An Efficient Federated Learning Human Activity Recognition with CNN and Pyramid Pooling

Xian Wu Ting1, b), Ying Han Pang1, 2, a), Zheng You Lim2, c), Shih Yin Ooi1, 2, d) and Fu San Hiew3, e)

1Faculty of Information Science and Technology, Multimedia University, Jalan Ayer Keroh Lama, 75450 Bukit Beruang, Melaka, Malaysia

2Centre for Advanced Analytics, CoE for Artificial Intelligence, Multimedia University, Jalan Ayer Keroh Lama, 75450 Bukit Beruang, Melaka, Malaysia

3Infineon Technologies Sdn Bhd, Free Trade Zone, Batu Berendam, 75350 Melaka, Malaysia

a) Corresponding author: yhpang@mmu.edu.myb)ting.xian.wu@student.mmu.edu.my

c)lim.zhengyou@mmu.edu.my

d)syooi@mmu.edu.my

e)fusan.hiew@infineon.com

**Abstract.** Sensor-based Human Activity Recognition (HAR) has been widely used in several fields, including interactive gaming, healthcare, and activity monitoring. Deep learning-based models, particularly Convolutional Neural Networks (CNNs), have advanced HAR by enabling automatic feature learning from raw sensor signals. Nevertheless, these conventional CNNs are suboptimal for effectively extracting multiscale features, thus limiting their capability to accurately recognize diversified activities. Furthermore, centralized training is a requirement of traditional deep learning-based HAR systems. There is a greater chance of data leakage because user data is sent to a central server. By storing data on client devices for decentralized model training, Federated Learning (FL) offers a superior solution while maintaining privacy. We suggest FedPyramidCNN, which uses a one-dimensional CNN architecture with a pyramid pooling strategy to overcome the privacy and feature extraction issues. By capturing multiscale features, the pyramid pooling module enhances HAR performance. In order to give the data model rich information and help it discover both short-term and long-term dependencies concerning the inertial data, this module specifically extracts data patterns at various scales. According to the collected testing data, the FedPyramidCNN that was proposed obtains an accuracy of 96.85%.

# Introduction

Human Activity Recognition (HAR) systems are widely implemented in various applications, including health monitoring, fitness tracking, smart homes, and rehabilitation [1-3]. Deep Learning (DL) approach has become a popular approach for HAR model training because it automatically learns and analyzes features from raw inertial data [4, 5]. However, conventional DL methods rely on centralized model training, which aggregates raw sensor data from numerous users on a central server for data learning. This centralized approach has significant privacy concerns. Specifically, the requirement that raw data be transmitted to a central server increases the possibility of data leaks and unauthorized exposure [6–8]. Federated Learning (FL) has become a viable way to deal with these issues. In FL-based HAR systems, data processing and model training occur in a decentralized manner, ensuring sensitive information remains on local devices [9, 10].

In the literature, one-dimensional Convolutional Neural Networks (1D CNNs) have been applied to HAR [11-13]. However, conventional 1D CNNs are suboptimal in capturing multiscale feature representations. To address privacy and feature extraction challenges, in this paper, we propose a privacy-preserving federated learning based smartphone-enabled HAR system for decentralized training. We develop a 1D CNN extension with a pyramid pooling module named FedPyramidCNN for extracting short and long-term activity patterns through multiscale pooling.

Several approaches have been proposed to improve HAR. For example, Sharen et al. introduced WISNet, a CNN-based model that uses identity blocks for stable training, attention modules for better feature extraction, and a Convolved Normalized Pooled block for refined representations, to outperform conventional models, though it relies on centralized training [14]. In contrast, Liu et al. developed the Privacy-Protected Federated Personalized Random Forest (PP-FPRF) model, which uses locality-sensitive hashing to group similar users for training and employs ensemble learning with differential privacy to protect user data [15].

Ouyang et al. introduced ClusterFL, which is a clustering-aware federated learning framework for HAR with performance improved with reduced communication overhead utilizing data clustering for discarding slower or less correlated nodes, accelerating convergence [16]. Shen et al. introduced FedMAT for HAR, addressing the privacy challenge by treating every user as a unique task. FedMAT uses a shared network for shared features and tailored attention modules for personalized information, achieving a balance between generalization and personalization to serve multiple users better [17].

