Learning By Machine

Orthogonal Contrastive Learning
for Multi-Representation fMRI Analysis

Tony M. Yousefnezhad

tony@learningbymachine.com

Learning By Machine

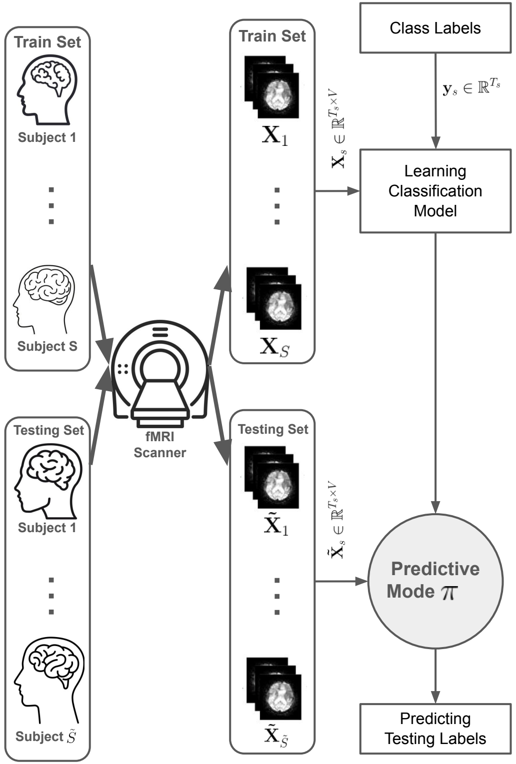
My website

OCL GitHub

NeurIPS 2025

easy fMRI

Motivation



- Developing a neural classification model from multi-subject task-based fMRI brain activation patterns.

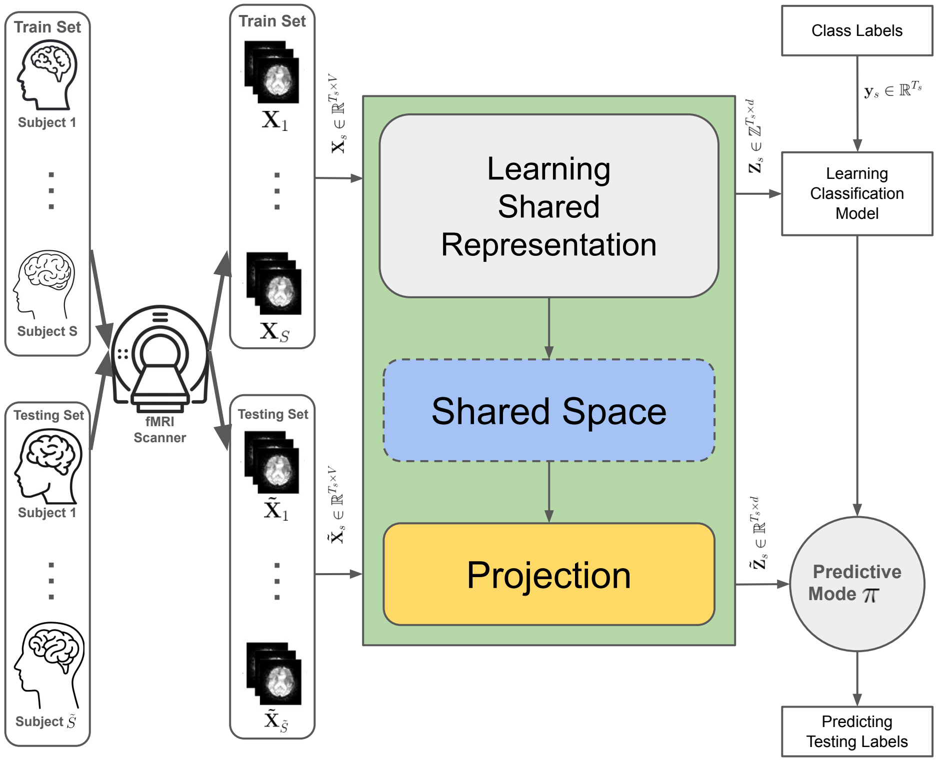
- Given labeled spatiotemporal fMRI data, we train a classification model capable of predicting task conditions for unseen test samples.

- These tasks may involve *visual or auditory stimuli, movie viewing, or other sensory experiments* performed by the subject during the fMRI scan.

- However, multi-subject neural responses must be functionally aligned due to variations in individual brain connectomes.

- From a machine learning perspective, each subject's data represents a distinct transformation of a shared underlying pattern, which must be accounted for before training a classifier.

Functional Alignment



- Functional alignment first learns a new representation from the training fMRI data, capturing a shared underlying pattern across subjects such that neural responses to the same stimuli are aligned among all individuals.

- We then train the classifier on the newly learned representation rather than on the raw neural responses.

