Human Activity Recognition with Wearables using Federated Learning

Borche Jovanovski[1], Stefan Kalabakov[1,2], Daniel Denkovski[1], Valentin Rakovic[1], Bjarne Pfitzner[3], Orhan Konak[3], Bert Arnrich[3] and Hristijan Gjoreski[1]
[1]*Faculty of Electrical Engineering and Information Technologies, "SS. Cyril and Methodius University" in Skopje, Rugjer Boshkovikj Str. 18, Skopje, N. Macedonia*
[2]*Jozef Stefan International Postgraduate School, Jamova Cesta Str. 39, Ljublana, Slovenia*
[3]*Digital Health – Connected Healthcare, Hasso Plattner Institute, University of Potsdam, Prof.-Dr.-Helmert-Str. 2-3, Potsdam, Germany*
kalabakovs@gmail.com, {borchej, valentin, danield, hristijang}@feit.ukim.edu.mk, {bjarne.pfitzner, orhan.konak, bert.arnrich}@hpi.de

Keywords: Activity Recognition, Machine Learning, Federated Learning, Deep Learning, Human Activity Recognition, Deep Neural Network.

Abstract: The increasing use of Wearable devices opens up the use of a wide range of applications. Using different models, these devices can be of great use in Human Activity Recognition (HAR), where the main goal is to process information obtained from sensors located in them, especially in eHealth. The high volume of data collected by various smart devices in contemporary ML scenarios, leads to higher processing consumption and in many cases results in compromised privacy. These shortcomings could be overcome by using Federated Learning (FL), a learning paradigm that allows for decentralized training of models such that user's personal data does not need to ever leave their devices, which substantially reduces to possibility of a breach. This paper analyses the behaviour and performances of FL when applied to the context of HAR. The obtained results show that FL can achieve comparable performances to those of centralized Deep Learning, while facilitating improved data privacy and diversity, as well as fostering real-time continuous learning.

1 INTRODUCTION

The ever-increasing trend of ubiquitous Internet of Things devices is paving the road for every day use and application of Wearable Sensor Technology (WST) [1]. WST has many potential benefits, such as: broad spectrum of usability, real-time information sharing, low cost of devices using this technology, custom hardware component of devices themselves, use for medical purposes in treatments and remote monitoring of patient condition, training and fitness, etc. The dominance of this technology is increasingly seen from a medical point of view. Specifically, the devices we carry enabled remote monitoring of the condition and health of patients. The data collected by the device is sent to the appropriate doctor, avoiding the unnecessary visit to the health institution. Continuous monitoring of patients' health helps to identify potential problems through preventive interventions, by improving the quality of care and saving money because the cost of prevention is often less than the cost of treatment [2]. The use of this type of devices, regardless of their purpose, is increasing day by day.Wearable Sensor Technology is a technology with a wide range of usability, which includes smart watches [3], fitness trackers, VR headsets, smart jewelry, webglasses and Bluetooth headsets. The way to use thesedevices is by attaching them to the human body, but there are some that do not have physical contact withthe user. A common factor among these WST devicesis the fact that each one monitors and collects datain real time. Due to the small form factor, these devices do not have much processing power or energy. This problem is exacerbated by the implementation of Machine Learning (ML) models, where the devices themselves are not able to store large amounts of data or apply complex DNNs [4]. Therefore, the information itself needs to be forwarded to a central server so that it can be processed. But there is a problem in terms of security and real-time continuous learning. Because of all this, when it comes to WST scenarios [5], the application will be most compatible with Fed erated Learning (FL). Analyzing FL with respect to HAR, the comparison between FL and DL scenarios, the influence of the number of users and the number of epochs/rounds in the models where the FL is used as well as its behavior for different situations are described in this paper.

The paper is organized as follows. Section 1 focuses on the implementation of WST using FL. Furthermore, Section 2 presents the related work in the area of FL and HAR, while Section 3 presents the data set background. Section 4 presents the methodology and describes the Feed-Forward Neural Network (FFNN) used in conjunction with the FL paradigm. Our evaluation setup, metrics as well as experimental scenarios are described in Section 5, while Section 6 presents the results and a discussion of them. Finally, we conclude the paper with our interpretation and reflection on the given question and outline the open perspectives and future work in Section 7 end spacing.

