

Yiming Shen, Ph.D.

Boston, MA | yiming.shen001@umb.edu
yimings.com.cn | [GitHub: Yiming-S](https://github.com/Yiming-S) | [LinkedIn](#)

SUMMARY

Computational Sciences Ph.D. with research experience in EEG-based brain-computer interfaces, statistical machine learning, and scientific software. Research focuses on evaluating EEG decoding systems, transfer across sessions and participants, and trade-offs between accuracy and computational cost. Develops Python and R research tools, with prior industry experience managing sensor time-series data in MySQL.

TECHNICAL SKILLS

EEG & signal processing: CSP/FBCSP, bandpower, covariance and tangent-space features, whitening, Euclidean Alignment.

Machine learning & decoding: LDA, SVM, EEGNet, EfficientNet, weighted ensembles; scikit-learn, PyTorch.

Evaluation & transfer: Nested cross-validation; participant-disjoint evaluation; bootstrap, mixed-effects models, domain adaptation.

Programming & data: Python, R, SQL, Bash; NumPy, SciPy, pandas; MySQL, Git, Linux.

Research tools: MNE-Python, MOABB, pyRiemann, statsmodels; Streamlit, Plotly.

EDUCATION

University of Massachusetts Boston

Ph.D. in Computational Sciences

Boston, MA

August 2026

Valparaiso University

M.S. in Analytics and Modeling

Valparaiso, IN

May 2018

Chongqing University

B.S. in Electronic Information Engineering

Chongqing, China

June 2014

SELECTED RESEARCH

Cross-session EEG transfer and source selection

[Published article](#) | [Research code](#)

- Benchmarked full pooling, distance-weighted pooling, and confidence-gated source selection for cross-session EEG decoding under matched feature, classifier, and adaptation configurations.
- On Stieger2021 (62 participants), the confidence-gated BDP strategy used approximately 37% of full pooling's complete benchmark-search runtime, with selected-configuration accuracy of 64.62%, compared with 64.56% for full pooling and 64.60% for distance-weighted pooling.
- Recorded 76,920 configuration-fold evaluations across Stieger2021 and BNCI2014-004 (9 participants), with explicit session-role records and configuration selection that excluded final target-test labels.

EEG system selection for new participants

- Tested whether small calibration cohorts can select EEG systems for unseen participants. Evaluated 12 configurations built from CSP+LDA, a Riemannian tangent-space classifier, and EEGNet across three datasets, with participant-disjoint training, calibration, and testing.
- Measured complete execution paths at batch size 1 on a single CPU thread. Within the 12 directly timed configurations, calibration-based selection did not outperform the fixed reference on held-out participants. Compared with solo EEGNet, routed systems occupied a lower-cost, lower-accuracy region.

Statistical screening of historical EEG sessions

Research manuscript

- Co-developed confidence-set methods for historical EEG source screening and implemented shared-target bootstrap calibration and simulations. Evaluated the methods on four datasets, including 625 rolling-origin targets across Stieger2021 and Ma2020.
- The refinement gate retained 40%–57% of candidate sessions. Three datasets met a two-percentage-point non-inferiority criterion for balanced accuracy versus all-source pooling; the fourth was inconclusive.

PUBLICATIONS & MANUSCRIPTS

Shen, Y., & Degras, D. (2026). A confidence-gated source selection strategy for cross-session transfer in brain-computer interfaces. *Frontiers in Human Neuroscience*, 20, 1895016. [Published article](#).

Shen, Y. *Small calibration cohorts do not reliably select motor-imagery EEG systems for new participants*. Under review, *Scientific Reports*.

RESEARCH SOFTWARE & EEG METHOD LIBRARIES

CrossDA | *Configurable transfer experiments*

[Project page](#) | [GitHub](#)

- Built command-line and Python interfaces for matched pooling, weighting, and selective-transfer experiments, integrating MOABB, MNE, pyRiemann, scikit-learn, and DA4BCI-Python.
- Exposed configurable feature, classifier, adaptation, and distance methods, with configuration summaries, detailed results, and explicit train, bridge, far, and target session records.

MSDA-Bench | *Benchmark inspection*

[Project page](#) | [GitHub](#)

- Built a Streamlit/Plotly dashboard for inspecting saved CrossDA benchmark results.
- Added participant- and target-session views of configuration effects, ranking stability, source roles, adaptation gains and losses, and accuracy-runtime trade-offs.

ShiftDx | *Offline drift diagnostics*

[Project page](#) | [GitHub](#)

- Developed offline dashboards and statistical reports relating EEG distribution-shift measures to performance under no-adaptation, adaptation, and retraining conditions.
- Integrated MMD, energy, Wasserstein, Mahalanobis, and Euclidean distance diagnostics, with separate workflows for recorded EEG analyses and synthetic-shift demonstrations.

BCIFeatR | *Training-fitted EEG features*

[Project page](#) | [GitHub](#)

- Developed separate training and test interfaces in R, fitting feature transformations on training EEG and reusing the fitted state on held-out trials.
- Implemented reusable feature-extraction routines for CSP/FBCSP, covariance representations, and tangent-space features to support comparative EEG decoding experiments.

eegwhiten | *Whitening and covariance diagnostics*

[Project page](#) | [GitHub](#)

- Implemented reusable whitening models with fit/predict interfaces, inverse transformations, and covariance diagnostics for PCA, ZCA, SVD, and Cholesky methods.
- Added shrinkage and robust covariance estimation, Euclidean alignment, and Riemannian tangent-space utilities for covariance-based analysis.

