

Contrastive Self-Supervised Learning for Motor Imagery: impact of the embedding size

V. Marissens Cueva^{1*}, L. Bougrain¹

¹Université de Lorraine, CNRS, LORIA, Neuro rhythms Team, 54000, Nancy, France

*LORIA, Campus Scientifique, BP 239, 54506 Vandœuvre-lès-Nancy, France, E-mail: valerie.marissens@loria.fr

Introduction: Due to the intra- and inter-individual variability of the electroencephalography (EEG) signals, brain-computer interfaces (BCI) require a daily user-specific calibration. This offline calibration step is necessary to set feature extraction, classification and pre-processing parameters. Yet, it is time consuming and might cause fatigue before the actual use of the BCI. Our goal is to reduce this time with a self-supervised classification method that achieves good detections with minimal calibration trials, for use in a motor imagery (MI)-based BCI that aims to enhance the rehabilitation of stroke patients. To process a small amount of labeled data, self-supervised learning (SSL) is currently the state-of-the-art method in the fields of vision and natural language processing [1], which makes it interesting to explore for EEG data.

Material, Methods and Results: Dataset 2a of the BCI competition IV [2] was used to estimate the capability of contrastive SSL. Two sessions of 72 trials each are available for training and testing. The classifier has to detect a right (or left) hand MI relative to a resting period. The approach uses a pretext task to create sample pairs from unlabeled EEG segments that are similar (close) or dissimilar (far) in time. SSL projects them in an embedding space accordingly, then reuses it to solve the real task. Our pretext task is based on Relative Positioning (RP) [3]. For T trials, SSL-RP produces 2T pairs of similar EEG windows if they belong to the same segment, and 4T(2T-1) pairs of dissimilar ones if they come from different segments, of which 2T are randomly selected. Segments are related to resting or MI periods. The feature extractor is EEGNet [4] without its classification layer, and both pretext and real task classifiers are logistic regressions. Fig. 1 presents the accuracy of the SSL-RP models among different percentages of the training set, as the number of features extracted, i.e., the size d of the embedding space, varies. The process was averaged across 10 repetitions. SSL-RP is compared to CSP+LDA with 6 filters.

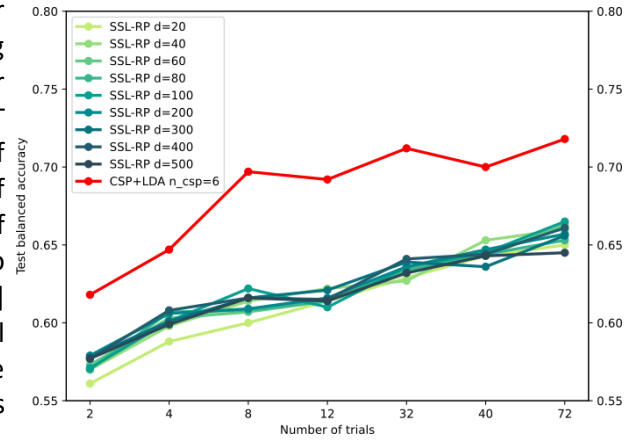


Figure 1. Test accuracy, averaged across all subjects, for the detections of MI vs rest (BCI Competition IV dataset 2a), obtained by SSL-RP models, as a function of the percentage of labeled training data. The embedding size is noted as d .

Discussion and Significance: As the number of training trials increases, the accuracy of SSL-RP models improves. The accuracy for $d=200$ is better than for $d=500$, meaning that the performance saturates as the number of features extracted d increases. An ANOVA test ($p < 0.001$) considers the different models statistically equal due to the small sample size. CSP+LDA shows slightly higher accuracies than SSL-RP, but is also statistically equal to the SSL-RP models. The negative pairs of EEG windows may be from the same class, which makes it difficult for a logistic regressor to associate them.

Acknowledgements: This work has been supported by the Agence Nationale de la Recherche (ANR), under grant ANR-19-CE33-0007 (project Grasp-IT). Experiments presented in this paper were carried out using the Grid'5000 testbed, supported by a scientific interest group hosted by Inria and including CNRS, RENATER and several Universities as well as other organizations (see <https://www.grid5000.fr>).

