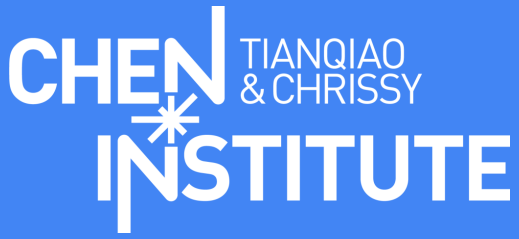
Chen Data Science & AI for Neuroscience Summer School



Caltech

Data Processing Principles

Sabera Talukder

What are your data processing principles?

- Visualize Your Data → Intuition Development

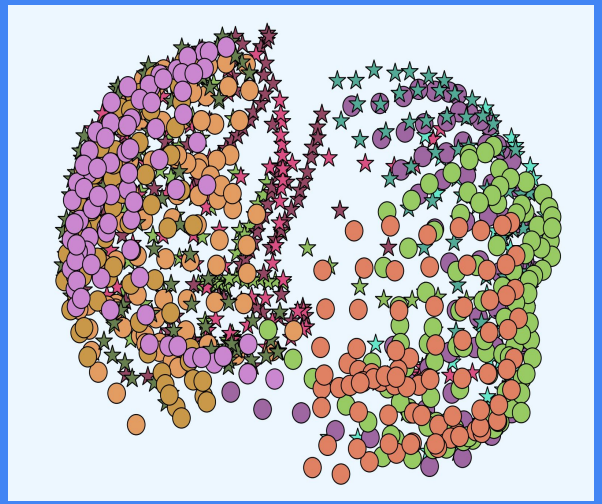
- Visualize Your Data → Intuition Development
- Signal Extraction (aka Denoising)

- Visualize Your Data → Intuition Development
- Signal Extraction (aka Denoising)
- Dataset Augmentation

- Visualize Your Data → Intuition Development
- Signal Extraction (aka Denoising)
- Dataset Augmentation
- Normalization / Standardization

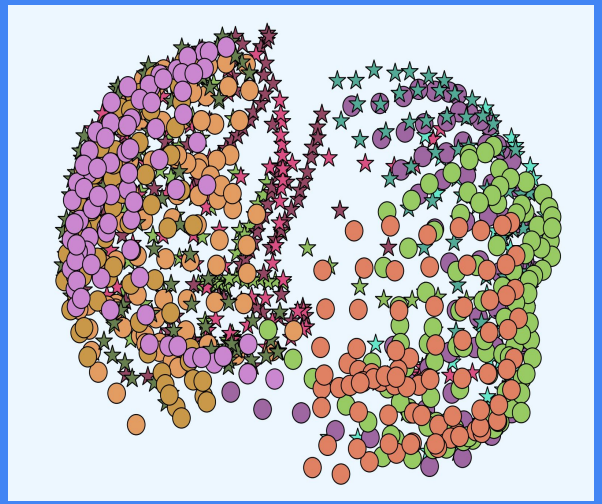
- Visualize Your Data → Intuition Development
- Signal Extraction (aka Denoising)
- Dataset Augmentation
- Normalization / Standardization
- Train / Validation / Test Splits