2 RELATED WORK

Advanced ML and DL solutions combine a large number of data sets collected by multiple sensor devices and therefore need to be located in one centralized location in order to achieve a single unit with high accuracy. The disadvantages of training centralized models appear in terms of the costs of establishing their own communication, as well as the security of privacy. Federated Learning (FL) can mitigate these disadvantages, by combining the information of locally trained models, obtained from the underlying clients. In [6] the authors evaluated FL as a kind of classifier for Recognizing Human Activities and compared performance with Centralized Learning based on a deep neural network and softmax regression, that are trained on synthetic and real data sets. The emphasis here is on the impact of erroneousclients with corrupted data on FL, as well as communication costs. It is concluded that FL manages to achieve a model with a worse but for now acceptable accuracy of 89% compared to 93% in Centralized Training when it comes to Deep Neural Networks. Inthis paper, the global model trained for FL in relation to faulty datasets shows that the accuracy here is comparable to the Centralized Learning model. A FL algorithm was further proposed that aims to first identify and reject users that contain corrupted data. Complex models that are built on the basis of Deep Neural Networks, and at the same time have implemented the HAR, the best benefits this solution wouldcertainly achieve by using FL. In [7], a new aggregation algorithm called FedDist is proposed here andimplemented to meet the needs of penetrating applications. This algorithm allows you to modify your model architecture by identifying differences between client-specific neurons. This allows the specifics of the customers to be taken into account without disturbing the generalization. In [8] the authors propose a new hybrid HAR method (FedHAR) that combinessemi-supervised learning and FL. The combinationof Active Learning and Spreading Labels to semiautomatically track the local flow of unlabeled sensor data relies entirely on FL to build a global modelof activity in a scalable and privacy-conscious manner. In [9], a new HAR FL system called FedDL has been proposed that captures basic user relationships and applies them to the dynamic learning of personalized models for different users. A dynamic layer sharing scheme is designed that learns the similarity between the weights of the user models to form the sharing structure and merges the iterative layer models from the bottom up. It also merges local models based on the dynamic sharing scheme, significantly accelerating convergence while maintaining high accuracy. In [10], a Meta-HAR federated representation learning framework is proposed in which the embedded signals are meta-learned by FL while the learnedsignal representations are entered into a personalizedclassification network for each user to predict activity. In [11], a new challenge has been explored that is of interest - tackling heterogeneity in users' activitiesusing FL. A framework with two different versions of activity-based federated aggregation is proposed, which uses overlapping gain of information across activities, with one using a model distillation update and the other using a weighted α-update. ClusterFL [12] is proposed as a FL system that is aware of similarities and can provide high model accuracy and low communication costs for HAR applications. In terms of the cluster connection learned, ClusterFL effectivelydiscards nodes that merge more slowly or have little correlation with the other nodes in each cluster, resulting in a significant acceleration of convergence whilemaintaining accuracy. Furthermore, our model will be presented, which consists of a set of data collected bytwo sensors, an accelerometer and a gyroscope placedin two locations on the human body on the wrist and thigh. Accuracy will be presented as a comparison when using Federated Learning versus Deep Learning for different scenarios, for example when using one or the other sensor at both locations, one or the other sensor separately for both locations and the twosensors together for both locations.

3 DATA

The dataset [13] consists of data from ten subjects, who performed the following activities of daily living: lying, walking, transition, sitting, standing, kneeling, all fours, bending, cycling, running. During the data collection process, subjects had two Inertial Measurement Units (IMUs) attached to their body, one on thewrist of the dominant hand and one on the thigh of thedominant leg. Both IMUs consist of: an accelerometer and a gyroscope. The sampling frequency of the IMUs is 50Hz. The number of samples in the training subset is 30607, while there are 6920 samples in the testing subset. The dataset contains features and logs that are organized into two directories: combined features and features. Both directories have the same data, organized in two slightly different ways. The Features-id and Labels-id files contain the extracted features (statistical + time domain + frequency domain) from the two IMU sensors (accelerometer + gyroscope) from the two IMU locations (wrist + thigh) for the object specified in their name. The data in the JSI-FOS database are approximately equally distributed among the ten subjects that participated in thedata collection, and the same can be said for the distribution of activities, hence we conclude that the distribution of activity and frequency of occurrence are the same in the test subsets and a train.

