

Yiming Shen, Ph.D.

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

SUMMARY

Ph.D. in Computational Sciences working across EEG and brain-computer interfaces (BCI), statistical machine learning, and research software. Studies model evaluation, transfer across sessions and participants, and accuracy-cost trade-offs. Builds Python/R research tools, with earlier industry experience managing sensor time-series data in MySQL.

TECHNICAL SKILLS

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

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

Evaluation & transfer: Nested and participant-disjoint validation, 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

Boston, MA

Ph.D. in Computational Sciences

August 2026

Valparaiso University

Valparaiso, IN

M.S. in Analytics and Modeling

May 2018

Chongqing University

Chongqing, China

B.S. in Electronic Information Engineering

June 2014

SELECTED RESEARCH

Cross-session EEG transfer and source selection

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

- Benchmarked strategies for reusing historical EEG sessions in cross-session decoding, comparing full pooling, distance weighting, and confidence-gated source selection under matched configurations.
- On Stieger2021 (62 participants), confidence-gated selection used approximately 37% of full pooling's complete benchmark-search runtime; selected-configuration accuracies across the three strategies were within 64.56%-64.62%.
- Recorded 76,920 configuration-fold evaluations across Stieger2021 and BNCI2014-004 (9 participants), with explicit session roles and no final target-test labels used for configuration selection; search runtime is not online inference latency.

EEG system selection for new participants

- Tested whether small calibration cohorts can select EEG systems for new participants, using 12 configurations built from CSP+LDA, a Riemannian tangent-space classifier, and EEGNet across three datasets with participant-disjoint training, calibration, and testing.
- Directly timed complete batch-one execution on one CPU thread. The calibration cohort's highest-accuracy choice did not outperform the fixed reference within the timed set; routing between decoders had lower execution cost and lower accuracy.

Statistical screening of historical EEG sessions

- Co-developed confidence-set methods to select historical EEG sessions under uncertainty; implemented shared-target bootstrap calibration, simulations, and longitudinal evaluations across four datasets and 625 rolling-origin targets.
- The manuscript reports retaining 40%-57% of candidate sessions. Three datasets met a two-percentage-point non-inferiority criterion 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*.

SELECTED SOFTWARE & APPLIED PROJECTS

CrossDA, MSDA-Bench, and ShiftDx | *Experiment tools and dashboards*

CrossDA: [GitHub](#) | MSDA-Bench: [GitHub](#) | ShiftDx: [GitHub](#)

- Built a configurable Python CLI/API for matched transfer experiments, recording configurations, detailed results, and train/bridge/far/target session roles.
- Developed dashboards for participant-level benchmark inspection and offline drift, adaptation, and retraining comparisons, including ranking stability and accuracy-runtime trade-offs.

EEGLayoutLab | *Electrode-layout simulation*

[Project page](#)

- Connected interactive 3D layouts and spacing constraints to candidate-exchange search, SimNIBS forward modeling, and LMMSE field reconstruction.
- Implemented reusable computation records and cache checks; evaluations used simulated sources and one specified head template.

BCIFeatR, eegwhiten, and DA4BCI | *R/Python research libraries*

BCIFeatR | [eegwhiten](#) | [DA4BCI-Python](#)

- Developed training-fitted feature interfaces, reusable whitening models, and covariance diagnostics; contributed common interfaces for standard domain-adaptation methods.

NLP and clustering of scientific literature

- Contributed to an EEG-literature analysis workflow using 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-labelled records from a retrieval capped at 3,000, with query and corpus limitations.

EXPERIENCE

Doctoral Researcher

Sep 2019 - Aug 2026

University of Massachusetts Boston

- Designed motor-imagery EEG experiments using five-fold outer and three-fold inner validation, separating feature-representation gains from fixed-feature classifier gains and measuring accuracy, latency, and model storage.
- Used mixed-effects analyses and participant trajectories to distinguish fixed-model degradation, adaptation benefit, and remaining retraining gains; retained dataset-specific counterexamples to drift as a loss proxy.

Data Engineer

Jul 2014 - Aug 2016

China Mobile IoT Company Limited, Chongqing, China

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

RECOGNITION & PRESENTATIONS

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

2024

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