EMBEDDINGS FOR MOTOR IMAGERY CLASSIFICATION

Quentin Langlois *Sébastien Jodogne*

Institute of Information and Communication Technologies, Electronics and
Applied Mathematics (ICTEAM), UCLouvain, 1348 Louvain-la-Neuve, Belgium

ABSTRACT

In recent years, deep learning techniques have been extensively applied to the analysis of electroencephalography (EEG) signals in various contexts, including motor imagery (MI) decoding. MI decoding has the potential for controlling devices or assisting in patient rehabilitation. A primary obstacle in implementing these algorithms in practical settings is the cost associated with model calibration, given their tendency to struggle with generalization to unseen subjects. To address this challenge, our work explores a new technique for representing EEG signals using embeddings. These embedding-based methods enable faster calibration without the need for model re-training. Experiments conducted on the Brain-Computer Interface Competition IV dataset 2a demonstrate the competitiveness of our approach compared to the state-of-the-art techniques in this field. It is also shown how embeddings offer additional opportunities to tackle EEG analysis challenges.

***Index Terms*—** Deep Learning, Embeddings, Neurophysiological Signals, Brain Computer Interface

1. INTRODUCTION

Motor Imagery (MI) classification is a key aspect of the Brain-Computer Interface (BCI) domain, where the objective is to decode the actions imagined by a subject without any physical movement [1]. This decoding is typically conducted by relying on the automated analysis of electroencephalograms (EEG). EEG captures electrical signals from electrodes placed on the scalp, which represents the brain activity of a subject. MI decoding can be applied to a wide range of practical applications, not only in the medical domain to aid in patient rehabilitation, but also for healthy subjects in video games or in device control.

In recent years, Deep Learning (DL) models have demonstrated impressive capabilities to analyze raw signal data [2]. This has prompted multiple research work aimed at applying DL models to EEG signal analysis, showing improved performance with respect to earlier Machine Learning (ML) techniques. Nevertheless, EEG signals are inherently non-stationary, which means that the signal statistics vary not only between subjects but also across longitudinal analyses for the same subject [3]. This reduces the performance of models on the long-term usage, as well as their generalization capabilities to unseen subjects.

To address the aforementioned challenges, researchers have explored the transfer learning between subjects [4] and the learning of new tasks using few-shot learning [5]. However, such methodologies represent a non-trivial procedure, as they require model fine-tuning, which may require a dedicated infrastructure and an adaptation of the model architecture.

In this paper, an alternative calibration strategy based on embeddings is explored, enabling continuous learning without the need for fine-tuning DL models [6]. This study serves as an initial exploration, introducing a novel approach for generating EEG embeddings. Our primary objective is to demonstrate the potential of the proposed pipeline compared to conventional state-of-the-art approaches. To this end, experiments have been conducted on the Brain-Computer Interface Competition IV, dataset 2a (BCI-IV 2a) [7], a commonly used benchmark for evaluating MI decoding techniques. The proposed embedding-based approach, built upon the ATC-Net [8] architecture, has been compared against its original implementation as well as against previous state-of-the-art models. The models have been evaluated in both the *within-subject* and *leave-one-subject-out* (LOSO) standard scenarios. The results demonstrate the competitiveness of the proposed method when compared to traditional methodologies. Moreover, embeddings facilitate the calibration of the models to new subjects without requiring fine-tuning. Additional opportunities to further leverage the potential of embeddings for EEG are also highlighted.

2. MATERIALS AND METHODS

This section introduces and discusses recent DL techniques used for EEG analysis, including MI classification, along with the experimental setup.

Quentin Langlois is funded by the Walloon Region through a F.R.S.-FNRS FRIA (Fund for Research training in Industry and Agriculture) grant.

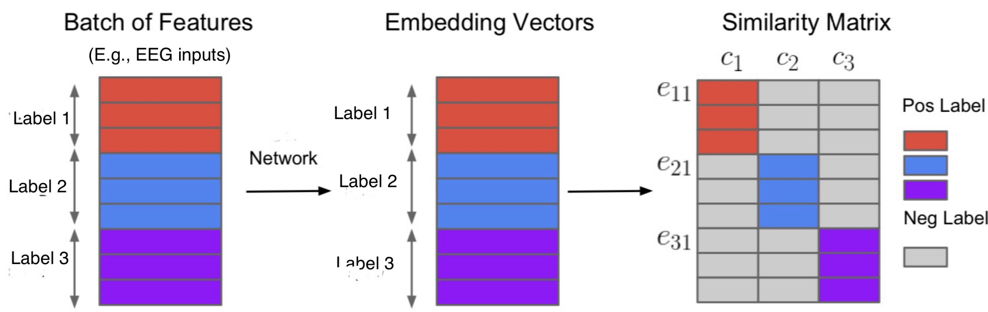

Fig. 1. GE2E loss matrix computation (adapted from the original paper [9]).

