Orthogonal Contrastive Learning
for Multi-Representation fMRI Analysis

Tony Muhammad Yousefnezhad
Learning By Machine
National Bank of Canada[1]
Edmonton AB Canada
Email: tony@learningbymachine.com
Website: https://www.yousefnezhad.com/

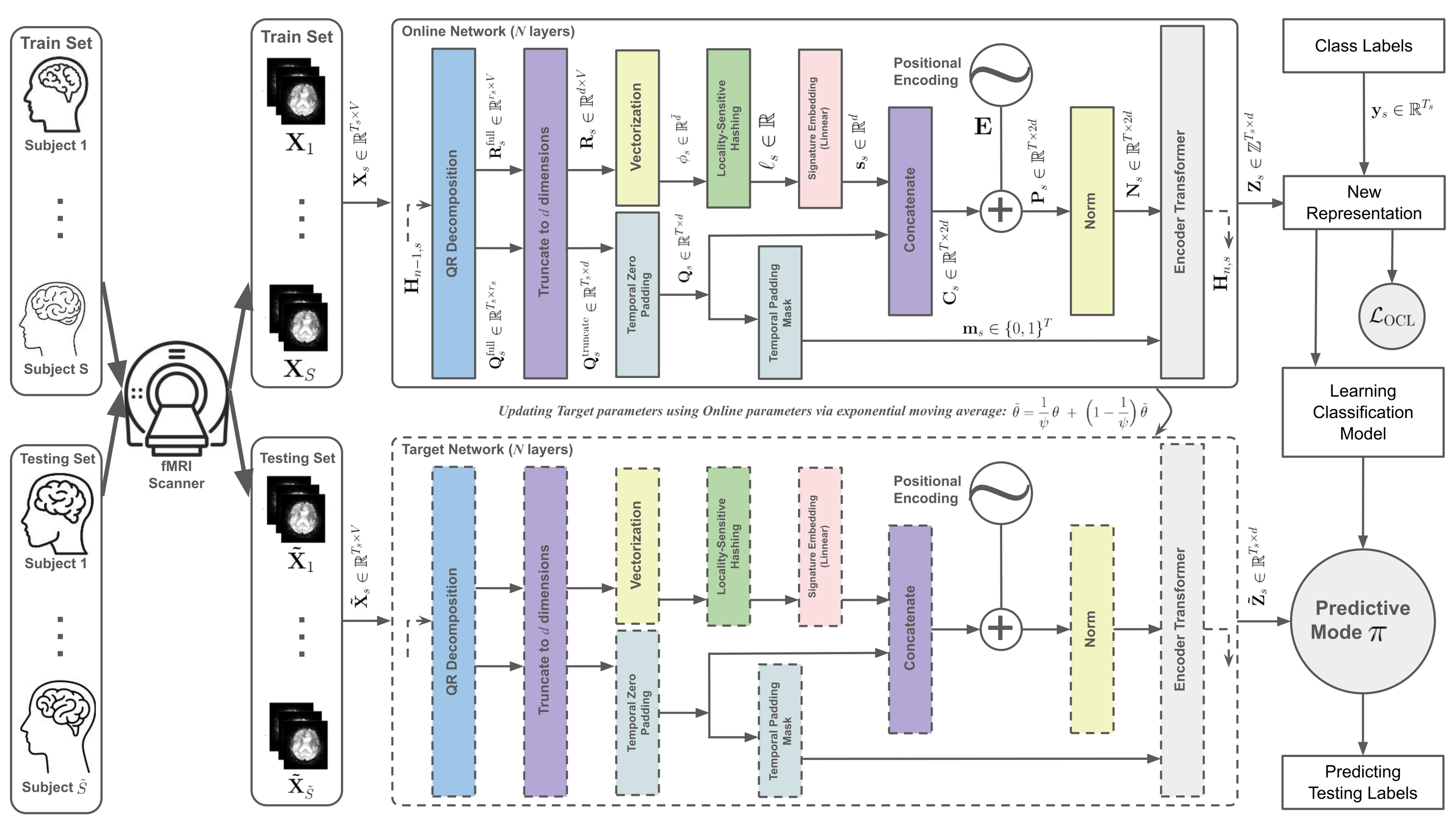
Motivation

Understanding brain activity during cognitive tasks is a key goal in neuroscience. Task-based fMRI enables this but faces **major challenges**—low signal-noise ratio, high-dimensionality, small sample sizes, and variability across subjects and sites.