- A commonly used classical approach is Generalized Canonical Correlation Analysis (GCCA), which has been developed in various deterministic and probabilistic forms:

Neural responses Shared Space

$$\arg\min_{\mathbf{R_s},\mathbf{G},\theta} \sum_{s=1}^{S} \| f_\theta(\mathbf{X}_s)\mathbf{R}_s - \mathbf{G} \|_F^2$$

Rotation Matrix

Subject to $\mathbf{G}^\top \mathbf{G} = \mathbf{I}$

- Function f could be a linear, classic kernel function, or even a deep neural network

Key Challenges

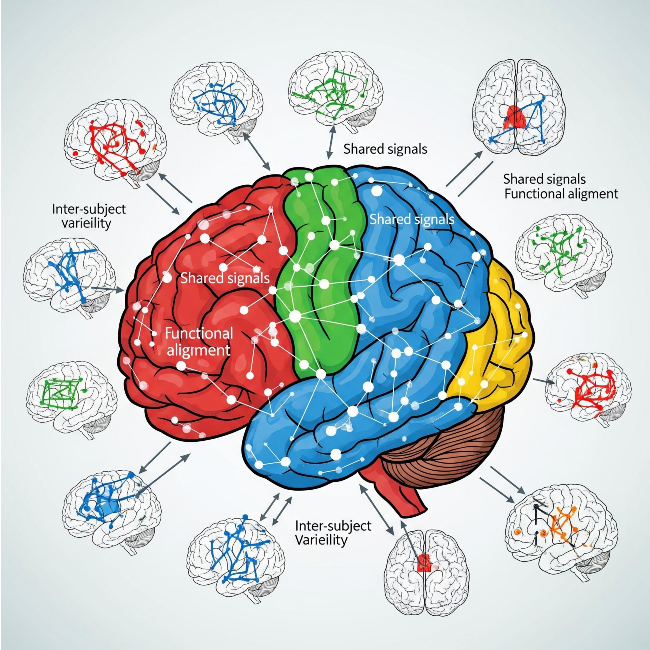
- Inter-subject variability leads to misaligned neural representations across individuals.

- High-dimensional, noisy, and limited-sample fMRI data hinder effective model training and generalization.

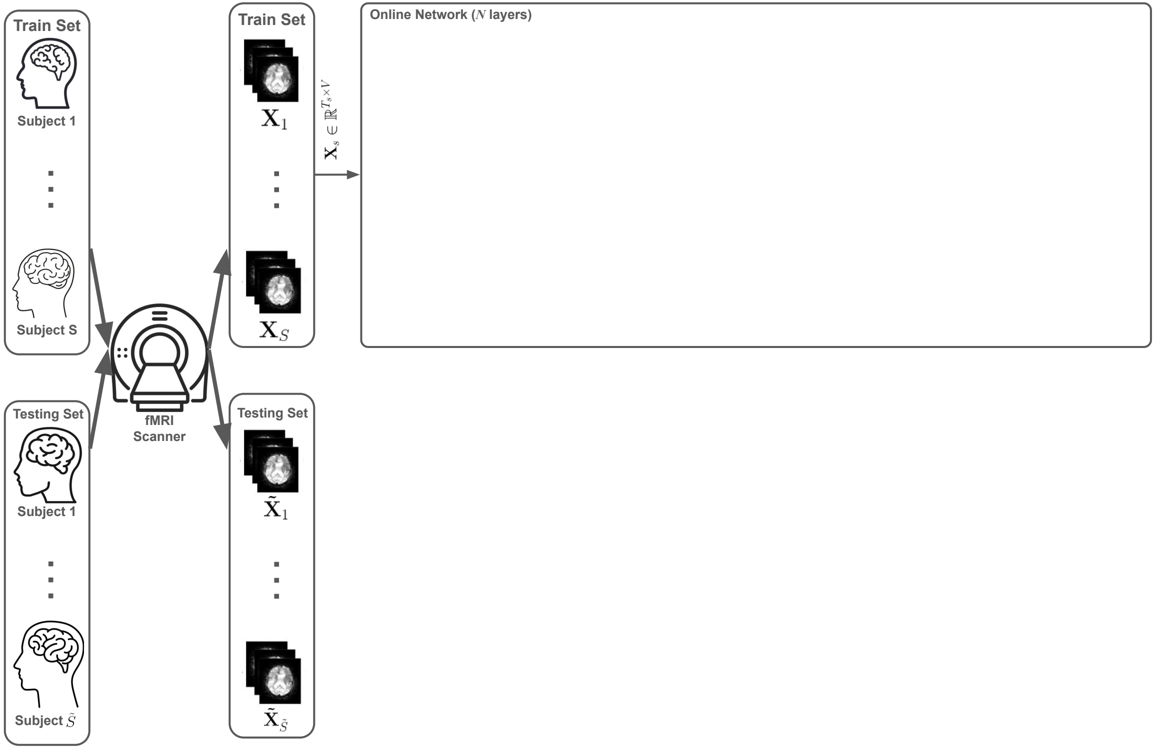
- Temporal misalignment of BOLD responses remains a major issue—most CCA-based alignment methods struggle with this due to their static shared-space structure, which cannot fully capture temporal dynamics across subjects.

