAR-Net (Fazli et al., 2020) | 0.852 |
| **MLP(44)** | **0.854** |
| **Fed-Net** | **0.836** |

## 6. DISCUSSION

In this work we present a method that can be seen as a compromise between using user-specific data and the benefit of more data and a wider range of labels in non-IID applications, like Human Activity Recognition. This approach also keeps users privacy and could, theoretically, deal with ”cold-start” problem, that is, a fairly good model can be achieved even when new clients with small personal datasets begin to use an application. The practical effectiveness of the method is subject to limitations of network band, latency, users' connectivity and personal devices' computational power. Particularly in this work, as we ran all distributed clients in a single server, we were not affected by such limitations. We sought to apply it in a ”on-device trainning” fashion, but we needed to better delimiter our proposal before adding possible issues, like hardware limitations.

We argue that the results show evidences about the use of Federated approaches in Human Activity Recognition while keeps users privacy. Also, as can be seen in table 3, our method achieved better or somewhat equivalent performance than previous works. This hybrid approach could also be extended to other ML-based applications that deals with sensitive data and need large amount of data to perform well.

## 7. FUTURE WORKS

As future works we expect to design and run experiments considering real Android devices or Android device farms to be able to measure parameters on those specific hardwares. We would like to collect network performance metrics in poorly or intermittent available connection and observe how these adversities affects performance. Also, we would like to produce more descriptive labels for future usage with Natural Language Processing (NLP) techniques for further investigation.

## ACKNOWLEDGEMENTS

This project is part of the Hub for Artificial Intelligence and Cognitive Architectures (H.IAAC- Hub de Inteligência Artificial e Arquiteturas Cognitivas). We acknowledge the support of PPI-Softex/MCTI by grant 01245.013778/2020-21 through the Brazilian Federal Government.

## REFERENCES

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., Kudlur, M., Levenberg, J., Monga, R., Moore, S., Murray, D.G., Steiner, B., Tucker, P., Vasudevan, V., Warden, P., Wicke, M., Yu, Y., and Zheng, X. (2016). TensorFlow: A system for Large-Scale machine learning. In *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)*, 265–283. USENIX Association, Savannah, GA. URL `https://www.usenix.org/conference/osdi16/technical-sessions/presentation/abadi`.

Beutel, D.J., Topal, T., Mathur, A., Qiu, X., Parcollet, T., de Gusmão, P.P., and Lane, N.D. (2020). Flower: A friendly federated learning research framework. *arXiv preprint arXiv:2007.14390*.

Bonawitz, K.A., Eichner, H., Grieskamp, W., Huba, D., Ingerman, A., Ivanov, V., Kiddon, C.M., Konečný, J., Mazzocchi, S., McMahan, B., Overveldt, T.V., Petrou, D., Ramage, D., and Roselander, J. (2019). Towards federated learning at scale: System design. In *SysML 2019*. URL `https://arxiv.org/abs/1902.01046`. To appear.

Choudhury, T., Borriello, G., Consolvo, S., Haehnel, D., Harrison, B., Hemingway, B., Hightower, J., Pedja, P., Koscher, K., LaMarca, A., et al. (2008). The mobile sensing platform: An embedded activity recognition system. *IEEE Pervasive Computing*, 7(2), 32–41.

Cisco (2018). Cisco annual internet report (2018–2023). URL `https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.pdf`. Last checked on Apr 25, 2022.

Ermes, M., Pärkkä, J., Mäntyjärvi, J., and Korhonen, I. (2008). Detection of daily activities and sports with wearable sensors in controlled and uncontrolled conditions. *IEEE transactions on information technology in biomedicine*, 12(1), 20–26.

Fazli, M., Kowsari, K., Gharavi, E., Barnes, L., and Doryab, A. (2020). Hhar-net: hierarchical human activity recognition using neural networks. In *International Conference on Intelligent Human Computer Interaction*, 48–58. Springer.

FEBRABAN (2021). Febraban - notícias. `https://portal.febraban.org.br/noticia/3648/pt-br/`. (Accessed on 04/25/2022).

Fintechnews (2021). 5 trends shaping brazil's fintech industry in 2021 | fintech schweiz digital finance news – fintechnewsch. `https://bityli.com/erlhYy`. (Accessed on 04/25/2022).

Garcia-Gonzalez, D., Rivero, D., Fernandez-Blanco, E., and Luaces, M.R. (2020). A public domain dataset for real-life human activity recognition using smartphone sensors. *Sensors*, 20(8), 2200.

Gjoreski, H., Ciliberto, M., Wang, L., Morales, F.J.O., Mekki, S., Valentin, S., and Roggen, D. (2018). The university of sussex-huawei locomotion and transportation dataset for multimodal analytics with mobile devices. *IEEE Access*, 6, 42592–42604.

Guiry, J.J., Van de Ven, P., and Nelson, J. (2014). Multisensor fusion for enhanced contextual awareness of everyday activities with ubiquitous devices. *Sensors*, 14(3), 5687–5701.