4 METHODOLOGY

4.1 Artificial Neural Networks

The inspiration from the way neurons in the human brain work is translated into these so-called Artificial Neural Networks [14], which are a kind of discipline when it comes to using Machine Learning. As technology breaks down, much more advanced machines with large capacities that can work with large amounts of data are being developed, rewarding this tin of Artificial Neural Networks (success is seen in the training of complex neural networks in some domains that require a lot of data and computing capacity). There are several types of Artificial Neural Networks, of which we can mention Feed-forward, recurring, etc. We will stick strictly to Feed-forward [15] which are successfully used for clustering, regression, template classification, optimization, control, etc. Characteristic of Feed-forward as an Artificial Neural Network is that its connections between the nodes themselves do not form a loop. On the opposite side of the Neural Networks moving forward, we have the Recurrent Neural Networks in which certain pathways are cyclically represented. The simplest form of Neural network is the retraction model because the information itself is processed in only onedirection, so the data passes through several hidden nodes, always moving in one direction and never going back. The Neural Network with more pronounced power is usually seen in its simplest form and is represented as a single-layer perceptron [16]. The modelcontains a series of inputs that are connected in a layer and multiplied by the corresponding weights. Each value is then added together to obtain the sum of the corresponding weighted input values. If the sum of the values is above a certain threshold, usually set to zero, the output value is often 1, while if the sum falls below the threshold, the output value is -1. This type of perceptron is very important in neural networks, especially when it comes to classification. Furthermore, the single-layer perceptron can incorporate aspects of machine learning, using a property known as the delta rule, the neural network can compare the outputof its nodes with predicted values, allowing the network to adjust its weights through training to produce more accurate output values. This way of training andlearning produces a form of gradient descent. In multilayered perceptrons, the process of updating weightsis represented by an analog, while the process is morespecifically defined as back-propagation. In that situation, the existence of a hidden layer in the network isadjusted according to the output values produced by the last layer. Feed Forward neural networks are characterized as quite simple in that their simplistic architecture takes advantage of certain applications when using machine learning.

4.2 Federated Learning

Federated learning [17] is a Machine Learning methodology in which all models experience is a multitude of situations with different data sets locatedin different locations, without disturbing the training sequences and data (Figure. 1). By sharing the device model, we can collectively train the model asa whole. Looking at this concept, anyone can participate directly or indirectly in the FL of their devices, e.g. edge devices, then smartphones and IoT devices, can benefit from device data without leaving the device, especially for devices that are confined to a computer where communication itself is presented as a bottleneck with smaller devices. The concept of moving data calculations is a powerful concept in terms of building any intelligent system, while protecting the privacy of all individuals. The FL paradigm downloads the current model and calculates the updated model on the device itself by utilizing its local data. These locally trained models are then sent from the devices back to the central server where they are collected, processed, commonly with an average weighted operation, and then a single consolidated and improved global model is sent back to the target devices. FL allows machine learning algorithms to gain experience from a wide range of data sets, even if they are present in different locations. This ensures that a number of different entities cooperate with each other and build a model that will be used by all stakeholders, and will not share their security information, but only the model they need will be strictly shared.

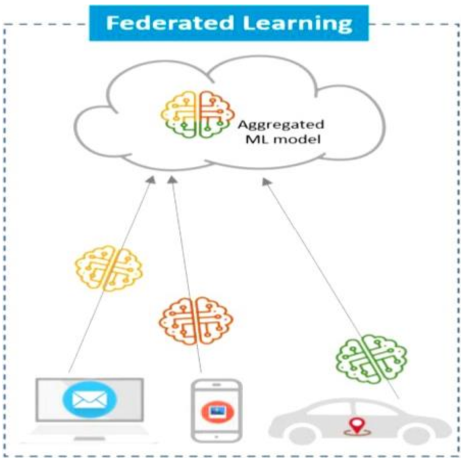

Figure 1: Federated Learning concept.