2.1. Data and evaluation

This study utilizes the BCI-IV 2a dataset [7] to assess the proposed technique in comparison to state-of-the-art models tailored for this task. The dataset contains 2 recording sessions of 9 subjects. Each session contains 72 trials of 7 seconds for each of the 4 MI classes (left/right hand, tongue, and feet), leading to a total of 288 trials per session. The input signal is collected with a 22-channels EEG headset, and 3 channels of electrooculography (EOG). The EOG channels have not been used for fair reproduction and comparison to the state-of-the-art methods used in this work.

To evaluate the models, the within-subject and the LOSO methodologies have been used, as described in the ATCNet paper [8]. On the one hand, the within-subject scenario is designed to assess long-term usage. It involves training the model with data from the first recording session of a subject and then evaluating its performance using data from the second session. On the other hand, the LOSO scenario evaluates the generalization capability to new subjects. This is done by training the model on both recording sessions of 8 out of the 9 subjects and then testing it on the second session of the subject that was excluded from training. The first recording session is occasionally employed for calibration, in which instance the methodology is referred to as *partial* LOSO.

2.2. DL for EEG classification

In recent years, multiple architectures combining convolutional neural networks (CNN) with recurrent neural networks (RNN) have been proposed, as reviewed in the EEGNex study [10]. These models are mainly composed of two core blocks: an EEGNet model [11], which combines the multiple channels with depth-wise 2D convolution, and a Temporal Convolution network (TCN) module [8], which is a regular 1D-CNN network focusing on temporal information of the signal. More recently, multi-scale fusion (such as TCNet-Fusion [12]) and Transformers-based models (like ATCNet [8, 13]) have extended these two core blocks, resulting in significant improvements in MI classification performance. The main limitation of these methods is their poor generalization capabilities to new subjects and their need for subject-specific fine-tuning to achieve qualitative results.

This paper leverages the ATCNet architecture, which has been independently re-implemented and validated, and extends it to the embeddings paradigm. For this purpose, the last classification layer is replaced by a n-dimensional linear layer. As a result, the model generates n-dimensional embeddings instead of a score for each label. The other models serve as baseline comparisons to evaluate the competitiveness of the proposed approach.

2.3. Embeddings for EEG analysis

To enhance generalization performance in the LOSO scenario, a pipeline combining embeddings with the EEGNet model has recently been proposed [4]. The primary limitation of this pipeline is the use of a subject-specific linear layer to classify new subject data. Consequently, this approach is closer to transfer learning with a fine-tuning strategy, like the one proposed by N. Mammone et al. [5], rather than to true embeddings.

In addition, a CNN architecture trained to represent medical EEG recordings with embeddings has been introduced [6]. Its training procedure relies on the triplet-loss [14] function and uses a large-scale medical EEG corpus [15]. The primary drawback of this approach is its intensive sampling strategy, which arises from the large amount of data involved. Specifically, the triplet-loss operates on triplets of data: the ”anchor”, a ”positive”, and a ”negative” sample. The loss aims to minimize (resp. maximizes) the distance between the embeddings of the anchor and the positive (resp. negative) samples. This results in an exponential number of candidate data triplets, leading to an aggressive sampling strategy that fails to fully exploit all available data.

2.4. Generalized End-to-End Loss

To mitigate the limitation of the triplet-loss, we propose to use the generalized end-to-end (GE2E) loss [9], which ex-