This study develops FedPyramidCNN, an enhanced 1D CNN-based model to improve local data training in a federated learning environment, motivated by the developments in federated learning. The architecture of the suggested method includes multiple clients, each with its own dataset, and uses a pyramid pooling strategy for local training to extract multiscale feature representation, as shown in Figure 1. A central server receives the generated model changes, aggregates them, and then sends the modified weights back to the clients. Because raw sensor data is stored on local devices, this decentralized procedure guarantees privacy preservation.

A diagram of a server

AI-generated content may be incorrect.

**Figure 1.** Architecture of FedPyramidCNN

The contributions made by this paper are:

* The development of a federated learning platform-based, privacy-preserving smartphone-based HAR system that decentralizes model training. While raw sensor data is stored on local devices, this model makes sure that the resulting model updates are shared.
* Incorporation of the pyramid pooling module in the deep learning architecture to produce a multiscale feature representation. With this, the model is able to capture data patterns through pyramid pooling operations of different scales.

# Methodology

## Dataset

The data utilized for the detection of human activity in this research is wireless sensor data mining (WISDM) [18]

The data is available at https://www.cis.fordham.edu/wisdm/dataset.php on the Wisdm website. It has six features: user, activity, timestamp, x-, y-, and z-accelerations, and 1,098,207 examples of different physical activities (sampled at 20 Hz). Walking, jogging, sitting, standing, and going upstairs and downstairs are the activities. The dataset is split into a 75 % training and a 25% testing set to facilitate model development and evaluation.

## Deep Learning Model

Due to data heterogeneity, Federated Learning in sensor-based HAR encounters challenges, which can degrade model performance. Traditional models such as 1D CNNs may struggle to extract the intrinsic complex features for reliable activity recognition. We introduce a federated learning HAR with an improved convolutional-based local training model (refer to Figure 2). In the proposed decentralized system, the convolutional feature extractor contains two 1D convolutional layers: one with 32 filters and another with 64 filters. Both of them are kernel size 4, followed by Batch Normalization, ReLU activation, and MaxPooling layers with a kernel size of 2 for downsampling the feature maps.

Each client trains the model using local inertial data. To effectively capture the multiscale features of the inertial data, the FedPyramidCNN incorporates the pyramid pooling module into its design. The changes made to the locally trained model are subsequently sent to a central server, where they are aggregated to create a global model. This collaborative learning enhances the generalizability of the global model while protecting data privacy by keeping the raw data locally.

A diagram of a process

AI-generated content may be incorrect.

**Figure 2.** Architecture of the model

FedPyramidCNN is able to capture distant and local features. It starts with two 4-sized kernel convolutional blocks, followed by batch normalization, ReLU, and 2-sized kernel max pooling. The pyramid pooling module then applies adaptive average pooling of scales 1, 2, and 4, stacks the outputs together, then flattens them and passes them through a fully connected layer with dropout, followed by a final classification layer.

The pyramid pooling module captures contextual information by applying adaptive average pooling at multiple scales [19]. It enhances feature extraction by capturing significant multiscale features from sensor data. Adaptive average pooling is applied at different scales, then each pooled output is upsampled to match the original feature map size before being concatenated. The original feature map is also included in the concatenation to preserve raw feature information alongside multiscale context. After concatenation, the feature map is flattened and processed by fully connected layers to produce the final classification output. Diversity of features enhances, helping distinguish between simultaneous activities and preservation of global and local contexts. In contrast to vanilla pooling, which can lose relevant features, pyramid pooling enriches the learned representation, avoiding overfitting and enhancing generalization.

## Model Aggregation

To address the deep learning challenge, aggregating weighting is used in global model aggregation to give equal contributions even in the case of non-IID and imbalanced data. Client contributions are weighted proportionally to their data sizes, ensuring that larger datasets have a stronger influence while maintaining model integrity. This decentralized aggregation technique ensures a well-representative global model across heterogeneous client environments. In the proposed model, the central server merges local updates proportionally to client data sizes [20] computing the global model iteratively.

|  |  |
| --- | --- |
|  | (1) |

where is the local model updates from client *k*, is the number of data samples on client *k*, and 𝑛 is the total number of data samples across all clients. The Equation (1) shows how the updates to the global model weight are performed by aggregating model weight updates contributed by each client, considering the contribution of data.