5 EXPERIMENTAL SETUP

This section presents the various scenarios and the model on which the tests were performed, as well as the behavior of the strategies we use during implementation.

5.1 Evaluation Setup

Behavior between FL and DL is presented in the following way: the collected data from the 10 users is divided into two possibilities, one for testing and the other for training. The test subset contains 20 percent of the original dataset, while the remaining 80 percent is intended for training. No validation set is used, and the goal is to directly report changes in performance made by different settings on the test subset. It was decided to use a more personalized evaluation setup instead of a Leave-One-Item-Out strategy, because FL provides a unique opportunity to train centralized models without explicitly sharing data between users. Splitting the data into two subsets strictly at each user serves to further merge with the corresponding subsets of the other users, with the final test and training subsets containing the data from all users, respectively. The way we use these two features during training and evaluation of DL and FL models is as follows: for DL models, the training feature updates the model in each epoch, while the test feature evaluates after each epoch and at the very end of training. On the other hand, in FL the training facility is used to train each local user model after each round, while testing it for evaluation. Finally, after each round we have an evaluation of the centralized model which contains the merged data from test subsets of all users in the database.

5.2 Metrics

Here, the macro F-score will be used to quantify the performance of each of our models, and this is done due to the unbalanced distribution of activity instances in the JSI-FOS database, in this study. The macro F-score allows avoiding the bias of metrics such as the accuracy of activities with a larger number of examples by calculating the F-score for each activity separately and reporting the average of the results obtained. This F-score metric is defined as the harmonic mean of the precision and recall metrics for a particular label. There is no value interpretation here, for example, precision metric, F-score, and macro F- score values closer to 1.0 represent better classification performance, while values closer to 0.0 represent worse classification performance. Note that the macro F-score and accuracy metrics typically report very close values on datasets with a balanced distribution of activity instances. Equations (1), (2), and (3) show how each of the precision, recall, and F-score metrics are calculated, respectively.

$$Precision = \frac{TP}{TP + FP} \quad (1)$$

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

$$F1 = \frac{2*Precision*Recall}{Precision + Recall} \quad (3)$$

5.3 DL and FL Models

The parallel between the DL and FL models is presented in terms of using a different number of training epochs/rounds. The main thing about this experiment is that we will look at the behavior of DL with respect to the FL model when we use the different number of epochs, that is, when the database will be fully used. The maximum number of epochs/rounds used for training is 50. Note that when comparing DLmodels and FL models, we treat one epoch (DL) andone round (FL) as equivalent. Here, we aim to understand the impact of individual FL parameters on the overall FL performance. To this end, we chose to observe the changes in performance produced by varying the number of epochs/rounds and the percentage of clients used for training during each round of the FL model trained when using the full dataset. We will include a metric called F-score, with which we will present the behavior of the model itself in relation to the number of clients per round, i.e. not only does it change the number of users who have access to each page, we also change the percentage of clients that are used for training in each round. The model fitted to this experiment is characteristic of 50 rounds.

6 RESULTS AND DISCUSSION

In this section, we present the results obtained from the various experiments regarding the behavior between the FL and DL models. FL is an online learning strategy and transfers the model weights to the FL centralized server in each round of operation, FL- based macro F-score curves are presented as continuous with respect to the data transfer volume. In contrast, since DL is an offline strategy and the data needs to be fully transferred to the centralized server to perform the model training. DL and FL strategies, multiple runs were conducted to calculate 95% confidence intervals for the macro F-score. We will look at the result for the macro F-score when using the full dataset (Figure 2). It is noteworthy that even in the case where the full dataset is used, the FL model slightly approaches the DL model in the case of 35 epochs, this is a result because the DL model has a fairly relative growth since the beginning compared to the FL model. The performance degradation of FL relative to macro F may come as a result of the small number of epochs used to train the local FL models. A slight degradation in terms of the performance itself is observed with the DL strategy, in the situation of the full feature dataset. We conclude that by increasing the number of local epochs in the FL strategy, it approaches the DL strategy, which shows the benefit of FL implementation.

