Evaluating Latent Space Robustness and Uncertainty of EEG-ML Models under Realistic Distribution Shifts

Neeraj Wagh[1], Jionghao Wei[2], Samarth Rawal[3], Brent Berry[4], Yogatheesan Varatharajah[1,2,4]

[1] *Dept. of Bioengineering, University of Illinois, Urbana, IL, USA*
[2] *Dept. of Electrical & Computer Engineering, University of Illinois, Urbana, IL, USA*
[3] *Carle Illinois College of Medicine, University of Illinois, Urbana, IL, USA*
[4] *Dept. of Neurology, Mayo Clinic, Rochester, MN, USA*

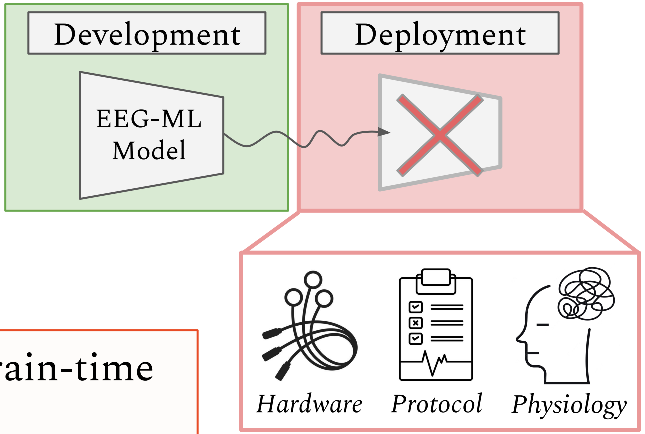
NeurIPS Conference, 2022

Real-world Robustness of EEG-ML Models

- EEG is a versatile tool for recording brain activity
  - Wide range of applications based on EEG and ML
- EEG-ML models fail in deployment[1]
  - Curated datasets, complex real-world shifts
- Can we predict deployment failures at train-time?
  - Existing approaches require data from target settings[2]
- Contribution: Approach to predict deployment failures at train-time
  - Domain knowledge to model realistic EEG shifts
  - Develop robustness measures to assess impact of shifts
  - Train-time analysis predicts in-the-wild performance

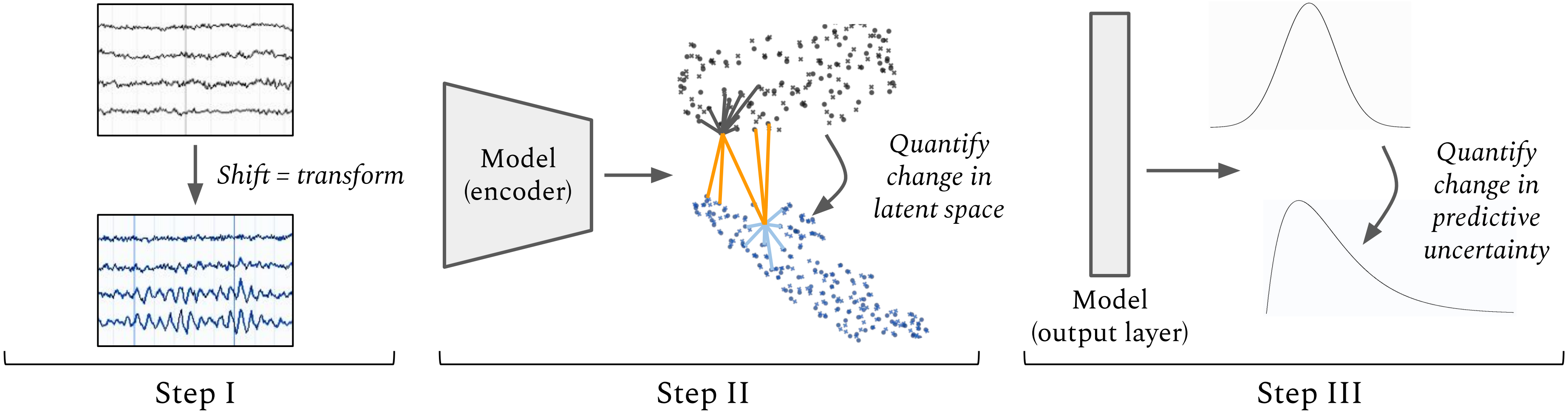
Clinical-grade

Consumer-grade



1. Xu, Lichao, et al. "Cross-dataset variability problem in EEG decoding with deep learning." Frontiers in human neuroscience 14 (2020): 103.
2. Saria, Suchi, and Adarsh Subbaswamy. "Tutorial: safe and reliable machine learning." arXiv preprint arXiv:1904.07204 (2019).