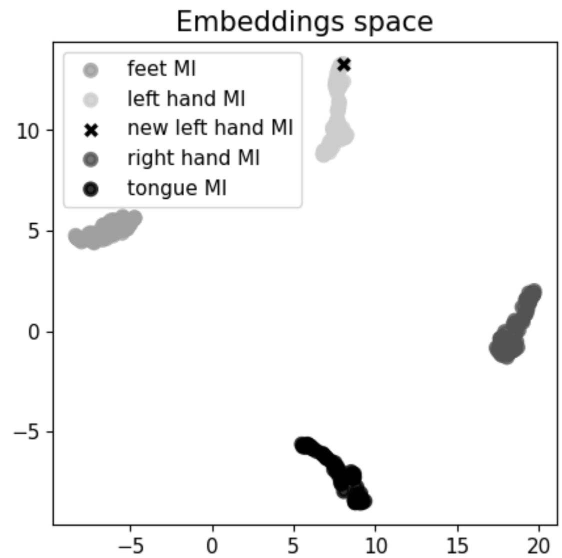

Fig. 2. UMAP [17] projection of reference points embeddings. The "x" represents a new data to classify that belongs to the left-hand MI class.

tends the triplet-loss mechanism by computing distance minimization/maximization per batch. The loss function operates on a batch of N samples of M distinct labels and computes the centroid for each label. The centroid for the i-th label, denoted c_i, is computed as follows:

$$c_i = \frac{1}{N} \cdot \sum_{j=0}^{N} e_{ij}, \quad (1)$$

where e_{ij} is the j-th embedding vector of label i. The distance matrix between each input and each centroid is computed, where d_{ij} denotes the distance between the j-th embedding of label i (i.e., e_{ij}) and the i-th centroid c_i, as illustrated in Figure 1. Finally, the distance between each embedding and its positive centroid is minimized, while being maximized for the other centroids.

The input format for the GE2E loss is more constrained in comparison to the triplet-loss, as the GE2E loss requires N samples of M labels at each batch. In this work, a custom data generation procedure, initially designed for speech verification, has been adapted [16]. The generator performs r iterations on all the available labels, and randomly selects N samples for each label. Each sample can therefore be selected multiple times, while being associated with various positive/negative samples. This procedure aims to maximize the number of samples seen during training, while tackling possible label unbalancing.

2.5. Training procedure

This section describes the hyperparameters used to train and evaluate the different models. The training process and implementation of models draw inspiration from the latest 2023 update of the Altaheri et al. [8] open project, which was independently reproduced for this study. The evaluation pipeline leverages a train/validation/test split that differs from their original paper, to be more aligned with the BCI-IV competition standard. This modification accounts for the variation in performance when compared to the results documented in the reference paper. However, these results align with those achieved using Altaheri et al.'s updated implementation if using the adjusted train/validation/test split.

All models have been trained for a maximum of 1000 epochs, with an early-stopping criterion that has a patience of 300 and a batch size of 64. The optimizer used is Adam [18] with an initial learning-rate of 0.001, which is reduced by a factor of 0.9 upon reaching a plateau. The preprocessing has been modified to perform per-trial, per-channel normalization by mean and standard deviation. This modification has a marginal impact (less than 1%) on performance, while being less data-dependent compared to the originally used global normalization. The results reported in this document are obtained using a 5-fold cross-validation evaluation strategy. This helps to mitigate the impact of randomness in model initialization and in the selection of training/validation data[1].

The prediction step for an embeddings-based model leverages a second ML classifier to classify the produced embedding for a given data. In this work, the Nearest Centroid (NC) classifier, a variant of the k-Nearest Neighbor ($k-$NN) classifier, has been used, as it is fast and easy to train. The NC classifier computes the centroid for each label, based on the reference points. To classify a new data point, it takes the label of the centroid with the lowest distance. This behavior is closer to the GE2E objective function compared to the standard k-NN classifier. The reference points provided to the NC classifier, which are used to compute the centroid of each label, consist in the embedded training samples used to train the DL model. This corresponds to 230 (resp. 3680) samples in the within-subject (resp. LOSO) scenario, balanced between the 4 labels. Figure 2 gives a comprehensive example of embeddings classification. In the partial LOSO scenario (i.e., with model calibration), the first recording session of the left-out subject is added to the NC points to calibrate it to the new subject.

3. RESULTS AND DISCUSSION

This section presents the experimental results by comparing state-of-the art architectures with the proposed embeddings-based approach. Additionally, a first attempt of model calibration without fine-tuning is explored. Finally, potential op-

[1] The test set is always the second session of the tested subject.