# Result

Table 1 summarizes the performance of the proposed FedPyramidCNN model alongside a baseline 1D CNN and other HAR approaches for comparison. Table 1 indicates the highest performance of FedPyramidCNN with an accuracy of 96.85%. This is due to the multiscale feature extraction through its pyramid pooling module that maintains local features while keeping the global context, unlike standard pooling that decreases the size of feature maps.

Jin et al. proposed a deep learning-based IoT system for post-stroke rehabilitation [21] achieving 94.49% accuracy. Although these models perform competitively, they require centralized data collection, which raises privacy concerns. FedPyramidCNN achieves higher accuracy in a federated learning environment without revealing unprocessed sensor data to centralized servers, thus enhancing data privacy.

Liu et al. developed a Privacy-Protected Federated Personalized Random Forest (PP-FPRF) [15], which attained 94.5% accuracy. On the other hand, some federated learning-based approaches showed lower performance. Sarkar et al. introduced GraFeHTy [22], a graph neural network that recorded 81.7% accuracy, while Yahya and Lee proposed Federated Model Contrastive Learning with Adaptive Control Variates [23] achieving 76.67% accuracy. In comparison, FedPyramidCNN leverages a pyramid pooling module to leverage multiscale feature extraction and thus forms a strong and privacy-supporting method for human activity recognition in decentralized settings.

**TABLE 1.** Accuracy performance between FedPyramidCNN and the existing method

|  |  |
| --- | --- |
| **System** | **Accuracy** |
| Deep Learning-Enhanced Internet [21] | 94.49 |
| PP-FPRF [15] | 94.5 |
| GraFeHTy [22] | 81.7 |
| FedCoad [23] | 76.67 |
| 1DCNN | 90.94 |
| FedPyramidCNN | 96.85 |

# CONCLUSION

This work devises a privacy-preserving smartphone-based HAR system by leveraging federated learning and a 1D CNN architecture enhanced with a pyramid pooling strategy. The proposed model, named FedPyramidCNN, decentralizes model training by keeping raw sensor data locally on devices. Furthermore, it enhances feature extraction through a pyramid pooling module based on adaptive average pooling to capture multiscale features. The proposed model is assessed using a publicly available dataset, and empirical results exhibit that FedPyramidCNN generally outperforms existing centralized deep learning models and other federated learning algorithms, achieving 96.85% accuracy. While this federated learning approach ensures privacy protection, it introduces additional communication costs due to the exchange of model updates between local clients and the global server. Future work will focus on exploring lightweight aggregation strategies to reduce dependence on the central server.

# Acknowledgement

Sincere gratitude is extended by the authors to the editor and all anonymous reviewers for their insightful comments that enhanced the manuscript. Multimedia University is the sponsor of the publication fund for this work..