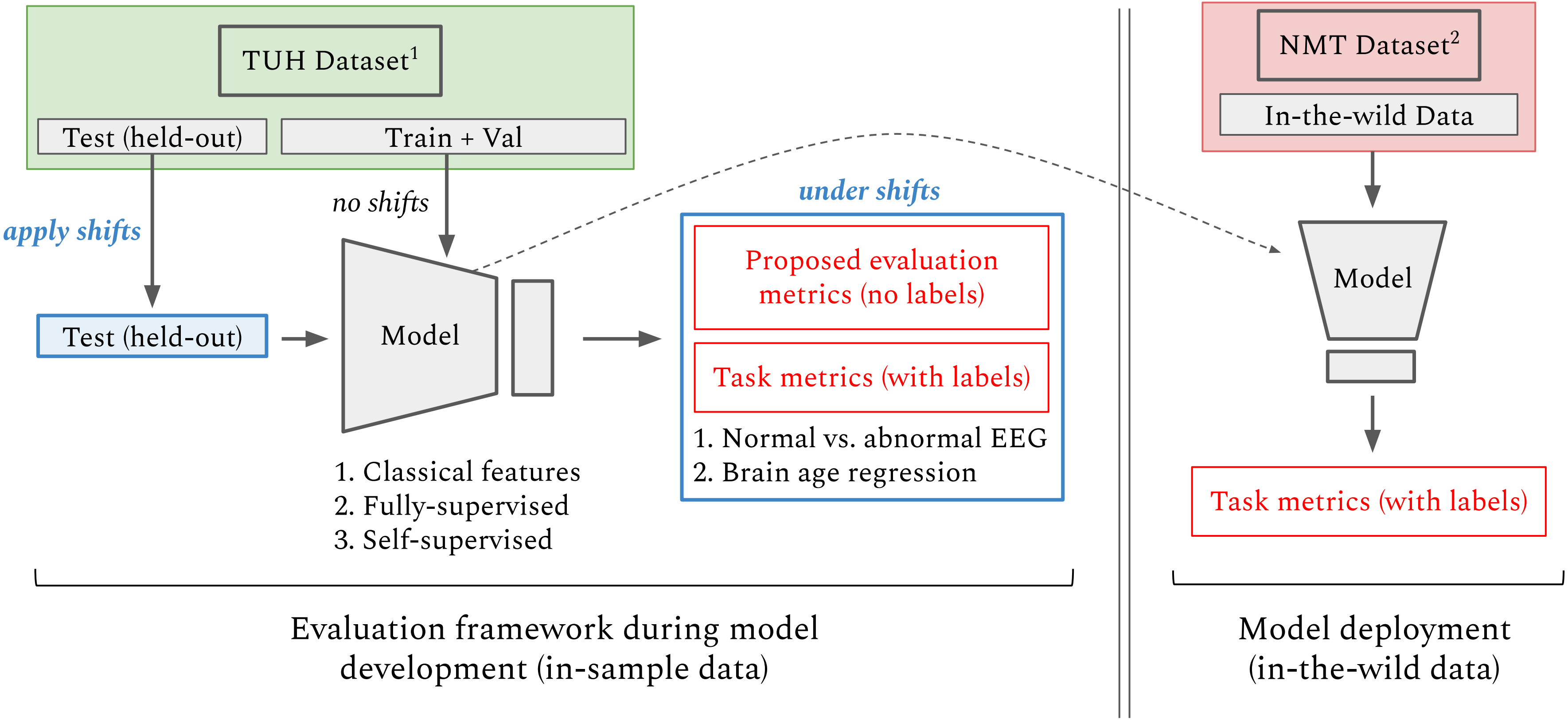
Evaluating Robustness to Realistic Distribution Shifts