Model	Within subject Accuracy (%)	LOSO Accuracy (%)
ATCNet [8]	**80.7 (+/- 9.8)**	**60.6 (+/- 13.4)**
ATCNet embeddings (*ours*)	79.3 (+/- 10.3)	58.9 (+/- 12.7)
TCNet-Fusion [12]	71.4 (+/- 10.9)	**60.8 (+/- 12.5)**
EEGNet [11]	69.9 (+/- 12.6)	58.8 (+/- 13.2)
EEG-TCNet [19]	66.7 (+/- 11.4)	59.2 (+/- 13.4)
EEGNex [10]	62.7 (+/- 6.87)	59.2 (+/- 6.0)

Table 1. Performance metrics in the within-subject and LOSO scenarios, without calibration.

Model	Full LOSO Accuracy (%)	Partial LOSO Accuracy (%)
ATCNet [8]	**60.6 (+/- 13.4)**	**60.6 (+/- 13.4)**
ATCNet embeddings (*ours*)	58.9 (+/- 12.7)	**60.7 (+/- 12.8)**

Table 2. Impact of calibration on the proposed method in the partial LOSO paradigm. *Left:* without calibration (full LOSO), which is an excerpt of Table 1. *Right:* with calibration (partial LOSO).

portunities for future work to expand upon this technique are discussed.

3.1. General results

Table 1 presents the general results, which highlights the competitiveness of the proposed approach when compared to the original ATCNet classifier. In both scenarios (within-subject and LOSO), the performance difference between the two methods is less than 2%. This indicates that both techniques are nearly equivalent in terms of performance. Additionally, the proposed method surpasses older techniques in the within-subject scenario and achieves similar performances in the LOSO scenario. These results demonstrate the potential of embeddings for EEG analysis.

3.2. Impact of calibration

Table 2 illustrates the impact of integrating subject-specific embeddings into the NC classifier. This integration simply involves gathering additional data and does not entail any fine-tuning of the DL model. This is why the original ATCNet classifier remains unaffected, as fine-tuning would have been necessary for calibration. Despite the limited improvement in performance, these results are promising for continuous learning strategies. Combining this technique with an initial fine-tuning offers a significant opportunity to address the non-stationary nature of EEG for cross-session applications. Further elaboration on this topic is provided in the subsequent section.

3.3. Open science

To facilitate the replication and future expansion of the proposed method, this study adheres to the principles of open science by providing open and free access to the source code and models[2]. The source code leverages the new multi-backend Keras 3 framework [20]. Moreover, the source code contains an independent re-implementation of the state-of-the art models evaluated in this work. The training and evaluation procedures have been optimized using the Python primitives for multiprocessing, which enables faster evaluation of the models under the k-fold cross-validation strategy.

3.4. Discussion and future work

This work serves as a proof-of-concept, aiming to demonstrate the competitiveness of embeddings compared to classifier methods in EEG analysis. Embeddings offer multiple benefits in tackling EEG analysis challenges, surpassing some limitations of the traditional approach. In this section, we propose several future research areas to further explore and exploit the potential benefits of embeddings in the context of MI decoding.

3.4.1. Continuous learning

As discussed in Section 3.2, the NC calibration offers limited improvements in performance. This indicates that further fine-tuning of the DL model remains necessary to attain performance levels comparable to those achieved in the within-subject scenario. However, maintaining consistent model performance across multiple sessions poses a significant challenge in EEG analysis. To tackle this issue, a promising strategy consists in combining an initial subject-specific fine-tuning of the DL model with a simpler NC calibration during subsequent sessions. This approach would facilitate continuous learning and adaptation without requiring extensive model refinement, rendering it more lightweight and practical. The Large Scale BCI database [21] is an ideal resource to evaluate this strategy, offering up to 4 sessions for the same subject for a given task. Consequently, exploiting the Large Scale BCI database would enable a comprehensive assessment of model performance across sessions.

3.4.2. "Label-free" approach

In BCI, and more broadly in EEG analysis, the ability to detect and adapt to new, unseen labels is crucial. This capability has diverse applications, including detecting new diseases in the medical field and dynamically incorporating new actions into application control. Traditional classifier approaches face

[2]Source code is available at: `https://forge.uclouvain.be/QuentinLanglois/mlsp-2024-embeddings-for-motor-imagery`