Figure 3 shows a heatmap that presents the achieved macro F-score of the FL model when varying the percentage of clients used for training in one round and the total number of rounds used for training. More specifically, on the horizontal axis, Figure 2 show the percentage of clients used for training in each epoch and on the vertical axis, it shows the maximum number of rounds. After roughly 20 rounds of training, the performance of the FL model plateaus and the achieved performance varies by at most two percentage points. Furthermore, Figure 3 also suggests that using a larger percentage of clients, i.e., above 50%, for training during one round produces better results. In our case, the best results seem to be produced by an FL model that uses 80% of all clients to update the centralized model in each training round.

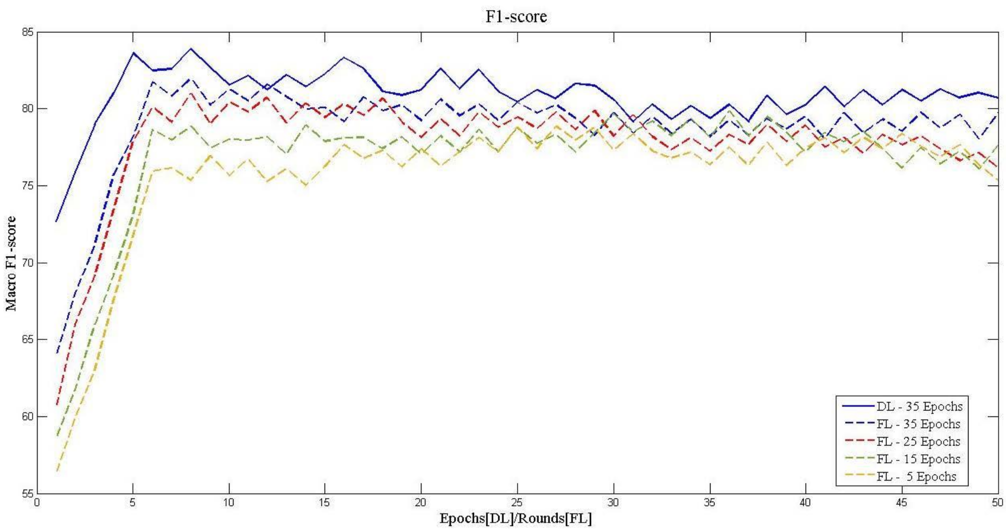

Figure 2: A comparison between FL and DL in terms of Full-featured dataset.