- **Step I:** Capture realistic EEG shifts as data transforms
  - Domain knowledge → effect of shift[1] → raw data transform
- **Step II:** Quantify change in the encoder's latent space
  - Neighboring points → graph[2] → graph-based measure
- **Step III:** Quantify change in predictive uncertainty at output
  - Monte Carlo dropout[3]-based measure

1. Kappenman, Emily S., and Steven J. Luck. "The effects of electrode impedance on data quality and statistical significance in ERP recordings." Psychophysiology 47.5 (2010): 888-904.
2. Poklukar, Petra, et al. "Delaunay component analysis for evaluation of data representations." arXiv preprint arXiv:2202.06866 (2022).
3. Gal, Yarin, and Zoubin Ghahramani. "Dropout as a bayesian approximation: Representing model uncertainty in deep learning." International Conference on Machine Learning. PMLR, 2016.

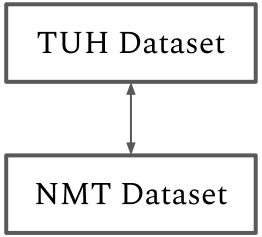
Experimental Setup: In-sample and Out-of-sample Evaluations

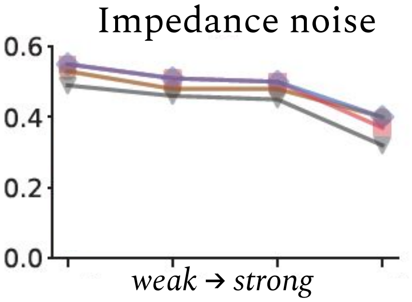


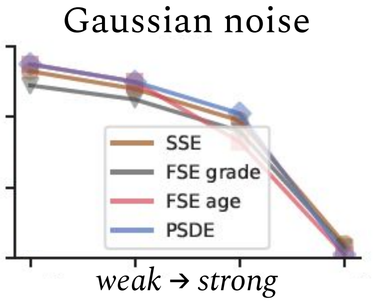
1. Obeid, Iyad, and Joseph Picone. "The temple university hospital EEG data corpus." Frontiers in neuroscience 10 (2016): 196.
2. Khan, Hassan Aqeel, et al. "The NMT Scalp EEG Dataset" Frontiers in Neuroscience 15 (2021).

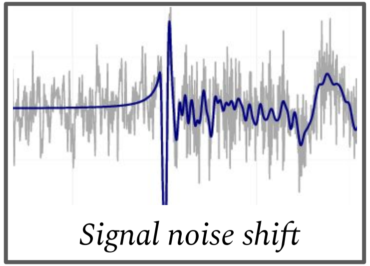
Strong Shifts During Development Predict Out-of-sample Performance

- Latent space robustness measure



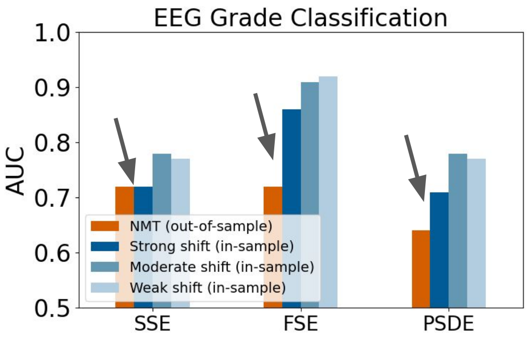


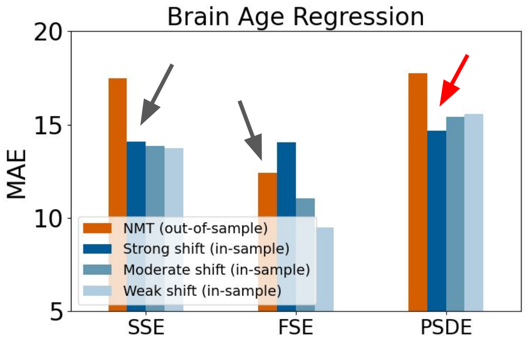




- Task performance

(Higher is better)





(Lower is better)

- Impact of strong shifts in development predicts in-the-wild performance

Future Directions

- Modeling additional EEG shifts
  - Physiological
  - Clinical protocols

- Extend approach to other healthcare data modalities
  - Imaging
  - Text

- How can we mitigate changes due to realistic shifts?
  - Training with shifts
  - Adversarial training

- Community benchmark for robustness
  - Establish robustness profiles of popular models

Code & pre-print