We introduce **Orthogonal Contrastive Learning (OCL)**, a unified framework that aligns multi-subject fMRI data **without requiring temporal** alignment or equal sequence lengths. OCL integrates **QR-based orthogonalization, locality-sensitive hashing, positional encoding, and transformer encoders** to learn robust, stimulus-aligned embeddings.

Trained through **contrastive learning and unsupervised pretraining** on fMRI-like synthetic data, OCL achieves state-of-the-art performance in both alignment quality and downstream classification across **multi-subject** and **multi-site** datasets.

The Proposed Orthogonal Contrastive Learning



❖ Locality-sensitive hashing (LSH): $\ell_s = \mathrm{lsh}_{\{\mathbf{a},b,w\}}(\phi_s) = \left\lfloor \dfrac{\langle \mathbf{a}, \phi_s \rangle + b}{w} \right\rfloor$

❖ Positional Encoding:

$$\mathbf{E} \in \mathbb{R}^{T \times 2d} = [e_{1,0}, \ldots, e_{T,2d}], \quad e_{t,2i} = \sin\!\left(\frac{t}{T^{2i/(2d)}}\right), \quad e_{t,2i+1} = \cos\!\left(\frac{t}{T^{2i/(2d)}}\right)$$

❖ OCL objective function:

$$\mathcal{L}_{\mathrm{OCL}\{\tau,\mu,\lambda\}}(\mathbf{Z}_s, \mathbf{y}_s) = -\frac{1}{T_s} \sum_{i=1}^{T_s} \log \frac{\sum_{j=1,\; j\neq i,\; y_{s,j}=y_{s,i}}^{T_s} \exp\left(\langle z_{s,i}, z_{s,j} \rangle / \tau\right)}{\sum_{k=1}^{T_s} \exp\left(\langle z_{s,i}, z_{s,k} \rangle / \tau\right)}$$

$$+\lambda \frac{1}{T_s^2} \sum_{i=1}^{T_s} \sum_{j=1,\; y_{s,j}\neq y_{s,i}}^{T_s} \log\left(1 + \exp\left(\langle z_{s,i}, z_{s,j} \rangle / \tau - \mu\right)\right).$$

❖ To initialize OCL on a multi-site setup, we aggregate the B learned sites via

$$\tilde{\theta}_{\mathrm{sites}} = \frac{1}{B} \sum_{b=1}^{B} \tilde{\theta}_b$$