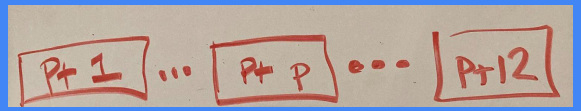
Visualize Your Data



Talukder et al. 2022

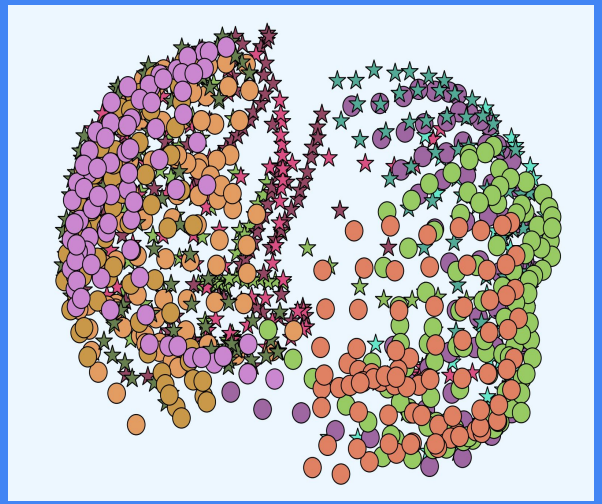
Visualize Your Data

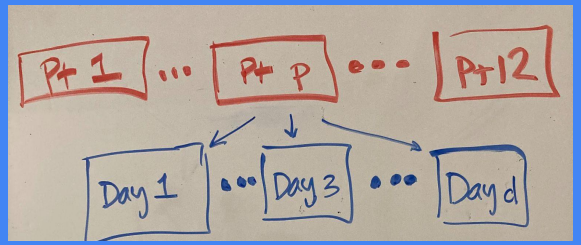




Talukder et al. 2022

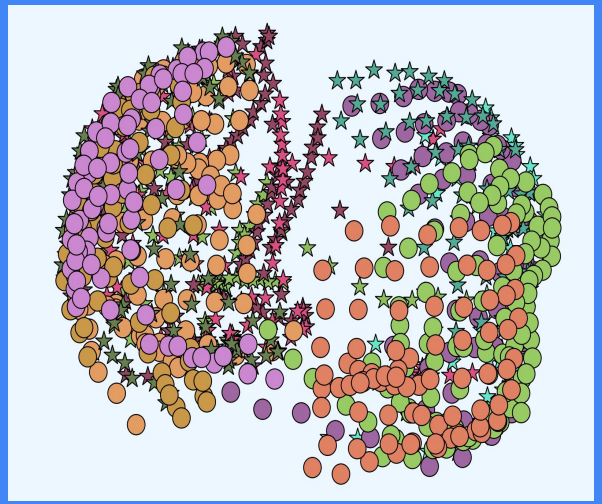
Visualize Your Data

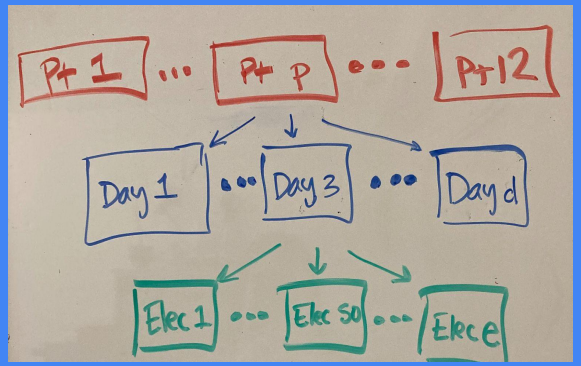




Talukder et al. 2022

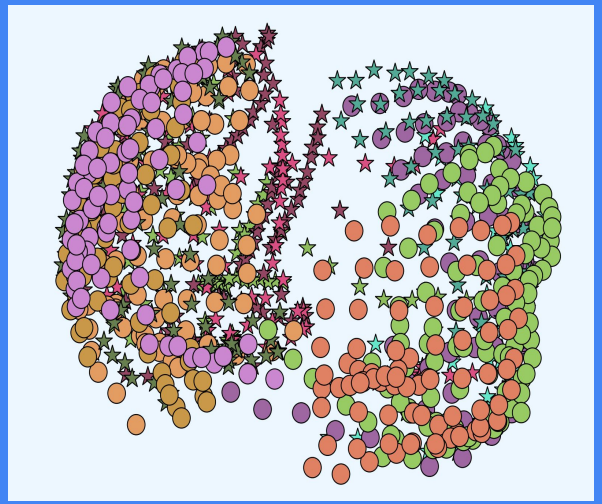