- Cross-site heterogeneity and scanner differences introduce domain shifts and batch effects. Lack of domain-adaptive, multi-representation learning frameworks limits robustness in large-scale, multi-subject fMRI analysis.

- Recent advances in Transformer architectures offer a promising solution for modeling long-term temporal dependencies and complex spatiotemporal patterns in fMRI analysis.

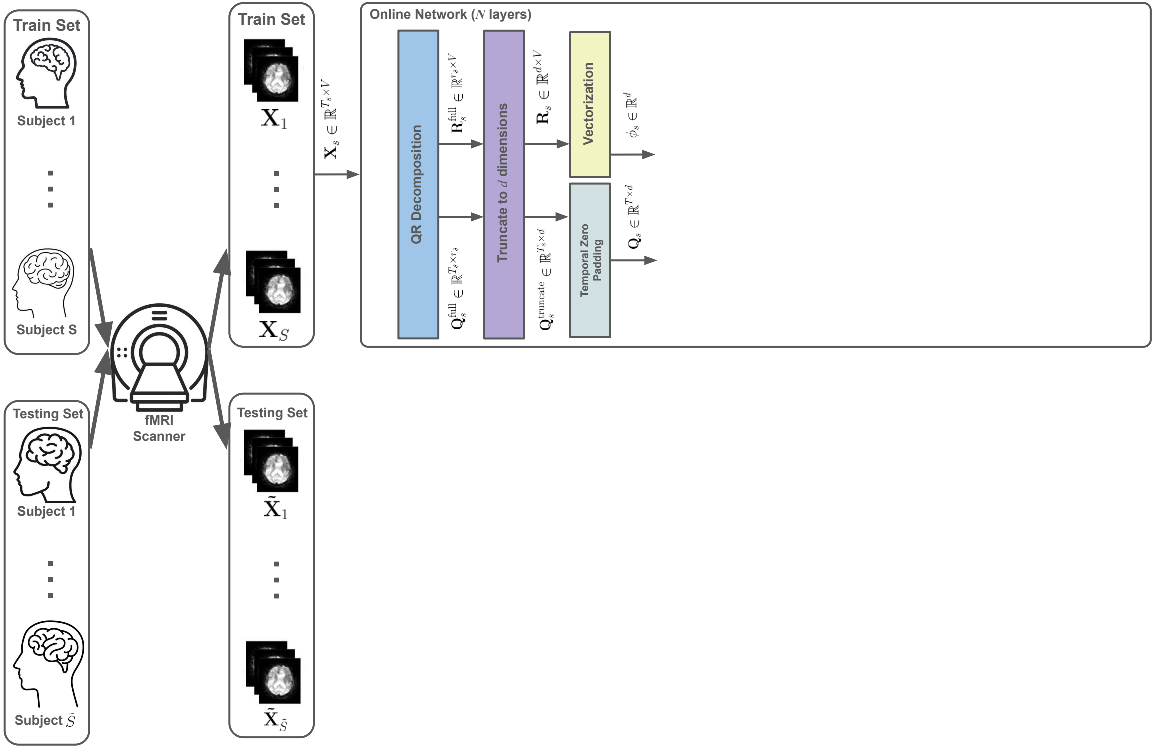


Orthogonal Contrastive Learning



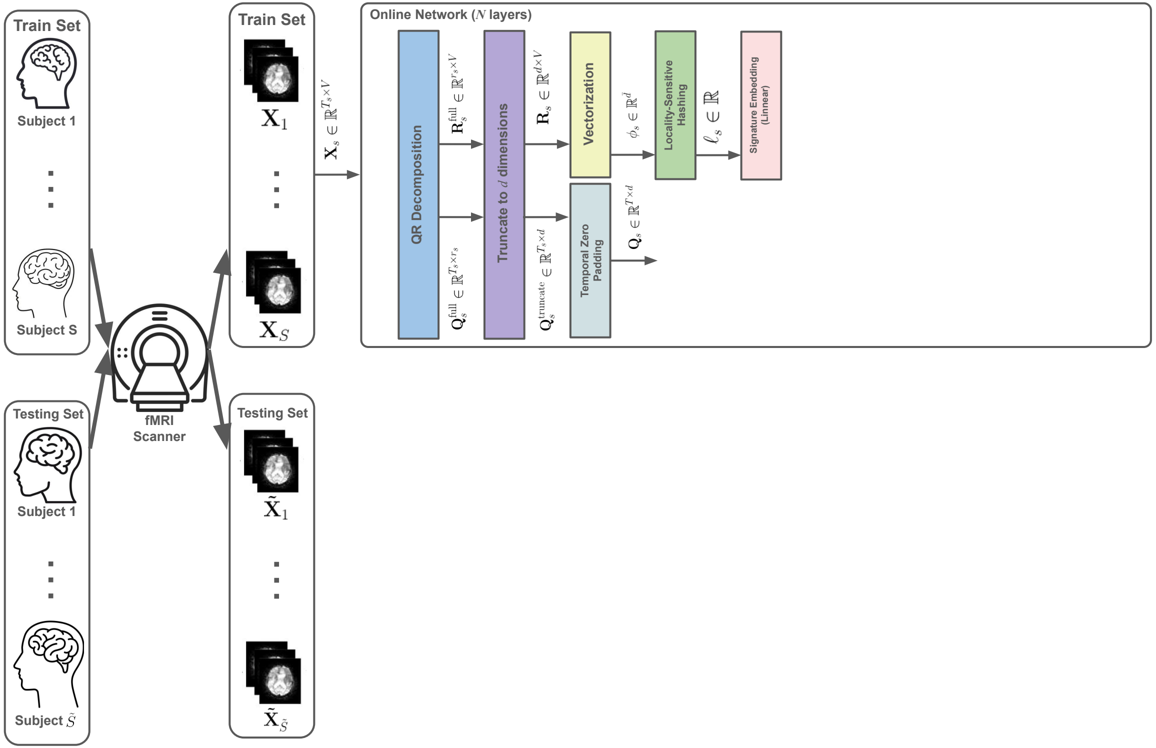
- Each OCL layer has four primary components:

Orthogonal Contrastive Learning - QR decomposition



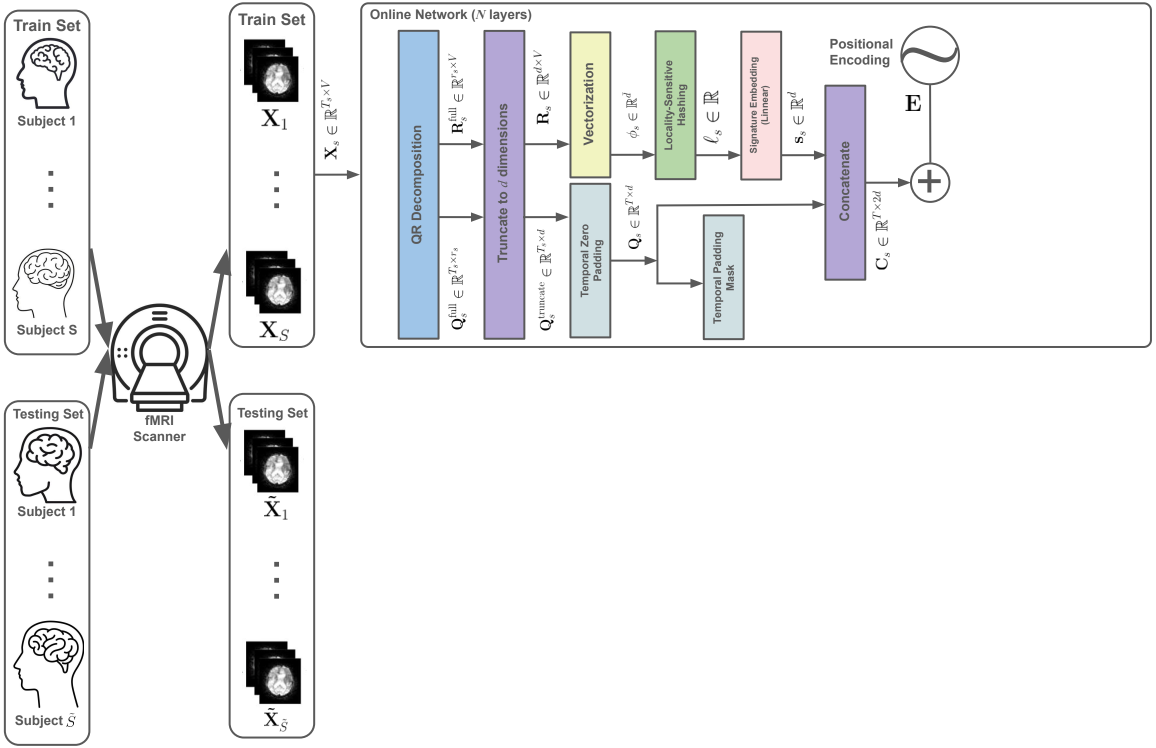
- Each OCL layer has four primary components:
- QR decomposition, which yields orthonormal feature bases to decorrelate signals and enhance the signal-to-noise ratio