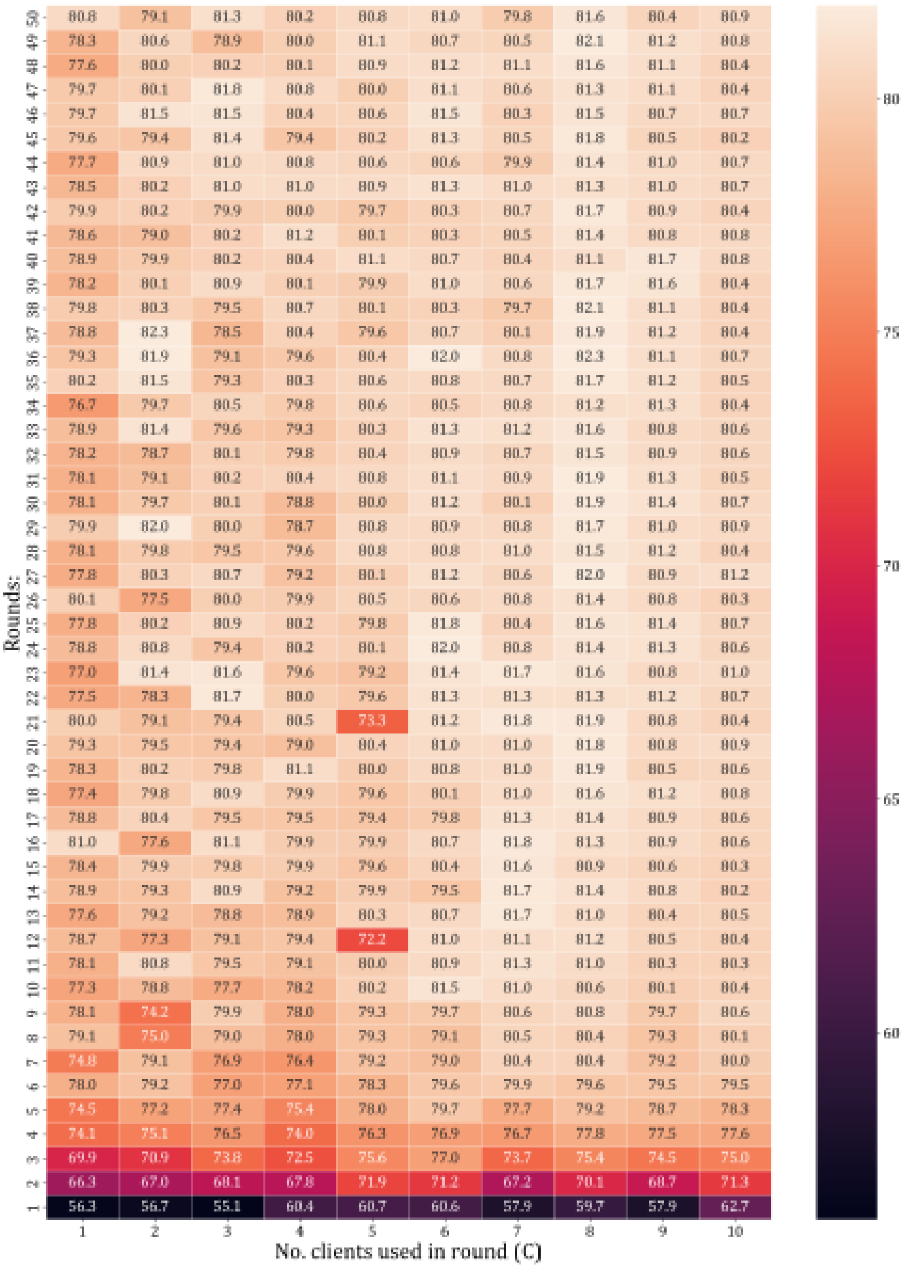

Figure 3: Macro F-scores achieved by an FL model when varying the number of clients used for training in each round and the total number of rounds used for training.

7 CONCLUSIONS

This paper presents the advantage that the FL strategy has in terms of IoT implementation, as well as the advantages that can be seen in terms of eHealth. The FL strategy is presented as a distributed collaborative approach to Artificial Intelligence - AI that is able to offer all of this to distributed IoT devices without the need for data sharing. The emphasis here is on the use of FL in scenarios used for HAR. What can be concluded is that the importance of the number of clients and local epochs for FL model training are key parameters in this experiment. The correct selection of the parameters, i.e. the optimal ones, leads to the fact that the performance of the FL strategy is quite close to that of the already existing DL strategies, which justifies the implementation of FL.

The future direction of research will continue to deepen the benefits and advantages that the FL strategy offers us, in terms of creating different models that will have different demands and testing their behavior in relation to DL.

ACKNOWLEDGMENTS

This work has been supported by the WideHealth project - European Union's Horizon 2020 research and innovation programme under grant agreement No. 952279. Additionally, Stefan Kalabakov would also like to thank the Slovene Human Resources Development and Scholarship Fund (Ad futura) for theirsupport.

REFERENCES

[1] A. Alsiddikya, W. Awwada, K. Bakarmana, H.Fouad, A. S.Hassanein, and A. M. Solimanc, "Priority-based data transmission using selective decision modes in wearable sensor based healthcare applications", Computer Communications, 1 July 2020, pp. 43-51.

[2] Z. Xiao, X. Xu, H. Xing, F. Song, X. Wang, and B. Zhao, "A federated learning system with enhanced feature extraction for human activity recognition", Knowl. Based Syst. 2021, vol. 229, p. 107338.