Experiments

❖ The fMRI datasets

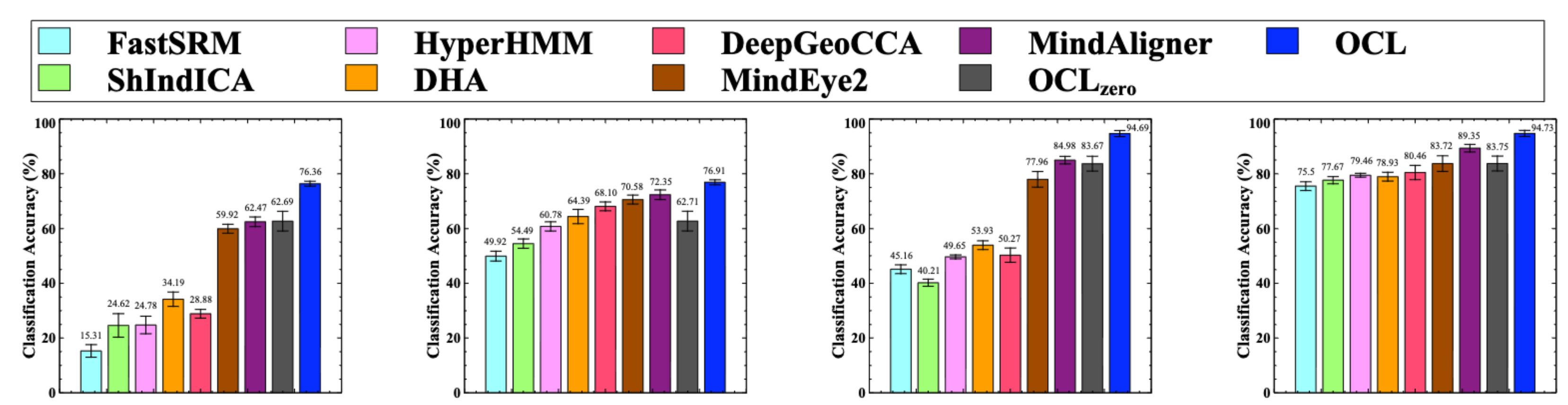
ID	Title	Type	S	$\lvert\mathbf{y}\rvert$	T_s	*Site(#)*
A*	Stop signal (DS007)	Decision	20	4	472	B (3)
B	Conditional stop signal (DS008)	Decision	13	4	317	A (1)
CMU	Meanings of Nouns	Semantic	9	12	402	
C	Simon task (DS101)	Simon	21	2	302	D (1)
D	Flanker task (DS102)	Flanker	26	2	292	C (1)
DS232	Face-coding models with individual-face	Visual	10	4	760	
E	Integration of sweet taste: Study 1 (DS229)	Flavour	15	6	580	F (1)
F	Integration of sweet taste: Study 3 (DS231)	Flavour	9	6	650	E (1)
Forrest	Forrest Gump movie	Visual	20	10	451	
Raiders	Raiders movie	Visual	10	7	924	

S is the number of subjects; $|\mathbf{y}|$ is the number of stimulus categories; T_s is the number of time points per subject; *Site* lists the other datasets whose neural responses can be transferred to this dataset. # represents the number of sites in the corresponding dataset. * this dataset is partitioned into three independent 'sites'—pseudo-word naming (A1), letter naming (A2), and manual response (A3)