Orthogonal Contrastive Learning - Locality-sensitive hashing



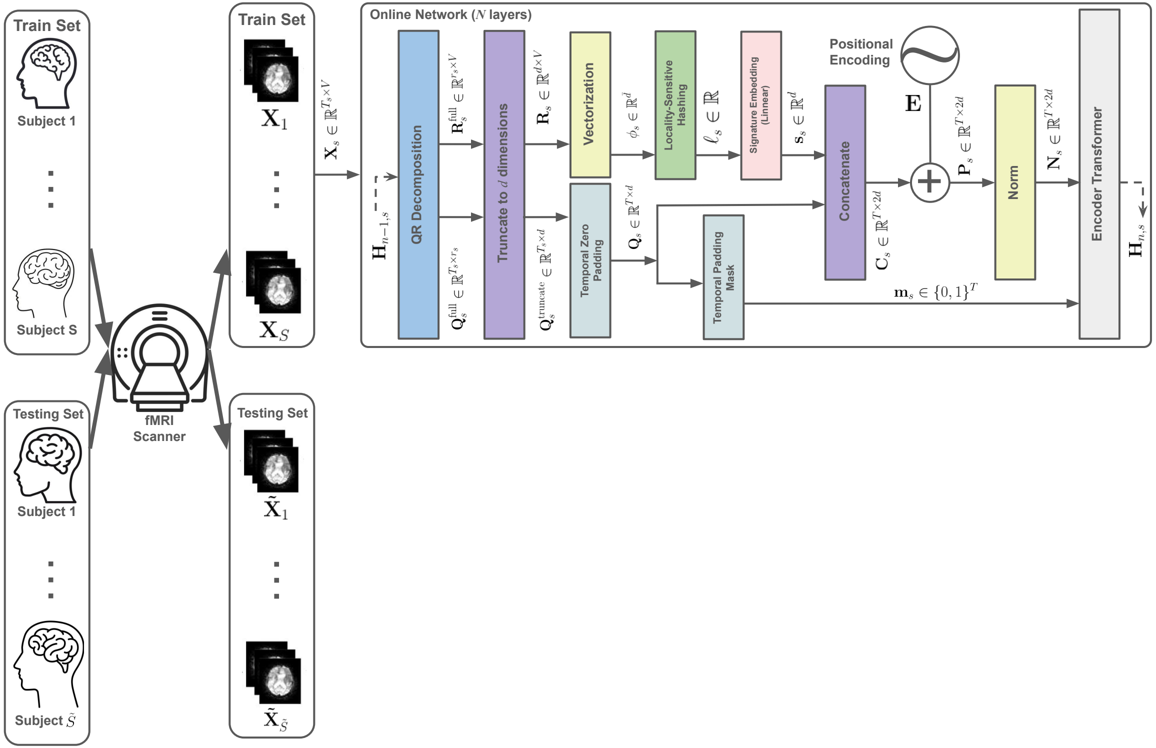
- Each OCL layer has four primary components:
- QR decomposition, which yields orthonormal feature bases to decorrelate signals and enhance the signal-to-noise ratio
- Locality-sensitive hashing (LSH), which produces compact subject-specific signatures

Orthogonal Contrastive Learning - Positional encoding



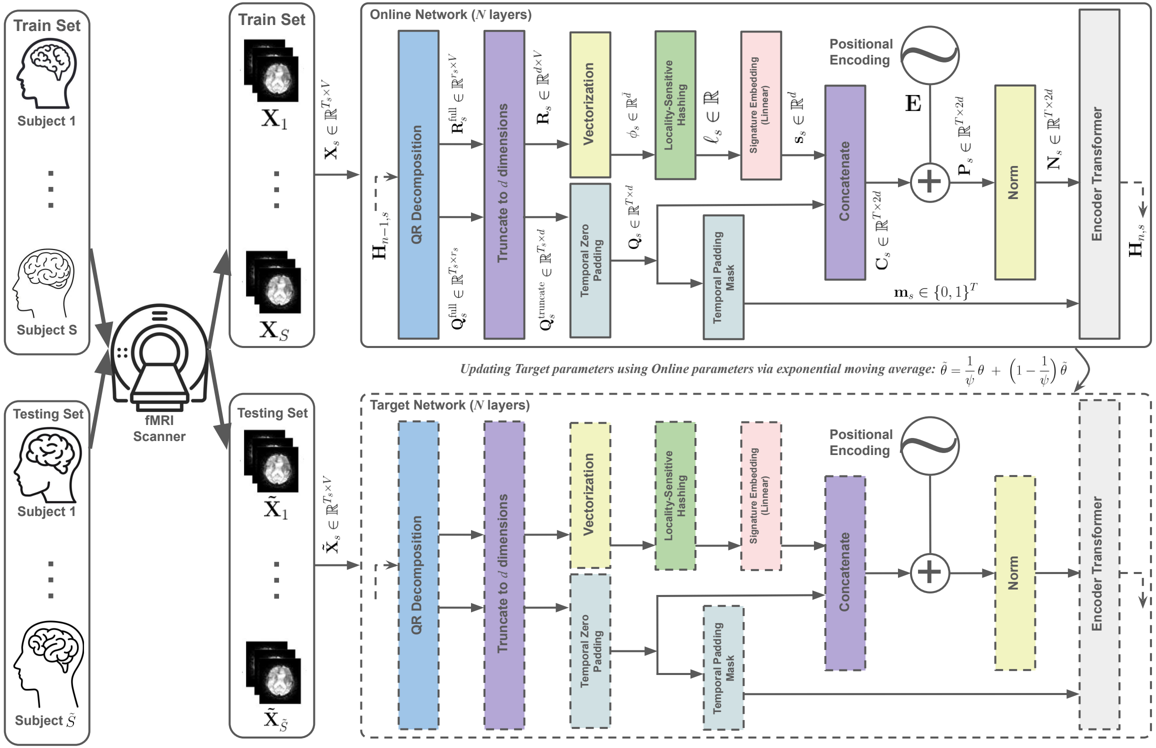
- Each OCL layer has four primary components:
- QR decomposition, which yields orthonormal feature bases to decorrelate signals and enhance the signal-to-noise ratio
- Locality-sensitive hashing (LSH), which produces compact subject-specific signatures
- Positional encoding, which integrates fMRI temporal information into spatial feature representations

