# Knowing When You Don't Know: Predictive Uncertainty Measures for Seizure Detection

**Nandita Bhaskhar[1] (nanbhas@stanford.edu), Daniel Rubin[2,3] & Christopher Lee-Messer[4]**

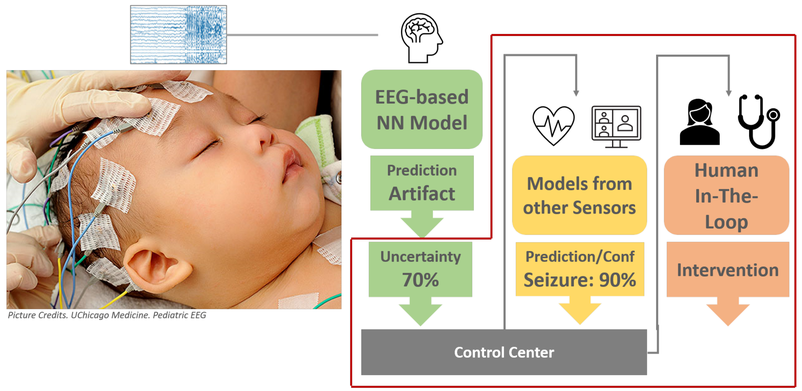
[1]Dept. of Electrical Engineering, [2,3]Depts. of Biomedical Data Science & Radiology, [4] Dept. of Neurology, Stanford University

## Introduction & Motivation

- **Modern neural networks** (NN) show **state-of-the-art performance** in various healthcare domains such as Radiology, Dermatology, Neurology, etc
- Despite their stellar performance on large-scale benchmarks, they are **not infallible** to errors
- Moreover, they are known to produce **highly confident predictions *even when wrong*** [1]
- This makes their adoption into clinical practice almost infeasible

## Uncertainty Prediction: Rationale for Neurology

- What if, instead, each NN's prediction comes with a measure of its **predictive uncertainty**?
- Consider an **EEG-based NN model** whose job is to predict if a child's brain activity constitutes a **seizure**, an **artifact** or **normal activity**
- Based on **robust uncertainty scores** produced by the model, it can alert humans or give control to other models when needed




Picture Credits. UChicago Medicine. Pediatric EEG

## Out-of-distribution (OOD) Detection

- Determining if **unseen input** belongs to the **same** distribution as **training data** (i.e. IN Distribution) or **not** (i.e. Out-of-distribution or OOD)

## Our approach: Key Insight

- NNs provide **lower-dimensional** representations of inputs we can use for OOD detection!!!

## Predictive Uncertainty (MScores): Generation Theory

**Generative Modelling: an LDA-based Softmax Classifier**

- The ***softmax*** classifier can be considered equivalent to the posterior distribution defined by a ***generative classifier under LDA (Linear Discriminant Analysis) with a shared covariance assumption*** [3]
- We thus fit class-conditional Gaussian distributions with a shared covariance matrix to training samples under the maximum likelihood estimator
- We estimate the empirical class-wise means $\mu_c$ and shared covariance matrix $\Sigma$ of the multivariate Gaussian using hidden layer activations $f(\mathbf{x})$ of the trained network as given by ➔

## Predictive Uncertainty (MScores): Generation Theory

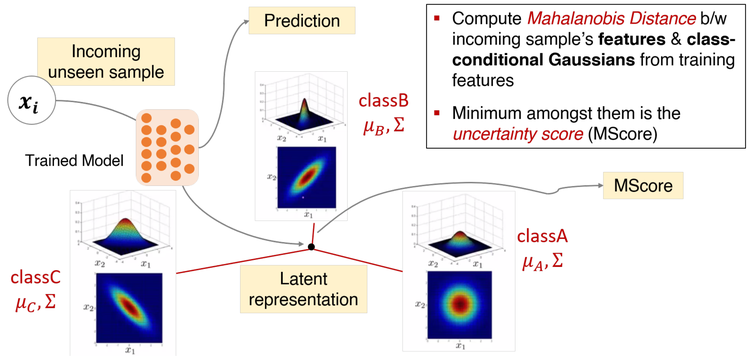
$$\hat{\mu}_c = \frac{1}{N_c}\sum_{i:y_i=c} f(\mathbf{x}_i)\ ; \quad \hat{\Sigma} = \frac{1}{N}\sum_{c}\sum_{i:y_i=c}(f(\mathbf{x}_i)-\hat{\mu}_c)(f(\mathbf{x}_i)-\hat{\mu}_c)^T \tag{1}$$