❖ We benchmark OCL against

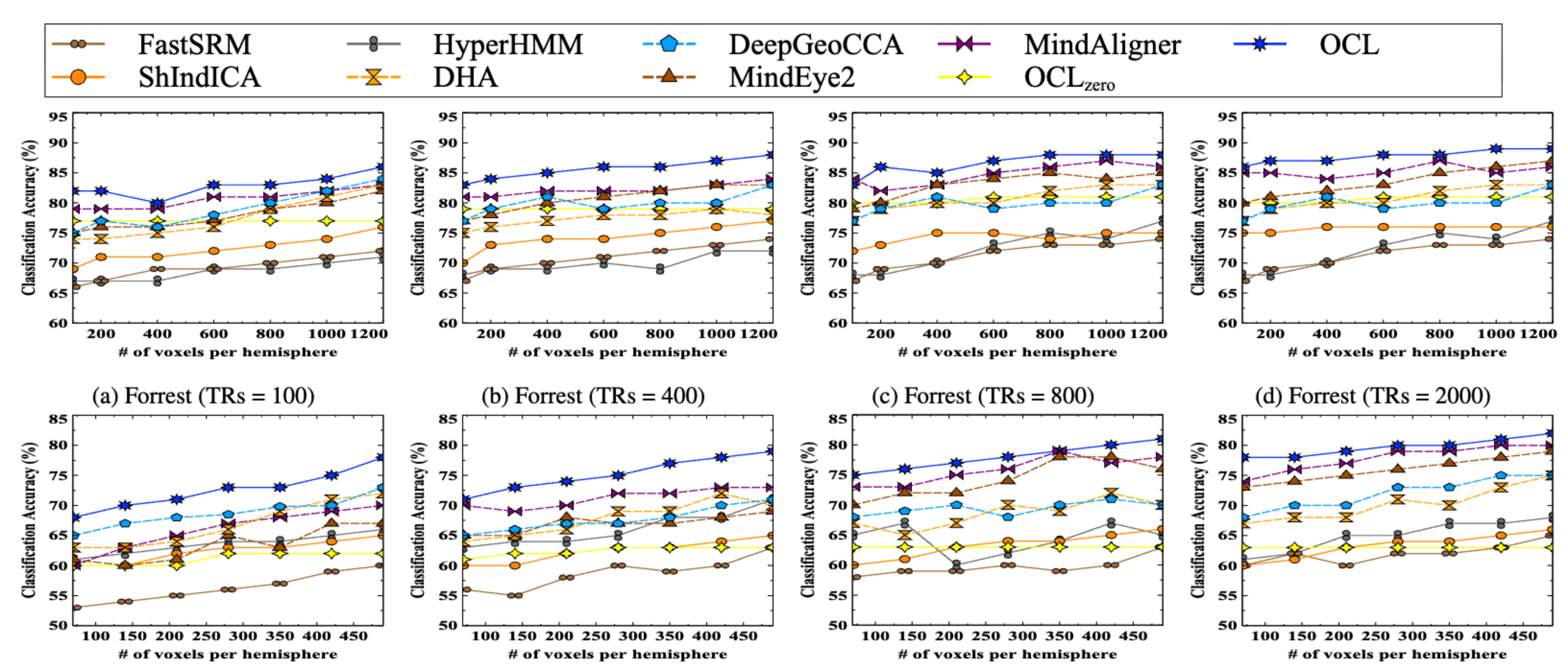
- **7 single-site** fMRI approaches: **FastSRM** and **HyperHMM** as baselines; **ShIndICA** as a non-CCA method; **DHA** and **DeepGeoCCA** as deep multi-view learning approaches; **MindEye2** and **MindAligner** as constructive learning approaches.
- **5 multi-site** techniques: **SSTL** as a baseline; **DeepSSTL** and **XG-GNN** as deep multi-site learning approaches; and **MindEye2** and **MindAligner** as self-supervised constructive methods.

Simple Cognitive Task Classification: *Temporally Aligned vs. Unaligned Data*



(a) CMU (Temporally Unaligned) (b) CMU (Temporally Aligned) (c) DS232 (Temporally Unaligned) (d) DS232 (Temporally Aligned)

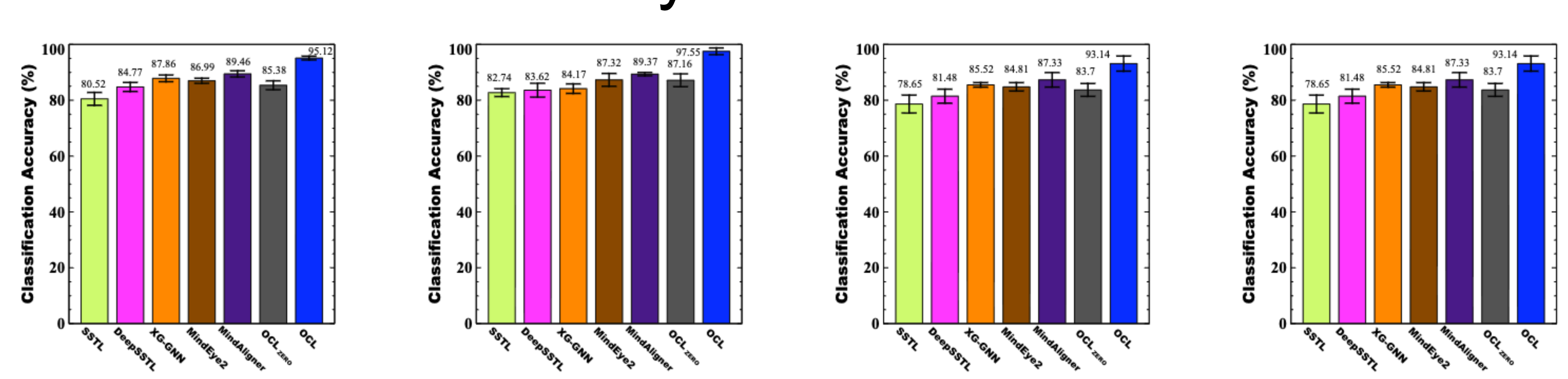
Classification Analysis on Movie Stimuli