Orthogonal Contrastive Learning - Transformer encoder



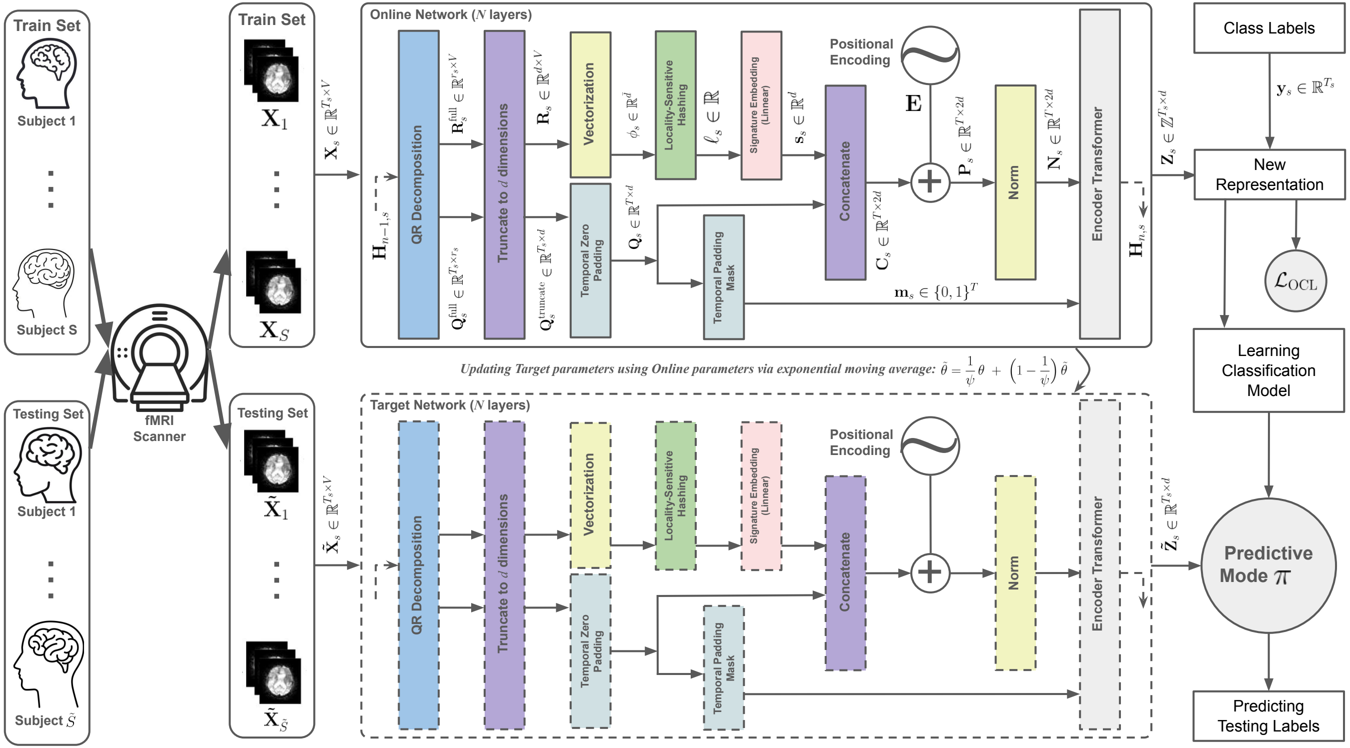
- Each OCL layer has four primary components:
- QR decomposition, which yields orthonormal feature bases to decorrelate signals and enhance the signal-to-noise ratio
- Locality-sensitive hashing (LSH), which produces compact subject-specific signatures
- Positional encoding, which integrates fMRI temporal information into spatial feature representations
- A transformer encoder integrates these inputs, ensuring that neural representations from the same stimulus become closely aligned, while representations of different stimuli remain distinct

