Internship proposal: Riemannian federated learning for EEG decoding with heterogeneous data

1 Context

Electroencephalography (EEG) records the brain activity of a subject through electrodes placed on the scalp. EEG decoding is a classification problem: a decoder maps each multi-channel trial, a few seconds of recorded signal, to a class. In motor imagery, the classes correspond to imagined movements, such as moving the left or the right hand.

EEG signals generally have a low signal-to-noise ratio and can vary strongly from one subject to another. A decoder therefore needs many trials to generalise, while a single subject provides a few hundred at most. Combining the recordings of many subjects, often collected at several sites such as hospitals or laboratories, is thus desirable. These recordings, however, contain medical information, which regulation and confidentiality prevent from pooling. Federated learning (FL) fits this setting [McMahan et al., 2017]. Without sharing any data, several sites collaboratively train a common model under the coordination of a server. Over several communication rounds, sites train the model on their own data and send back their parameters, which the server aggregates. The model thus learns from all recordings, while each recording stays on its site. FL has already been applied to EEG decoding [Ju et al., 2020; Liu et al., 2024].

The architecture of the model to be federated depends on the local network used on each client; this network must be lightweight to avoid costly communications. One way to develop these networks is to transform the raw data during preprocessing. In EEG, it is common to do this using the spatial covariance matrix. These matrices are symmetric positive definite (SPD) and belong to a Riemannian manifold. Riemannian classifiers can operate directly on covariance matrices and respect their geometry [Barachant et al., 2012]. They are robust across subjects and remain lightweight. However, performance can be significantly improved by applying Riemannian neural networks to these covariance matrices, such as SPDNet [Huang and Van Gool, 2017], which mimics a deep learning network.

Federating these networks requires adapting both the local training and the server aggregation to the geometry. Client updates must stay on the manifold, and the average of the client weights must be mapped back onto it. The team has proposed FedSPDNet, a federated SPDNet with geometry-aware aggregation [Pautrel et al., 2026a], and has studied it on several motor-imagery datasets, in standard and personalised FL [Pautrel et al., 2026b], with the Euclidean EEGNet as a baseline. These studies rely on public datasets from the MOABB benchmark [Jayaram and Barachant, 2018].

These studies assume that within a dataset, all trials are recorded with a common pattern, so all covariance matrices have the same size.

2 Problem statement

This assumption often fails in practice, for two reasons. First, each site may use its own headset, with a different number of electrodes. Second, faulty electrodes are often discarded from some recordings. Covariance matrices then differ in size across clients, and a common model becomes inapplicable. Restricting all clients to their shared electrodes restores a common size, but discards a large part of the recorded information [Mellot et al., 2024].

Electrode positions follow international standards where each electrode has a name and

a fixed location on the scalp. For each client, we thus know which electrodes of a reference pattern it holds, and where they lie. This knowledge suggests two strategies. The first learns an input layer specific to each client, which maps its data to a representation of common size [Ju et al., 2020; Liu et al., 2024]. The second expands each client's covariance matrices to the reference pattern. The entries of the missing electrodes are set to a fixed block [Rodrigues et al., 2021], interpolated from neighbouring electrodes [Mellot et al., 2024], or estimated from the observed data [Hippert-Ferrer et al., 2022]. These works only give partial answers. None of them combines FL, a Riemannian network and heterogeneous patterns. Integrating and comparing both strategies in FedSPDNet thus remains open.

3 Objectives

The internship aims to train a federated SPDNet efficiently when clients provide covariance matrices of heterogeneous sizes. It will also study how the strategy chosen to handle this size heterogeneity affects performance.

1. **Controlled setting.** Assign each subject of a dense-pattern MOABB dataset a subset of electrodes, so that each strategy can be compared with the full-pattern case.
2. **Methods.** In the first strategy, each client keeps its own input layer, which maps its covariance matrices to a common size, and only the following layers are aggregated. In the second strategy, each client expands its matrices to the reference pattern, estimating the missing entries from the electrode positions or from statistics shared across clients, and all layers are aggregated.
3. **Comparison.** Compare both strategies in accuracy, convergence speed and communication cost, under standard and personalised FL, with EEGNet as a Euclidean reference.
4. **Theory and generalisation.** Use the experimental findings to formulate the general problem of FL with SPD inputs of heterogeneous sizes, and, depending on the candidate's profile, study how missing electrodes degrade the learned representation.