(a) Forrest (TRs = 100) (b) Forrest (TRs = 400) (c) Forrest (TRs = 800) (d) Forrest (TRs = 2000)

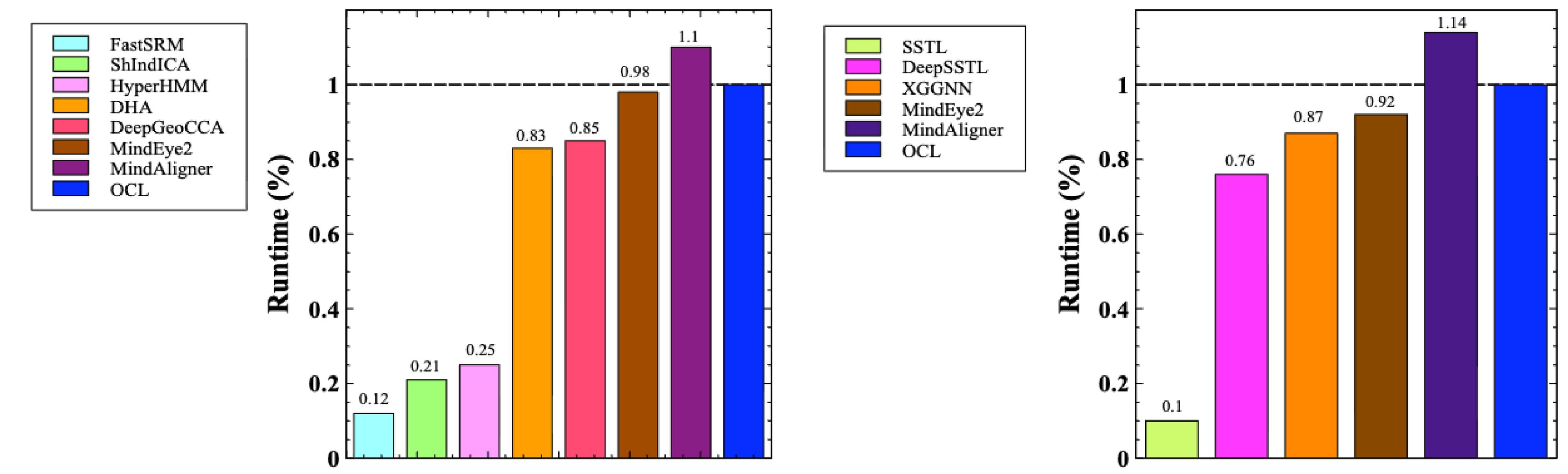
(e) Raiders (TRs = 100) (f) Raiders (TRs = 400) (g) Raiders (TRs = 800) (h) Raiders (TRs = 2000)