Orthogonal Contrastive Learning - Exponential moving average



- We design a dual-encoder architecture:
    - An online network, which actively learns representations through a contrastive objective
    - A target network, whose parameters are gradually updated as a moving average of the online network's parameters.
- This dual-network setup stabilizes training and ensures consistent representations across subjects

Orthogonal Contrastive Learning



- OCL objective function:

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

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

- We learn a contrastive loss that pulls same-stimulus responses together and pushes different-stimulus responses apart

- Trained OCL models can facilitate multi-site fMRI analysis through transfer learning based on orthogonal contrastive embeddings:

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

Single-site parameters

Experimental Setup and Methodology

- We first pretrained OCL using 2 million synthetic fMRI-like data

- We compare OCL with 7 single-site methods:
  - FastSRM and HyperHMM as baselines
  - ShIndICA as a non-CCA method;
  - DHA and DeepGeoCCA as deep multi-view learning approaches
  - MindEye2 and MindAligner as self-supervised constructive learning approaches.

- We compare OCL with 5 multi-site techniques
  - SSTL as a baseline
  - DeepSSTL and XG-GNN as deep multi-site learning approaches
  - MindEye2 and MindAligner as self-supervised constructive methods.

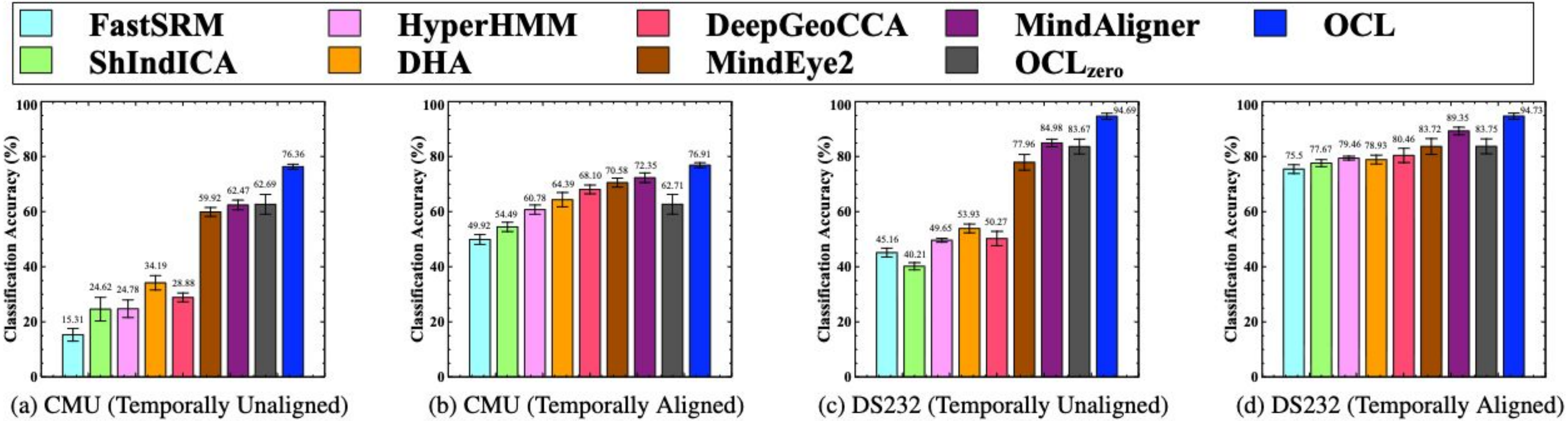
The fMRI datasets

- We used whole-brain ROI
- Each scan was registered to the MNI152 T1-weighted template
- We used 10 datasets to benchmark OCL performance