He, C., Annavaram, M., and Avestimehr, S. (2020). Group knowledge transfer: Federated learning of large cnns at the edge. *Advances in Neural Information Processing Systems*, 33, 14068–14080.

IBGE (2019). Pnad contínua tic 2019: internet chega a 82,7% dos domicílios do país | agência de notícias. `https://bityli.com/IBmdX`. (Accessed on 04/25/2022).

Jobanputra, C., Bavishi, J., and Doshi, N. (2019). Human activity recognition: A survey. *Procedia Computer Science*, 155, 698–703.

Kairouz, P., McMahan, H.B., Avent, B., Bellet, A., Bennis, M., Bhagoji, A.N., Bonawitz, K., Charles, Z., Cormode, G., Cummings, R., et al. (2021). Advances and open problems in federated learning. *Foundations and Trends® in Machine Learning*, 14(1–2), 1–210.

Kerr, J., Patterson, R.E., Ellis, K., Godbole, S., Johnson, E., Lanckriet, G., and Staudenmayer, J. (2016). Objective assessment of physical activity: classifiers for public health. *Medicine and science in sports and exercise*, 48(5), 951.

Khan, A.M., Tufail, A., Khattak, A.M., and Laine, T.H. (2014). Activity recognition on smartphones via sensor-fusion and kda-based svms. *International Journal of Distributed Sensor Networks*, 10(5), 503291.

Kwapisz, J.R., Weiss, G.M., and Moore, S.A. (2011). Activity recognition using cell phone accelerometers. *ACM SigKDD Explorations Newsletter*, 12(2), 74–82.

Li, Q., Wen, Z., and He, B. (2020). Practical federated gradient boosting decision trees. In *AAAI*.

Mantyjarvi, J., Himberg, J., and Seppanen, T. (2001). Recognizing human motion with multiple acceleration sensors. In *2001 ieee international conference on systems, man and cybernetics. e-systems and e-man for cybernetics in cyberspace (cat. no. 01ch37236)*, volume 2, 747–752. IEEE.

Mathur, A., Beutel, D.J., de Gusmao, P.P.B., Fernandez-Marques, J., Topal, T., Qiu, X., Parcollet, T., Gao, Y., and Lane, N.D. (2021). On-device federated learning with flower. *arXiv preprint arXiv: 2104.03042*.

McMahan, B., Moore, E., Ramage, D., Hampson, S., and y Arcas, B.A. (2017). Communication-efficient learning of deep networks from decentralized data. In *Artificial intelligence and statistics*, 1273–1282. PMLR.

Natarajan, A., Angarita, G., Gaiser, E., Malison, R., Ganesan, D., and Marlin, B.M. (2016). Domain adaptation methods for improving lab-to-field generalization of cocaine detection using wearable ecg. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 875–885.

Nishio, T. and Yonetani, R. (2019). Client selection for federated learning with heterogeneous resources in mobile edge. In *ICC 2019-2019 IEEE international conference on communications (ICC)*, 1–7. IEEE.

Saha, S.S., Rahman, S., Rasna, M.J., Islam, A.M., and Ahad, M.A.R. (2018). Du-md: An open-source human action dataset for ubiquitous wearable sensors. In *2018 Joint 7th International Conference on Informatics, Electronics & Vision (ICIEV) and 2018 2nd International Conference on Imaging, Vision & Pattern Recognition (icIVPR)*, 567–572. IEEE.

Shoaib, M., Bosch, S., Scholten, H., Havinga, P.J., and Incel, O.D. (2015). Towards detection of bad habits by fusing smartphone and smartwatch sensors. In *2015 IEEE international conference on pervasive computing and communication workshops (PerCom Workshops)*, 591–596. IEEE.

Statista (2022). Social media usage in brazil – statistics & facts | statista. `https://www.statista.com/topics/6949/social-media-usage-in-brazil/`. (Accessed on 04/25/2022).

Vaizman, Y., Ellis, K., and Lanckriet, G. (2017). Recognizing detailed human context in the wild from smartphones and smartwatches. *IEEE pervasive computing*, 16(4), 62–74.

Vaizman, Y., Ellis, K., Lanckriet, G., and Weibel, N. (2018a). Extrasensory app: Data collection in-the-wild with rich user interface to self-report behavior. In *Proceedings of the 2018 CHI conference on human factors in computing systems*, 1–12.

Vaizman, Y., Weibel, N., and Lanckriet, G. (2018b). Context recognition in-the-wild: Unified model for multimodal sensors and multi-label classification. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 1(4), 1–22.

Wang, S., Tuor, T., Salonidis, T., Leung, K.K., Makaya, C., He, T., and Chan, K. (2019). Adaptive federated learning in resource constrained edge computing systems. *IEEE Journal on Selected Areas in Communications*, 37(6), 1205–1221.

Zhao, L., Ni, L., Hu, S., Chen, Y., Zhou, P., Xiao, F., and Wu, L. (2018). Inprivate digging: Enabling tree-based distributed data mining with differential privacy. In *IEEE INFOCOM 2018 - IEEE Conference on Computer Communications*, 2087–2095. doi:10.1109/INFOCOM.2018.8486352.