Defining Out-of-Distribution detection for EEG-BCIs

Ivo Pascal de Jong*, Matias Valdenegro-Toro, Andreea Ioana Sburlea
Bernoulli Institute, University of Groningen, Groningen, The Netherlands
*Nijenborgh 9, 9747 AG, Groningen, The Netherlands, E-mail: ivo.de.jong@rug.nl

Introduction: Out-of-Distribution (OoD) detection methods are variations on Machine Learning (ML) models that can detect when a sample is different from the data the model is trained on [1]. OoD is minimally explored in BCI research [2], but it has the potential to make a BCI system that detects problems and delivers corrective action. OoD detection may detect: variabilities that cause decreases in performance, bugs in preprocessing that cause bad predictions, destructive artifacts that obfuscate the signal, and even off-task neural patterns. Traditional ML would give arbitrary guesses in these cases.
The broad definition of OoD makes the effectiveness of general OoD-detection hard to measure. In Computer Vision research this definition is worked out into established benchmarks covering a broad range of causes. For evaluating universal OoD detection in BCIs we should define BCI-specific causes that generate OoD data. This requires an employable definition for OoD specific to (EEG-based) BCIs. Our aim is to create a dataset of a broad range of artifacts, variabilities and other problems that can arise in BCIs that should be detectable by a good OoD detection system. At the 2025 BCI Meeting, we will establish with the community a shared definition of what OoD data in BCIs really is. We will then continue to develop a matching benchmark to evaluate various OoD detection methods.

Material, Methods and Results: As an example we consider a common OoD detection task called *leave-one-class-out*. Here, a model is trained on $C-1$ classes, and is tested on all C classes. It should report high uncertainty for the excluded class, which would represent off-task behaviour as OoD. We previously presented results on (non-BCI) datasets with various models, where the uncertainty is higher for the excluded class. This allowed us to detect them with AUROC scores between 0.65 and 0.98 [4]. However, in Fig. 1 we show that in a within-subject four-class Motor Imagery classification task [5] these methods fail with an AUROC score of ~ 0.5.

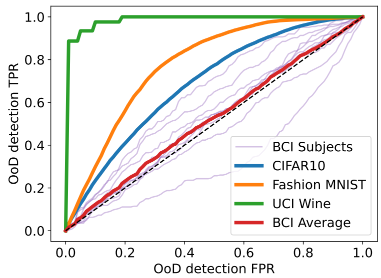

Figure 1: *Leave-one-class-out OoD detection works well on non-BCI datasets, but not for four-class Motor Imagery.*

Leave-one-class-out OoD detection is a moderately difficult OoD detection task. Some large artifacts or bugs in preprocessing may be more easily detected, whereas variabilities could be more difficult. A good and community-supported understanding of what OoD detection can mean for BCIs is needed to construct proper benchmarks

Conclusion: OoD detection identifies when traditional ML methods may fail. However, it has a very broad definition. To be able to work with OoD detection and advance the BCI field, our community needs to define an agreed set of examples that are to be worked out into benchmarks. We aim to implement such benchmarks in an open-source repository using public datasets. We already have plans for defining OoD data (including off-task behavior, movement artifacts, disconnected electrodes, hardware and preprocessing differences, and cross-subject BCIs), but by getting input from the community we can create a complete and appropriate benchmark. A definition of OoD will include a better understanding of variabilities [6], which is an established need in BCI research.

References:

[1] Yang J, Zhou K, Li Y, Liu Z. Generalized out-of-distribution detection: A survey. In *International Journal of Computer Vision*, 2024.

[2] de Jong IP, Sburlea AI, Valdenegro-Toro MA. Uncertainty Quantification in Machine Learning for Biosignal Applications–A Review. *arXiv preprint arXiv:2312.09454*, 2023.

[3] Manivannan P, de Jong IP, Valdenegro-Toro M, Sburlea AI. Uncertainty Quantification for cross-subject Motor Imagery classification. In *9th Graz Brain-Computer Interface Conference*, 2024.

[4] de Jong IP, Sburlea AI, Valdenegro-Toro MA. How disentangled are your classification uncertainties?. *arXiv preprint arXiv:2408.12175*, 2024.

[5] Brunner C, Leeb R, Müller-Putz G, Schlögl A, Pfurtscheller G. BCI Competition 2008–Graz data set A. *Institute for knowledge discovery (laboratory of brain-computer interfaces), Graz University of Technology*, 2008.

[6] Riascos J, Molinas M, Lotte F. Machine Learning Methods for BCI: challenges, pitfalls and promises. In *ESANN 2024-European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*, 2024.