The internship builds on the team's existing code and on public MOABB data.

Possible extensions

Depending on progress, two directions may be explored.

Pre-training and transfer. A federated model is trained on a source dataset, recorded with a rich headset. It is then reused on a target dataset whose electrodes and classes are subsets of those of the source, for instance a lighter headset with fewer motor-imagery tasks. The model then handles the missing electrodes with the strategies developed during the internship, and keeps only the outputs of the target classes. The goal is to reduce the number of trials needed to calibrate a new subject.

Multimodal learning. Other data often come with EEG, such as eye activity (EOG), functional brain imaging (fMRI) or text metadata. These heterogeneous inputs need to be processed separately, then combined, so that the decoder benefits from each modality. This calls for new architectures that merge data of different natures, in a federated setting where clients may hold different subsets of modalities.

4 Requirements

Master or engineering student in applied mathematics, machine learning or signal processing, with a solid background in linear algebra, optimisation and statistics. Good coding skills in

Python, ideally with PyTorch. Prior experience with Riemannian geometry, federated learning or EEG is a plus.

5 Internship details

Location	L2S laboratory, CentraleSupélec, Université Paris-Saclay, Gif-sur-Yvette
Duration	5–6 months, starting ideally in March 2027
Supervisors	Thibault Pautrel, thibault.pautrel@centralesupelec.fr Florent Bouchard, florent.bouchard@centralesupelec.fr Guillaume Ginolhac, guillaume.ginolhac@univ-smb.fr
To apply	Send a CV and a short cover letter to the supervisors.

References

Barachant, A., Bonnet, S., Congedo, M., and Jutten, C. (2012). Multiclass brain–computer interface classification by Riemannian geometry. *IEEE Transactions on Biomedical Engineering*, 59(4):920–928.

Hippert-Ferrer, A., Mian, A., Bouchard, F., and Pascal, F. (2022). Riemannian classification of EEG signals with missing values. In *30th European Signal Processing Conference (EUSIPCO)*.

Huang, Z. and Van Gool, L. (2017). A Riemannian network for SPD matrix learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 2036–2042.

Jayaram, V. and Barachant, A. (2018). MOABB: trustworthy algorithm benchmarking for BCIs. *Journal of Neural Engineering*, 15(6):066011.

Ju, C., Gao, D., Mane, R., Tan, B., Liu, Y., and Guan, C. (2020). Federated transfer learning for EEG signal classification. In *42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 3040–3045.

Liu, R., Chen, Y., Li, A., Ding, Y., Yu, H., and Guan, C. (2024). Aggregating intrinsic information to enhance BCI performance through federated learning. *Neural Networks*, 172:106100.

McMahan, H. B., Moore, E., Ramage, D., Hampson, S., and Agüera y Arcas, B. (2017). Communication-efficient learning of deep networks from decentralized data. In *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS)*, volume 54 of *PMLR*, pages 1273–1282.

Mellot, A., Collas, A., Chevallier, S., Engemann, D., and Gramfort, A. (2024). Physics-informed and unsupervised Riemannian domain adaptation for machine learning on heterogeneous EEG datasets. In *32nd European Signal Processing Conference (EUSIPCO)*, pages 1367–1371.

Pautrel, T., Bouchard, F., Mian, A., and Ginolhac, G. (2026a). FedSPDnet: geometry-aware federated deep learning with SPDnet. In *34th European Signal Processing Conference (EUSIPCO)*. Accepted for publication. Preprint: arXiv:2604.22494.

Pautrel, T., Bouchard, F., Mian, A., and Ginolhac, G. (2026b). Personalised federated learning for Riemannian and Euclidean EEG decoding. Preprint: arXiv:2609.29037.

Rodrigues, P. L. C., Congedo, M., and Jutten, C. (2021). Dimensionality transcending: a method for merging BCI datasets with different dimensionalities. *IEEE Transactions on Biomedical Engineering*, 68(2):673–684.