References:

- [1] Bommasani R et al. On the Opportunities and Risks of Foundation Models, 2021.
- [2] Tangermann M et al. Review of the BCI Competition IV. *Frontiers in Neuroscience*, vol. 6, 2012.
- [3] Banville H et al. Uncovering the structure of clinical EEG signals with self-supervised learning, 2020.
- [4] Lawhern VJ et al. EEGNet: a compact convolutional neural network for EEG-based brain-computer interfaces. *Journal of Neural Engineering*, vol. 15, no. 5, p. 056013, 2018.

Contrastive self-supervised learning for motor imagery: impact of the embedding size

Valérie Marissens Cueva^{1*}, Laurent Bougrain¹

¹Neurorhythms team, Université de Lorraine, CNRS, LORIA, Vandœuvre-lès-Nancy, France

* Email: valerie.marissens@loria.fr

INTRODUCTION

Due to the intra- and inter-individual variability of the electroencephalography (EEG) signals, brain-computer interfaces (BCI) require a daily user-specific calibration. This offline calibration step is necessary to set feature extraction, classification and pre-processing parameters, yet it restricts users from directly using the BCI and might induce fatigue.

Self-Supervised Learning (SSL) is currently the state-of-the-art method in the fields of vision and natural language processing [1]. It learns the features of a small amount of labeled data by generating a significant number of self-labeled data.

- Can SSL improve the accuracy obtained by the standard method Common Spatial Pattern with Linear Discriminant Analysis (**CSP+LDA**) for motor imagery (MI) detection with a **few training trials** by generating more?
- What is the **impact of the embedding size**?

MATERIAL

Dataset 2a of the BCI competition IV:

- 9 healthy subjects, 22 electrodes, sampling rate of 250 Hz
- 4 MI tasks, but only right and left hand MI were used for this study
- 2 sessions (training and testing) composed of 6 runs, each of 12 trials per MI task
- From 2 to 72 trials were used to learn the model

METHODS

SSL-RP uses a pretext task called **Relative Positioning (RP)** [2], **EEGNet** [3] as embedder, and logistic regressions as classifiers for both pretext and downstream tasks :