Classification Analysis on Movie Stimuli



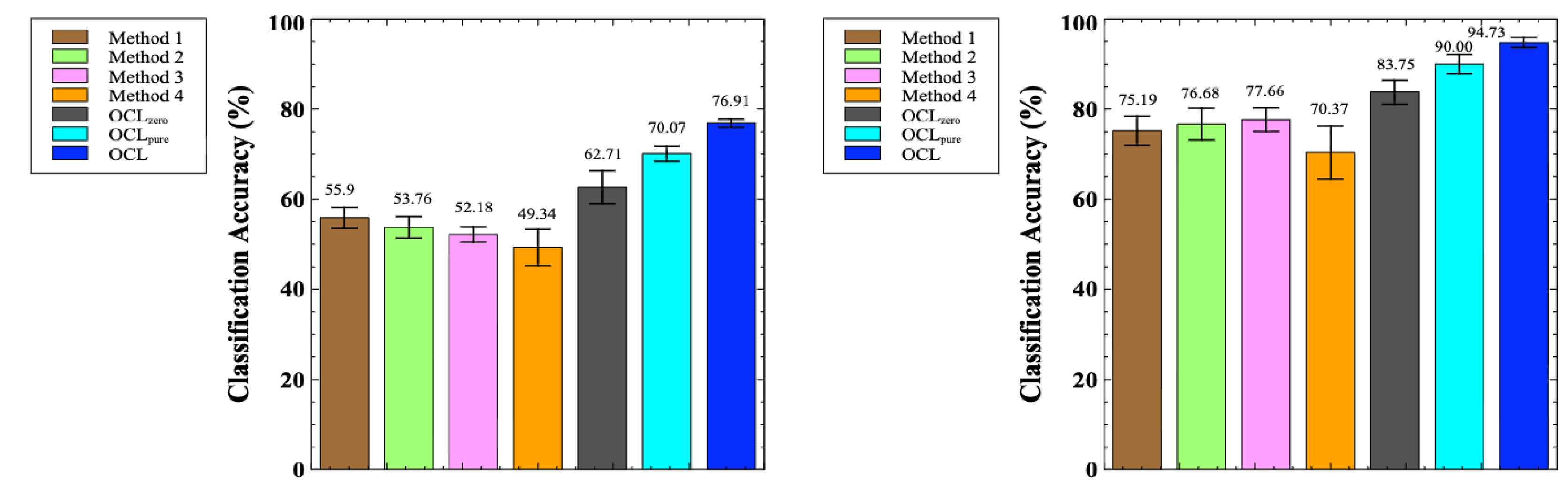
(a) A ⇌ B (b) C ⇌ D (c) E ⇌ F (d) Pairs of {A1, A2, A3, B}

Runtime



(a) CMU (b) E ⇌ F

OCL Component Ablation Study



(a) CMU (b) DS232

- **Method 1**: Skip both QR and LSH. The raw input Xs is passed directly through positional encoding and then the Transformer encoder.
- **Method 2**: Skip LSH. Apply thin QR decomposition to obtain (Qs, Rs), then concatenate Qs and Rs, followed by positional encoding and the Transformer.
- **Method 3**: Skip Positional Encoding. Compute QR and then LSH to get signature φs; concatenate Qs with the MLP-embedded s_s, and feed directly into the Transformer.
- **Method 4**: Skip Encoder Transformer. Use the full QR + LSH + Positional Encoding pipeline, but replace the Transformer encoder with a bidirectional LSTM of comparable capacity.
- **OCL$_{pure}$** : trained on fMRI datasets without the pretraining procedure

Conclusion

Orthogonal Contrastive Learning (OCL) offers a robust, scalable solution for multi-subject and multi-site fMRI analysis. By combining orthogonal feature extraction, contrastive learning, and transformer-based representations, OCL eliminates the need for temporal alignment while improving cross-subject consistency. Experiments show that OCL outperforms existing alignment methods and enables effective transfer learning across diverse datasets—advancing the integration of machine learning and neuroscience.

[1]This research study is conducted independently and is not connected to the author's role at the *National Bank of Canada.*

My website OCL GitHub easy fMRI