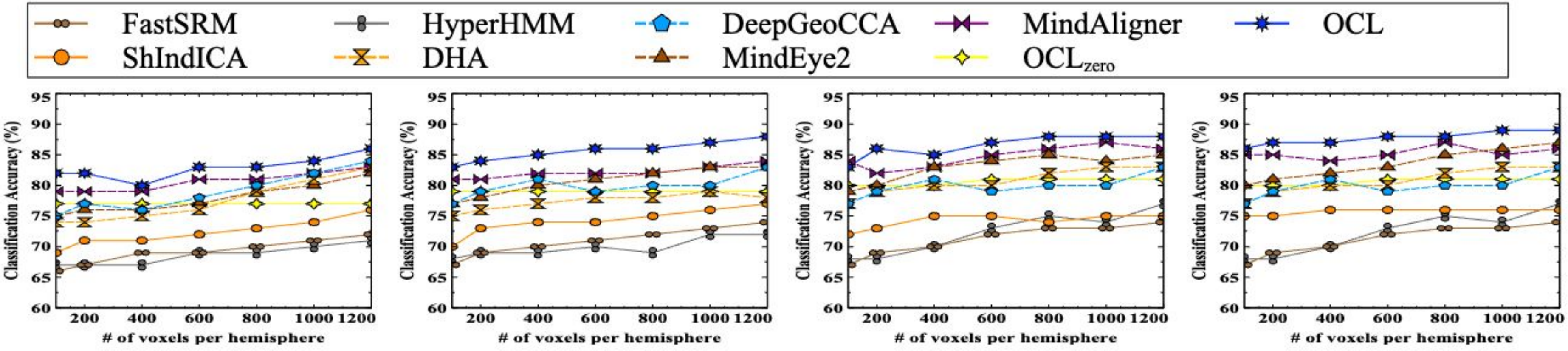
ID	Title	Type	S	$\|\mathbf{y}\|$	T_s	*Site(#)*
A*	Stop signal (DS007)	Decision	20	4	472	B (3)
B	Conditional stop signal (DS008)	Decision	13	4	317	A (1)
CMU	Meanings of Nouns	Semantic	9	12	402	
C	Simon task (DS101)	Simon	21	2	302	D (1)
D	Flanker task (DS102)	Flanker	26	2	292	C (1)
DS232	Face-coding models 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)

Classification Analysis: Temporally Aligned versus Temporally Unaligned Data



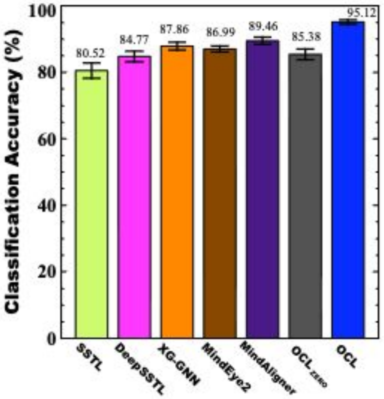
Classification Analysis: Movie stimuli



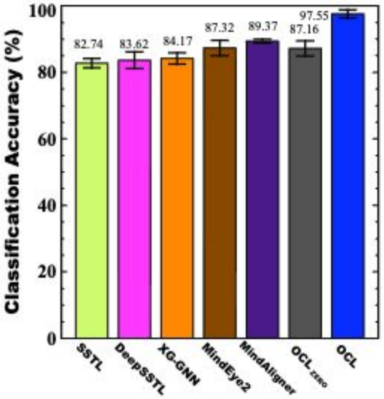
(a) Forrest (TRs = 100) (b) Forrest (TRs = 400) (c) Forrest (TRs = 800) (d) Forrest (TRs = 2000)

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

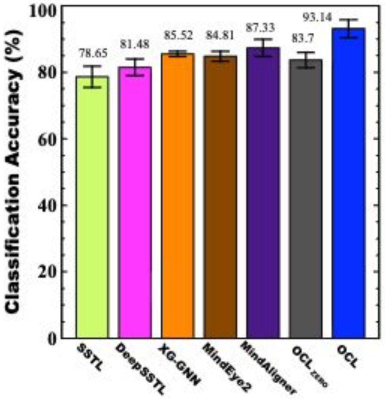
Multi-site Classification Analysis



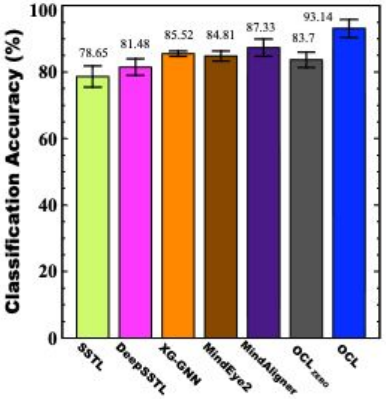
(a) A ⇌ B



(b) C ⇌ D



(c) E ⇌ F



(d) Pairs of {A1, A2, A3, B}

Conclusion

- We introduced orthogonal contrastive learning (OCL), a unified framework that addresses task-based fMRI's key challenges:
  - Low signal-to-noise ratio
  - High dimensionality
  - Variable time-series lengths

- OCL employs a dual-encoder design: an online network and a target network whose weights track the online network via exponential moving average to stabilize learning.

- Each OCL network layer combines
  - QR decomposition for orthogonal feature extraction
  - Locality-sensitive hashing (LSH) to produce compact subject-specific signatures
  - Positional encoding to embed temporal structure alongside spatial features
  - A transformer encoder to generate discriminative, stimulus-aligned embeddings, trained with a contrastive loss that pulls together same-stimulus responses and pushes apart different-stimulus responses.

Learning By Machine

Thank You

Tony M. Yousefnezhad

tony@learningbymachine.com

Learning By Machine

My website

OCL GitHub

NeurIPS 2025

easy fMRI