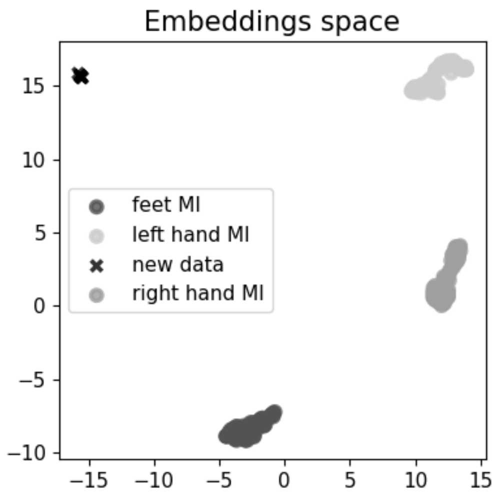

Fig. 3. Example for the label-free property. In this case, the new data point should be associated with a new cluster corresponding to the ”tongue MI” class.

limitations in this aspect. Classifiers are constrained to classify new data based on the initial classes defined during their training. Incorporating new labels thus requires adapting the architecture and fine-tuning it [5]. This contrasts with embeddings models that are trained to represent EEG signals using embedding vectors, regardless of their label. Therefore, integrating new, unseen labels only requires simple calibration of the NC classifier, as already demonstrated in the context of sensor-based activity recognition [22]. Figure 3 depicts the presence of 3 clusters of known labels, along with some new data to classify located far from each known cluster. The thresholding technique proposed by Yu et al. [22] allows to automatically associate these new data points with a new cluster.

3.4.3. Multi-task Federated Learning

In the Federated Learning (FL) paradigm [23], various entities, such as hospitals, industries, or individuals, train a local copy of a global model using their own private data. Once the local training is completed, the updated models are combined onto a central server, and the aggregated model is distributed back to the participants. This iterative procedure allows one to benefit from training the model on a large amount of data, without any data exchange, which enhances data privacy. One major limitation in deploying such systems in the context of EEG analysis is the necessity for all participants to define and share a fixed set of labels to initialize the shared, global classification model.

Embeddings represent a strong candidate for addressing this limitation, thanks to the label-free property discussed in the Section 3.4.2. Indeed, embeddings enable the definition of a global model without predefined labels. As a result, each participant can train local copies of the model on their own data, regardless of their specific classification task. This approach would offer increased flexibility and could potentially foster stronger collaboration among participants in a FL coalition, as they could contribute to the training of a global model without the need for predefined task-related labels.

4. CONCLUSION

This study introduces deep learning architectures for the automated analysis of EEG signals using embeddings. It is demonstrated that embeddings offer an alternative to traditional classifier approaches on the task of MI decoding. The experiments conducted on the widely used BCI-IV benchmark dataset 2a have demonstrated that embeddings are competitive compared to the state-of-the art models, in both within-subject and leave-one-subject out scenarios. To further explore the potential of embeddings, a first attempt of model calibration without fine-tuning has been evaluated, highlighting the potential for continuous learning strategies. Finally, further discussions have been proposed to explore the potential of embeddings in addressing the non-stationary nature of EEG signals.

5. REFERENCES

[1] Theo Mulder, “Motor imagery and action observation: Cognitive tools for rehabilitation,” *Journal of neural transmission (Vienna, Austria : 1996)*, vol. 114, pp. 1265–78, 02 2007.

[2] Yann LeCun, Y. Bengio, and Geoffrey Hinton, “Deep learning,” *Nature*, vol. 521, pp. 436–44, 05 2015.

[3] Alexandre Gramfort, Daniel Strohmeier, Jens Haueisen, Matti Hämäläinen, and Matthieu Kowalski, “Time-frequency mixed-norm estimates: Sparse M/EEG imaging with non-stationary source activations,” *NeuroImage*, vol. 70, 01 2013.

[4] Pierre Guetschel, Thoedore Papadopoulo, and Michael Tangermann, “Embedding neurophysiological signals,” in *2022 IEEE International Conference on Metrology for Extended Reality, Artificial Intelligence and Neural Engineering (MetroXRAINE)*, 2022, pp. 169–174.

[5] Nadia Mammone, Cosimo Ieracitano, Rossella Spataro, Christoph Guger, Woosang Cho, and Francesco Morabito, “A few-shot transfer learning approach for motion intention decoding from electroencephalographic signals,” *International Journal of Neural Systems*, vol. 34, no. 02, pp. 2350068, 2024, PMID: 38073546.

[6] R. Thiyagarajan, C. Curro, and S. Keene, “A learned embedding space for EEG signal clustering,” in *2017*

IEEE Signal Processing in Medicine and Biology Symposium (SPMB), 2017, pp. 1–4.