DA4BCI | *R/Python adaptation interfaces*

[R: GitHub](#) | [Python: GitHub](#)

- Contributed reusable R and Python interfaces for established domain-adaptation methods and distribution-distance metrics.
- Integrated subspace-alignment, covariance-alignment, and distribution-distance methods into research workflows used by CrossDA and ShiftDx.

RESEARCH APPLICATIONS & SIMULATION

EEGLayoutLab | *Interactive numerical simulation*

[Project page](#)

- Connected a React/Three.js electrode-layout interface to Python candidate-exchange search through a Node.js host, with region and spacing constraints.
- Integrated SimNIBS forward modeling and LMMSE reference-field reconstruction, with localization evaluated after layout selection.
- Implemented computation records and cache validation to reuse matching simulation requests. Evaluation used simulated sources on a single specified head template.

NeuroStream | *Recorded EEG replay*

[Project page](#) | [GitHub](#)

- Built a Python desktop application for sample-paced replay of recorded EEG, integrating existing decoding pipelines and Euclidean alignment with background training and inference.
- Added progressive prediction, confidence, bandpower, trial-history, and confusion-matrix displays, with 0.5-second checkpoints for scheduling prediction updates.

TensorEEG | *R/Python simulation and augmentation*

[R: GitHub](#) | [Python: GitHub](#)

- Developed R and Python tools for tensor-form EEG simulation and covariance-based augmentation, with covariance-fidelity audits and manifest-based replay.

ShiftLens | *Interactive distribution-shift exploration*

[Project page](#) | [Demo](#)

- Built a JavaScript/HTML canvas application for comparing domain-adaptation methods and distribution-shift measures on two-dimensional toy data.
- Added CSV input, shareable configuration URLs, and PNG/WebM exports for reusable method demonstrations.

GimbalPaw / OM3 Lab | *Experimental macOS vision/BLE application*

[Project page](#) | [GitHub](#)

- Built an experimental macOS application integrating SwiftUI, AVFoundation camera capture, Apple Vision person tracking, and CoreBluetooth control of a DJI OM3 gimbal.
- Implemented target reacquisition, One Euro smoothing, frame-freshness checks, software motion limits, and command coordination using a community-documented device protocol.
- Wrote XCTest unit tests and mocked integration tests for application logic.

COLLABORATIVE ALGORITHMS & ADDITIONAL RESEARCH

TMCCA / tensorMCCA | *Tensor multiview methods*

GitHub: [project repository](#)

Collaborative contribution; package author and maintainer: David Degras.

- Contributed software development, numerical simulations, and data analyses to tensor multiple canonical correlation analysis, studying recovery and optimization behavior.
- Analyzed a single-participant, multi-session EEG example using four aligned views, including task labels, with five components fixed in advance.

matchFeat | *One-to-one feature matching*

GitHub: [project repository](#)

Collaborative contribution; package author and maintainer: David Degras.

- Contributed to software development, validation, and data curation for one-to-one matching of unlabeled feature vectors across data units.
- The package provides empirical-loss, clustering, and probabilistic approaches to one-to-one feature matching.

tensorICA | *Collaborative algorithm design and validation*

- Designed and validated tensor independent component analysis algorithms in collaborative doctoral research, using R prototypes and source-separation checks.

NLP and clustering of scientific literature

- Contributed to an EEG-literature analysis workflow combining OpenAlex retrieval, abstract reconstruction, MiniLM embeddings, UMAP, HDBSCAN, and BERTopic.
- Documented corpus filtering and sensitivity analyses. The manuscript reports 1,986 usable abstracts and 1,671 cluster-labeled records from a retrieval capped at 3,000.

EXPERIENCE

Doctoral Researcher

Sep 2019 - Aug 2026

University of Massachusetts Boston

- Designed nested cross-validation experiments with five outer and three inner folds to distinguish representation gains from classifier gains on fixed features; measured held-out accuracy, inference latency, and serialized model size.
- Used mixed-effects models and participant-level trajectories to distinguish fixed-model degradation, adaptation benefits, and remaining gains from retraining; documented dataset-specific limits of drift as a proxy for performance loss.

Data Engineer

Jul 2014 - Aug 2016

China Mobile IoT Company Limited, Chongqing, China

- Managed gigabyte-scale sensor time-series data in MySQL using indexing, partitioning, and aggregate queries. Built network-health dashboards and automated reports within established SLAs.

RECOGNITION & PRESENTATIONS

Kaggle HMS – Harmful Brain Activity Classification

2024

Silver Medal; 98th of 2,767 teams (top 4%); solo competitor.

Neuroergonomics, Boston – Poster: EEG accuracy-complexity benchmarking.

July 2026

ENAR Spring Meeting, Indianapolis – Poster: cross-session BCI transfer.

March 2026

Inria MIND Seminar, France – Invited talk: EEG classification pipelines.

June 2025