# References

1. C. Shiranthika, N. Premakumara, H.-L. Chiu, H. Samani, C. Shyalika, and C.-Y. Yang, “Human Activity Recognition Using CNN & LSTM,” in *2020 5th International Conference on Information Technology Research (ICITR)*, (IEEE, Moratuwa, Sri Lanka, 2020), pp. 1–6.
2. M.M. van der Linden Thesis, prof P. dr Havinga J Klein Brinke, and dr Y. Miao, Federated Learning for Indoor Human Activity Recognition: Adapting to Changing Realistic Environments (2023).
3. D. Dayakaran, and N. Kadiresan, “Federated Learning Framework for Human Activity Recognition Using Smartphones,” Procedia Computer Science **235**, 2069–2078 (2024).
4. F. Mireshghallah, M. Taram, P. Vepakomma, A. Singh, R. Raskar, and H. Esmaeilzadeh, “Privacy in Deep Learning: A Survey,” (2020).
5. L. Deng, and D. Yu, “Deep Learning: Methods and Applications,” SIG **7**(3–4), 197–387 (2014).
6. S.E. Whang, Y. Roh, H. Song, and J.-G. Lee, “Data collection and quality challenges in deep learning: a data-centric AI perspective,” The VLDB Journal **32**(4), 791–813 (2023).
7. Q.W. Khan, A.N. Khan, A. Rizwan, R. Ahmad, S. Khan, and D.-H. Kim, “Decentralized Machine Learning Training: A Survey on Synchronization, Consolidation, and Topologies,” IEEE Access **11**, 68031–68050 (2023).
8. S. Abdulrahman, H. Tout, H. Ould-Slimane, A. Mourad, C. Talhi, and M. Guizani, “A Survey on Federated Learning: The Journey From Centralized to Distributed On-Site Learning and Beyond,” IEEE Internet Things J. **8**(7), 5476–5497 (2021).
9. Y. Liu, Y. Kang, T. Zou, Y. Pu, Y. He, X. Ye, Y. Ouyang, Y.-Q. Zhang, and Q. Yang, “Vertical Federated Learning: Concepts, Advances, and Challenges,” IEEE Trans. Knowl. Data Eng. **36**(7), 3615–3634 (2024).
10. K. Zhang, X. Song, C. Zhang, and S. Yu, “Challenges and future directions of secure federated learning: a survey,” Front. Comput. Sci. **16**(5), 165817 (2022).
11. A.O. Ige, and M. Sibiya, “State-of-the-Art in 1D Convolutional Neural Networks: A Survey,” IEEE Access 12, 144082–144105 (2024).
12. R. Lateef, and A. Abbas, “Tuning the Hyperparameters of the 1D CNN Model to Improve the Performance of Human Activity Recognition,” ETJ **40**(4), 547–554 (2022).
13. M.-K. Yi, W.-K. Lee, and S.O. Hwang, “A Human Activity Recognition Method Based on Lightweight Feature Extraction Combined With Pruned and Quantized CNN for Wearable Device,” IEEE Trans. Consumer Electron. **69**(3), 657–670 (2023).
14. H. Sharen, L. Jani Anbarasi, P. Rukmani, A.H. Gandomi, R. Neeraja, and M. Narendra, “WISNet: A deep neural network based human activity recognition system,” Expert Systems with Applications **258**, 124999 (2024).
15. S. Liu, J. Wang, and W. Zhang, “Federated personalized random forest for human activity recognition,” Mathematical Biosciences and Engineering 19(1), 953–971 (2022).
16. X. Ouyang, Z. Xie, J. Zhou, J. Huang, and G. Xing, “ClusterFL: a similarity-aware federated learning system for human activity recognition,” in *Proceedings of the 19th Annual International Conference on Mobile Systems, Applications, and Services*, (ACM, Virtual Event Wisconsin, 2021), pp. 54–66.
17. Q. Shen, H. Feng, R. Song, S. Teso, F. Giunchiglia, and H. Xu, “Federated Multi-Task Attention for Cross-Individual Human Activity Recognition,” in *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence*, (International Joint Conferences on Artificial Intelligence Organization, Vienna, Austria, 2022), pp. 3423–3429.
18. J.R. Kwapisz, G.M. Weiss, and S.A. Moore, “Activity recognition using cell phone accelerometers,” SIGKDD Explor. Newsl. 12(2), 74–82 (2011).
19. Y.-H. Wu, Y. Liu, X. Zhan, and M.-M. Cheng, “P2T: Pyramid Pooling Transformer for Scene Understanding,” IEEE Trans. Pattern Anal. Mach. Intell. **45**(11), 12760–12771 (2023).
20. H.B. McMahan, E. Moore, D. Ramage, S. Hampson, and B.A. y Arcas, “Communication-Efficient Learning of Deep Networks from Decentralized Data,” (2016).
21. F. Jin, M. Zou, X. Peng, H. Lei, and Y. Ren, “Deep Learning-Enhanced Internet of Things for Activity Recognition in Post-Stroke Rehabilitation,” IEEE J. Biomed. Health Inform. **28**(7), 3851–3859 (2024).
22. A. Sarkar, T. Sen, and A.K. Roy, “GraFeHTy: Graph Neural Network using Federated Learning for Human Activity Recognition,” in *2021 20th IEEE International Conference on Machine Learning and Applications (ICMLA)*, (IEEE, Pasadena, CA, USA, 2021), pp. 1124–1129..
23. I. Iwan, B. Yahya, and S. L. Lee, Federated model contrastive learning with adaptive control variates for human activity recognition. SSRN Paper No. 4810020 (2024).