[7] Michael Tangermann, Klaus-Robert Müller, Ad Aertsen, Niels Birbaumer, Christoph Braun, Clemens Brunner, Robert Leeb, Carsten Mehring, Kai Miller, Gernot Müller-Putz, Guido Nolte, Gert Pfurtscheller, Hubert Preissl, Gerwin Schalk, Alois Schlögl, Carmen Vidaurre, Stephan Waldert, and Benjamin Blankertz, "Review of the BCI competition IV," *Frontiers in neuroscience*, vol. 6, pp. 55, 07 2012.

[8] Hamdi Altaheri, Ghulam Muhammad, and Mansour Alsulaiman, "Physics-informed attention temporal convolutional network for EEG-based motor imagery classification," *IEEE Transactions on Industrial Informatics*, vol. 19, no. 2, pp. 2249–2258, 2023.

[9] Li Wan, Quan Wang, Alan Papir, and Ignacio Lopez Moreno, "Generalized end-to-end loss for speaker verification," 2017.

[10] Xia Chen, Xiangbin Teng, Han Chen, Yafeng Pan, and Philipp Geyer, "Toward reliable signals decoding for electroencephalogram: A benchmark study to eegnex," *Biomedical Signal Processing and Control*, vol. 87, pp. 105475, 01 2024.

[11] Vernon Lawhern, Amelia Solon, Nicholas Waytowich, Stephen Gordon, Chou Hung, and Brent Lance, "EEGNet: A compact convolutional network for eeg-based brain-computer interfaces," *Journal of Neural Engineering*, vol. 15, 11 2016.

[12] Yazeed Musallam, Nasser AlFassam, Ghulam Muhammad, Syed Amin, Mansour Alsulaiman, Wadood Abdul, Hamdi Altaheri, Mohamed Bencherif, and Mohammed Algabri, "Electroencephalography-based motor imagery classification using temporal convolutional network fusion," *Biomedical Signal Processing and Control*, vol. 69, pp. 102826, 08 2021.

[13] Hamdi Altaheri, Ghulam Muhammad, and Mansour Alsulaiman, "Dynamic convolution with multilevel attention for EEG-based motor imagery decoding," *IEEE Internet of Things Journal*, vol. 10, no. 21, pp. 18579–18588, 2023.

[14] Elad Hoffer and Nir Ailon, *Deep Metric Learning Using Triplet Network*, p. 84–92, Springer International Publishing, 2015.

[15] Iyad Obeid and Joseph Picone, "The temple university hospital EEG data corpus," *Frontiers in Neuroscience*, vol. 10, 05 2016.

[16] Quentin Langlois and Sébastien Jodogne, "Practical study of deep learning models for speech synthesis," in *Proceedings of the 16th International Conference on PErvasive Technologies Related to Assistive Environments*, New York, NY, USA, 2023, PETRA '23, p. 700–706, Association for Computing Machinery.

[17] Leland McInnes, John Healy, Nathaniel Saul, and Lukas Grossberger, "UMAP: Uniform manifold approximation and projection," *The Journal of Open Source Software*, vol. 3, no. 29, pp. 861, 2018.

[18] Diederik Kingma and Jimmy Ba, "Adam: A method for stochastic optimization," *International Conference on Learning Representations*, 12 2014.

[19] Thorir Ingolfsson, Michael Hersche, Xiaying Wang, Nobuaki Kobayashi, Lukas Cavigelli, and Luca Benini, "EEG-TCNet: An accurate temporal convolutional network for embedded motor-imagery brain-machine interfaces," 2020.

[20] François Chollet et al., "Keras," `https://keras.io`, 2015.

[21] Murat Kaya, Mustafa Binli, Erkan Ozbay, Hilmi Yanar, and Yuriy Mishchenko, "A large electroencephalographic motor imagery dataset for electroencephalographic brain computer interfaces," *Scientific Data*, vol. 5, pp. 180211, 10 2018.

[22] Hongzheng Yu, Zekai Chen, Xiao Zhang, Xu Chen, Fuzhen Zhuang, Hui Xiong, and Xiuzhen Cheng, "FedHAR: Semi-supervised online learning for personalized federated human activity recognition," *IEEE Transactions on Mobile Computing*, vol. 22, no. 6, pp. 3318–3332, 2023.

[23] Chen Zhang, Yu Xie, Hang Bai, Bin Yu, Weihong Li, and Yuan Gao, "A survey on federated learning," *Knowledge-Based Systems*, vol. 216, pp. 106775, 2021.