- We define the confidence metric (MScore or $M(\mathbf{x})$) given to each test sample $\mathbf{x}_i$ under this induced generative classifier to be the Mahalanobis distance between the sample and the closest class-conditional Gaussian distribution

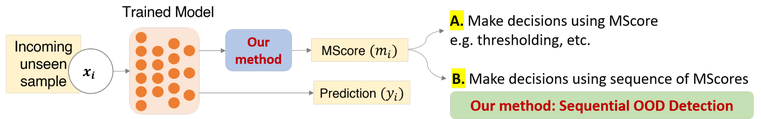
$$M(\mathbf{x}_i) = \min_c \ (f(\mathbf{x}_i)-\hat{\mu}_c)^T\hat{\Sigma}^{-1}(f(\mathbf{x}_i)-\hat{\mu}_c) \tag{2}$$

$$\hat{y}(\mathbf{x}_i) = \operatorname*{argmin}_c \ (f(\mathbf{x}_i)-\hat{\mu}_c)^T\hat{\Sigma}^{-1}(f(\mathbf{x}_i)-\hat{\mu}_c) \tag{3}$$

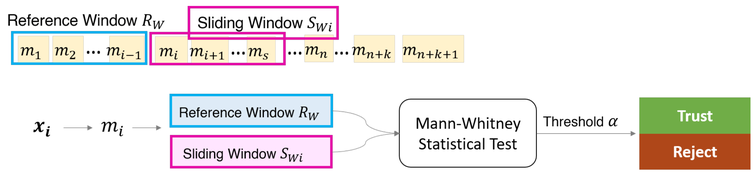
- This generative classifier thus classifies the incoming test sample as per Eq. 3.




## Sequential OOD Detection: Motivation & Theory



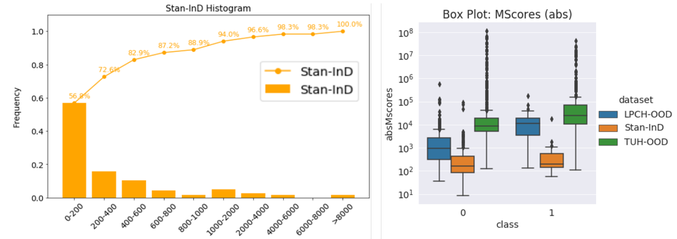

- A trained model **deployed** to a clinic encounters **continuously evolving stream of medical data** (e.g. EEG that changes with age, physical condition, etc)
- We detect such changes over time by assessing similarity between MScore distributions for in-distribution data and MScore distributions for unseen test samples during deployment
- We put forth an ***unsupervised, sliding-window based algorithm*** building on work done by Kifer et al. [2] to identify ***when*** the model should indicate that it is no longer certain of its predictions
- Consider $n$ test samples to the model, $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ with MScores, $\{m_1, m_2, \ldots, m_n\}$



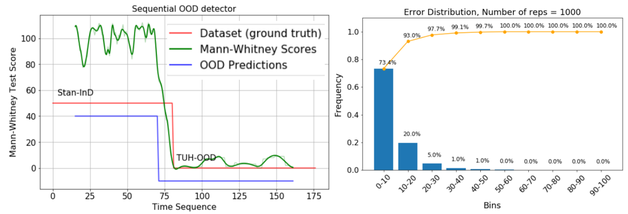

## Methods

- **Datasets used**: EEGs from Stanford Hospital (**Stan-InD**), Lucile Packard Children's Hospital (**LPCH-OOD**) and Temple University public EEG dataset (**TUH-OOD**)
- Stan-InD, LPCH-OOD vary in age distributions. TUH-OOD is from a different institution
- **Task**: Seizure detection, **Model**: **Dense-inception** [4] trained on **12s** clips from **Stan-InD**

## MScores: Results




## Sequential OOD Detection: Results




## Conclusions

- High quality of MScores indicative of distribution shifts generated
- Novel sequential detection framework introduced. Makes NO assumptions on data
- Methodology generalizable to all kinds of data, clinical and non-clinical use cases

**Future Directions**: Synergies & Applications to Online learning, Active learning, Federated learning, Learning with feedback, Clinical Decision-Making, etc

**References**:

1. Guo2017, ICML 2. Kifer2004, VLDB 3. Lee2018, NeurIPS 4. Saab2020, npj Digital Medicine