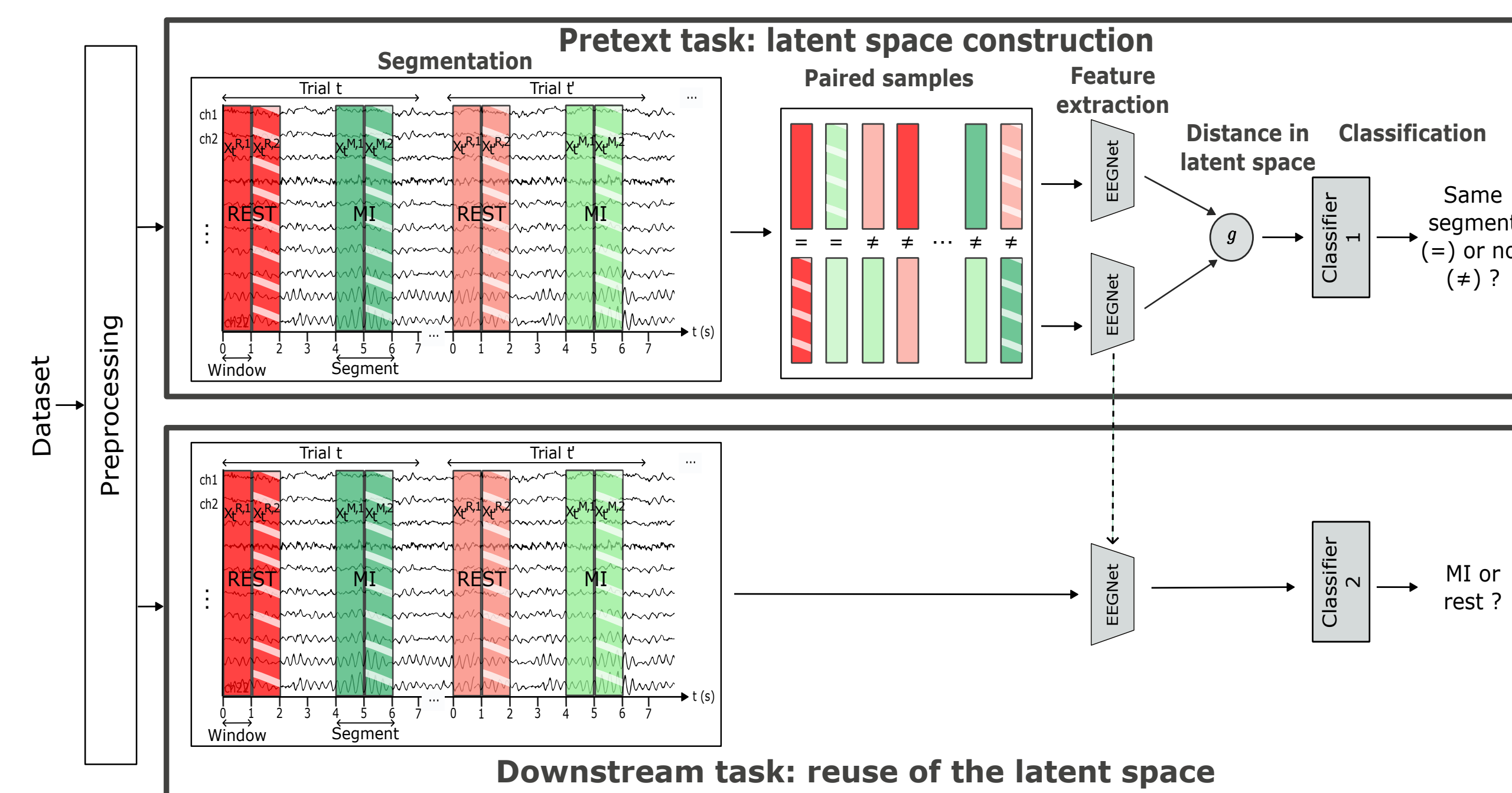


Fig. 1: Structure of the SSL-RP model (figure inspired by [4] and adapted for this MI window sampling). MI and rest segments are selected from each trial and divided into windows. Positive and negative pairs of windows are used for the pretext task, to train the features' extractor (here EEGNet). During the downstream task, EEGNet is reused to project the data, where the second classifier will determine if the window corresponds to MI or rest class.