[3] L. Tu, X. Ouyang, J. Zhou, Y. He, and G. Xing, "FedDL: Federated Learning via Dynamic Layer Sharing for Human Activity Recognition", in Proceedings of the 19th ACM Conference on Embedded Networked Sensor Systems, Coimbra Portugal, 15-17 November 2021.

[4] Y. Liu, J. Peng, J. Kang, A. M. Iliyasu, D. Niyato (IEEE Member), and A. A. Abd El-Latif (IEEE Fellow), "A Secure Federated Learning Framework for 5G Networks", IEEE Wireless Communication Magazine, 12 May 2020.

[5] M. Wasilewska, H. Bogucka, and A. Kliks, "Federated Learning for 5G Radio Spectrum Sensing", Sensors 2022.

[6] K. Sozinov, V. Vlassov, and S. Girdzijauskas, "Human activity recognition using federated learning", in Proceedings of the 2018 IEEE Intl Conf on Parallel, Distributed Processing with Applications, Ubiquitous Computing Communications, Big Data Cloud Computing, Social Computing Networking, Sustainable Computing Communications.

[7] S. Ek, F. Portet, P. Lalanda, and G. Vega, "A Federated Learning Aggregation Algorithm for Pervasive Computing: Evaluation and Comparison", 2021 IEEE International Conference on Pervasive Computing and Communications (PerCom), 25 May 2021.

[8] C. Bettini, G. Civitarese, and R. Presotto, "Personalized Semi-Supervised Federated Learning for Human Activity Recognition", arXiv 2021, arXiv:2104.08094.

[9] L. Tu, X. Ouyang, J. Zhou, Y. He, and G. Xing, "FedDL: Federated Learning via Dynamic Layer Sharing for Human Activity Recognition", in Proceedings of the 19th ACM Conference on Embedded Networked Sensor Systems, Coimbra Portugal, 15-17 November 2021.

[10] C. Li, D. Niu, B. Jiang, X. Zuo, and J. Yang, "Meta-HAR: Federated Representation Learning for Human Activity Recognition", in Proceedings of theWeb Conference 2021 (WWW'21); Association for Computing Machinery: Ljubljana, Slovenia.

[11] G.K. Gudur and S.K. Perepu, "Resource- constrained federated learning with heterogeneous labels and models for human activity recognition", in Proceedings of the Deep Learning for Human Activity Recognition: Second InternationalWorkshop, DL-HAR 2020, Kyoto, Japan, 8 January 2021, Springer: Berlin/Heidelberg, Germany, 2021.

[12] X. Ouyang, Z. Xie, J. Zhou, J. Huang, and G. Xing, "ClusterFL: A Similarity-Aware Federated Learning System for Human Activity Recognition", MobiSys '21: Proceedings of the 19th Annual International Conference on Mobile Systems, Applications, and Services, June 2021.

[13] S. Kozina, H. Gjoreski, M. Gams, and M. Lustrek, "Three-layer Activity Recognition Combining Domain Knowledge and Meta- classification", Journal of Medical and Biological Engineering, vol. 33(4), pp. 406-414.

[14] R. Dastres and M. Soori, "Artificial Neural Network Systems", International Journal of Imaging and Robotics (IJIR), 2021, vol. 21 (2), pp.13-25.

[15] Al-Z. Malek, S. Almajali, and A. Awajan, "Experimental evaluation of a multi-layer feedforward artificial neural network classifier for network intrusion detection system", International Conference on New Trends in Computing Sciences (ICTCS) 2017.

[16] J. Singh and R. Banerjee, "A Study on Single and Multi-layer Perceptron Neural Network", 2019 3rd International Conference on Computing Methodologies and Communication (IC-CMC), 27-29 March 2019.

[17] Q. Li, Z. Wen, Z. Wu, and et al., "A Survey on Federated Learning Systems: Vision, Hype and Reality for Data Privacy and Protection", IEEE Transactions on Knowledge and Data Engineering (TKDE), 5 Dec 2021.