RESULTS: pretext task

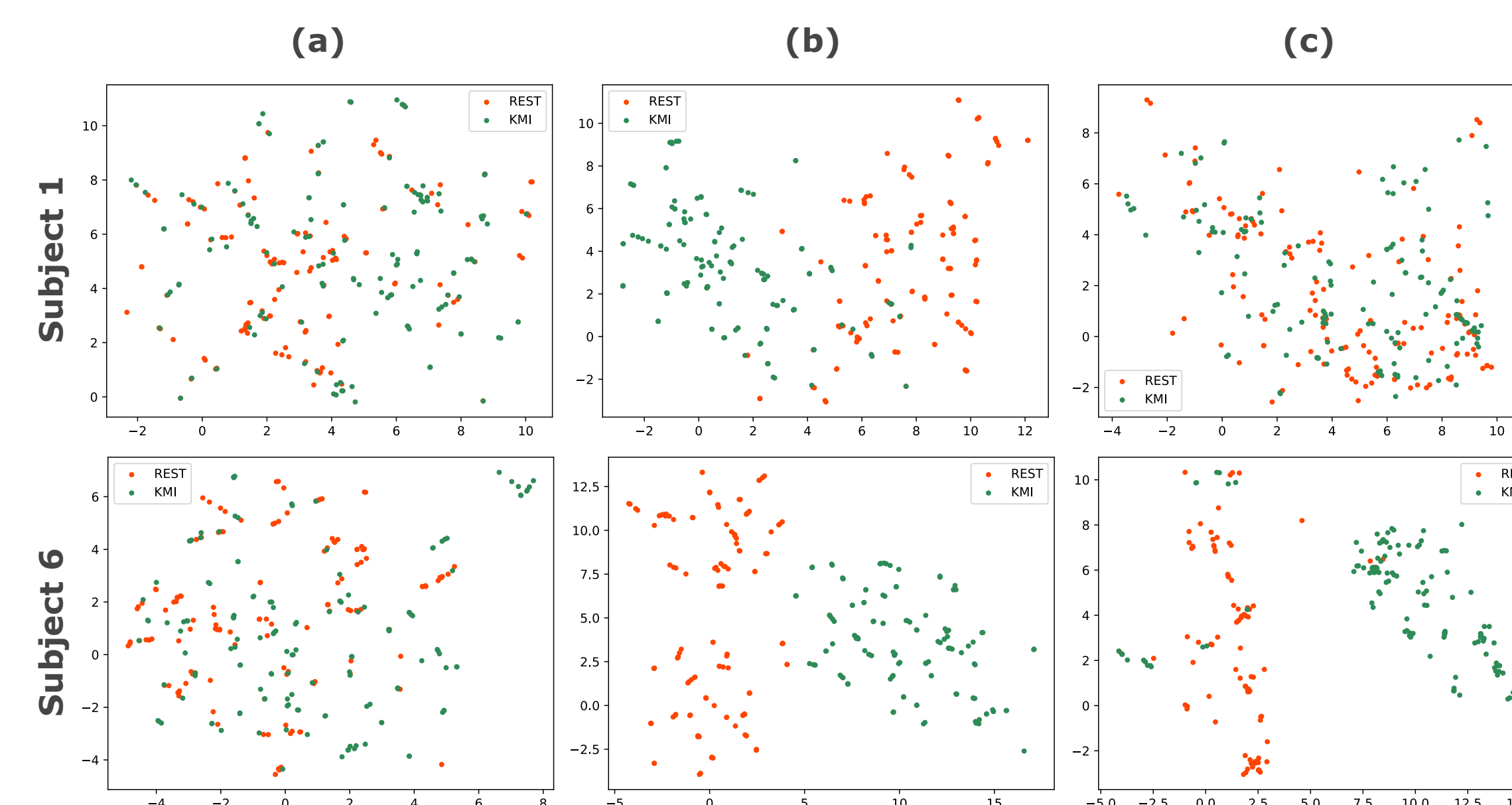


Fig. 2: Projections in a 2D-space of the latent space, obtained by the UMAP [5] algorithm when 72 trials were used by SSL-RP with $d=100$ features extracted, for subject 1 (lower accuracy) and subject 6 (higher accuracy). The projections correspond to the training set (a) before the pretext task, (b) after the pretext task, and (c) to the testing set after the pretext task.

- The pretext task enabled the embedder to create a **useful latent space** that could effectively distinguish between MI and rest without relying on the actual labels
- This latent space may **not always be generalizable**.

RESULTS: downstream task

Comparison of SSL-RP models, for different embedding sizes d , with CSP+LDA:

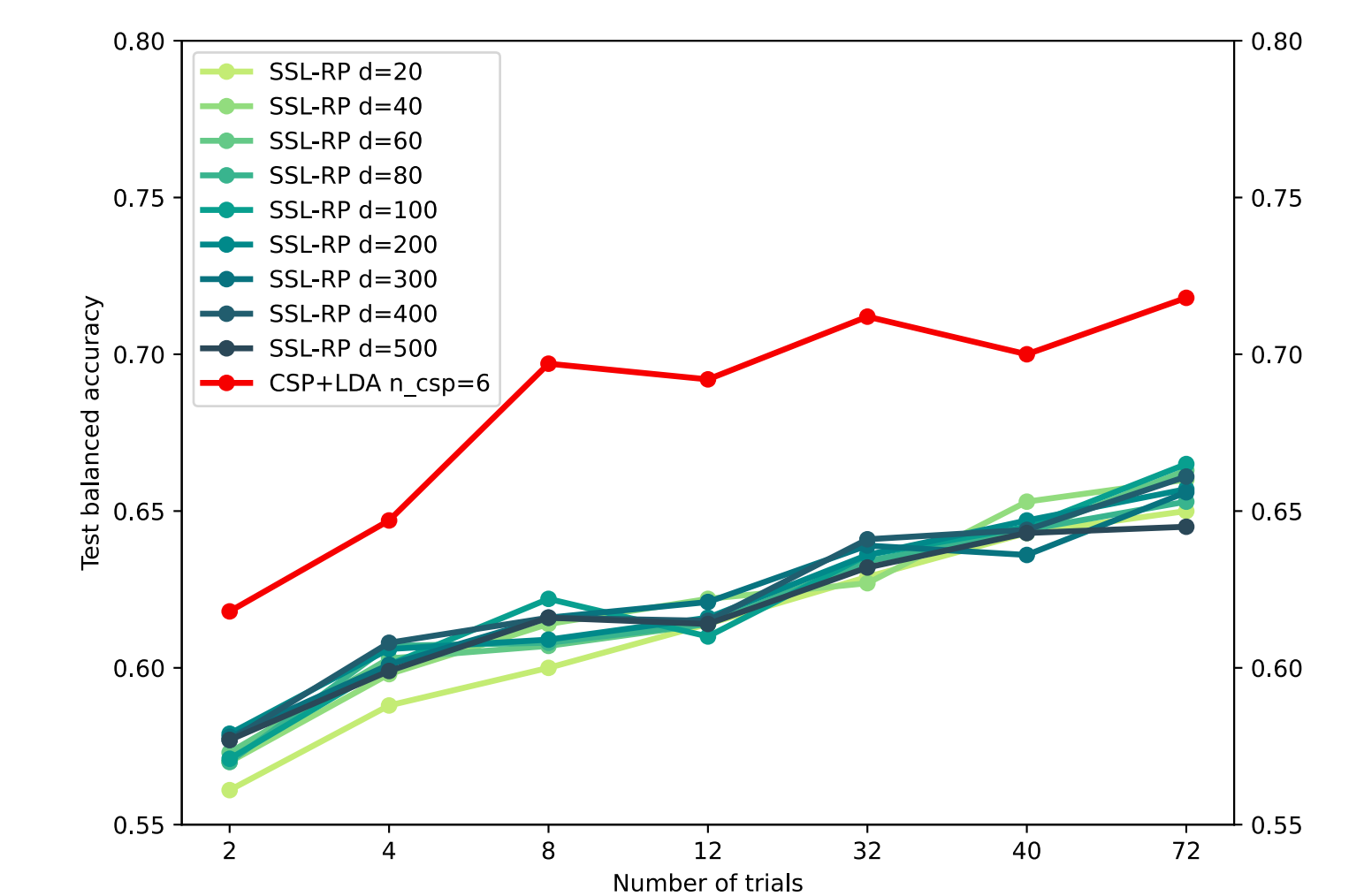


Fig. 3: Test accuracy averaged across all 9 subjects, for the detections of right or left hand MI vs rest, according to the number of training trials, obtained by the models (Self-Supervised Learning-Relative Positioning for various embedding sizes d and Common Spatial Pattern with a Linear Discriminant Analysis). The sequential training trials were randomly selected and the process was averaged for 10 repetitions.

Impact of the number of training trials:

- For all models, 12 training trials seem to be enough to calibrate the BCI.

Impact of the embedding sizes:

- According to an ANOVA test with $p < 0.001$, all models are statistically equivalent to one another, for each number of training trials.
- The performance saturates as the number of features extracted d reaches 200, and $d=40$ seems to be a good lower limit.

CONCLUSIONS

- CSP+LDA has better performance than SSL-RP overall by 4.5% when only 2 training trials are used, up to 8.5% with 8. The **negative pairs of EEG windows may be from the same class**, which makes it difficult for a logistic regressor to associate them.
- Increasing the embedding size does not significantly enhance the recognition of MI periods.

REFERENCES

- [1] R. Bommasani et al. *On the Opportunities and Risks of Foundation Models*, 2021.
- [2] H. Banville et al. *Uncovering the structure of clinical EEG signals with self-supervised learning*, 2020.

- [3] V. J. Lawhern et al. *EEGNet: a compact convolutional neural network for EEG-based brain-computer interfaces*, 2018.
- [4] Y. Ou et al. *An improved self-supervised learning for EEG classification*, 2022.
- [5] L. McInnes et al. *UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction*, 2018.

ACKNOWLEDGMENTS

The authors acknowledge the support of the French National Research Agency (ANR), under grant ANR-19-CE33-0007 (project Grasp-IT). The experiments were carried out using the Grid'5000 testbed, supported by a scientific interestgroup hosted by Inria and including CNRS, RENATER and several Universities as well as other organizations (see <https://www.grid5000.fr>).