Visualize Your Data

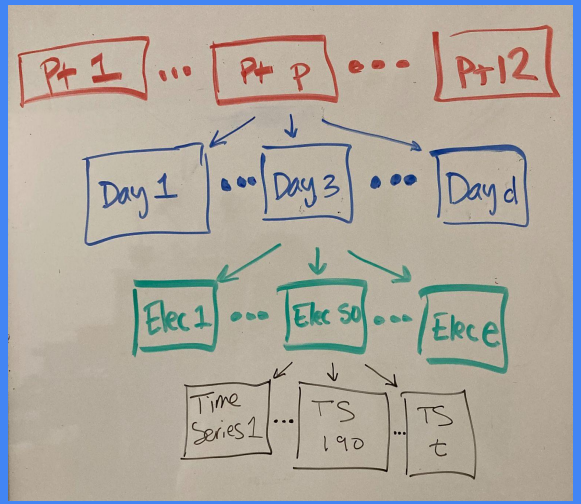




Talukder et al. 2022

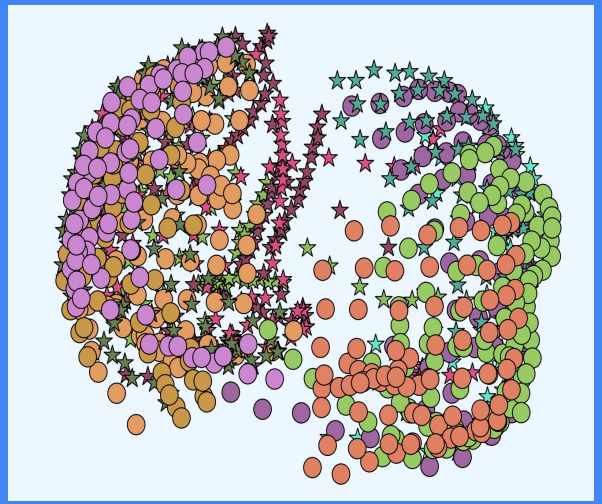
Visualize Your Data

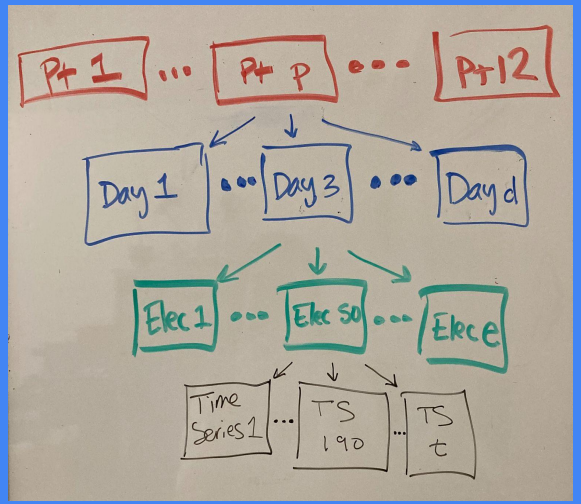


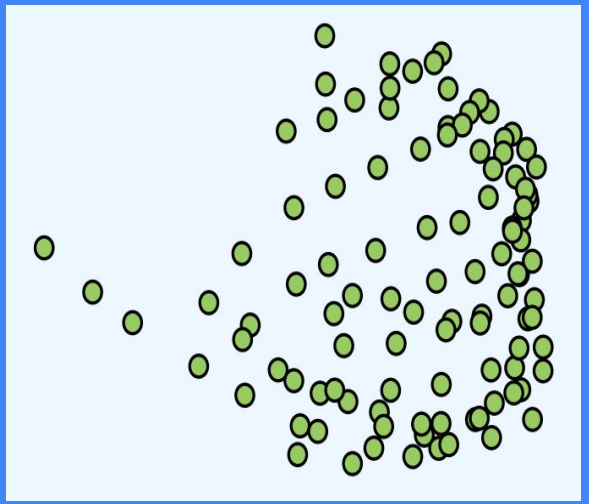


Talukder et al. 2022

Visualize Your Data

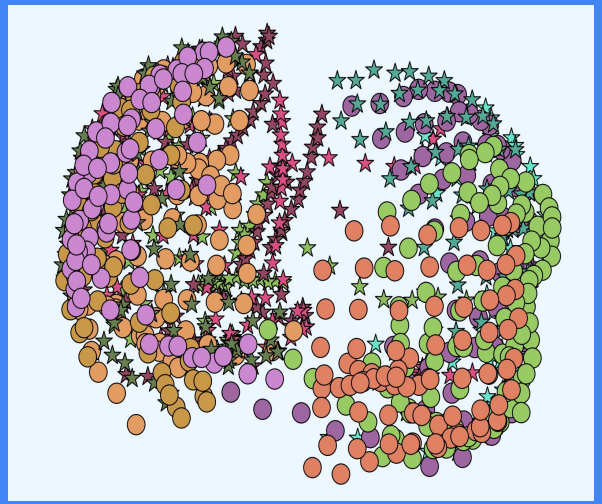


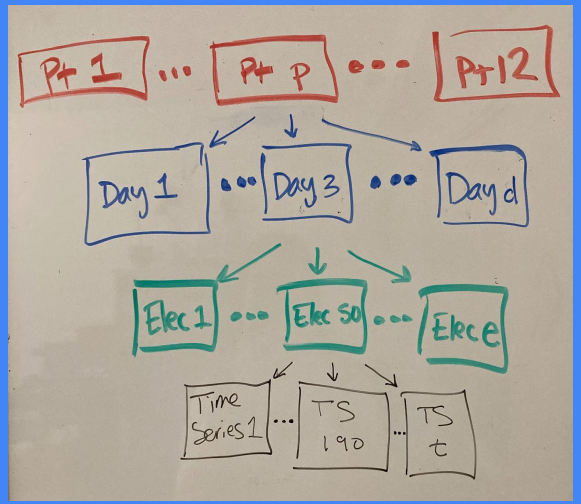


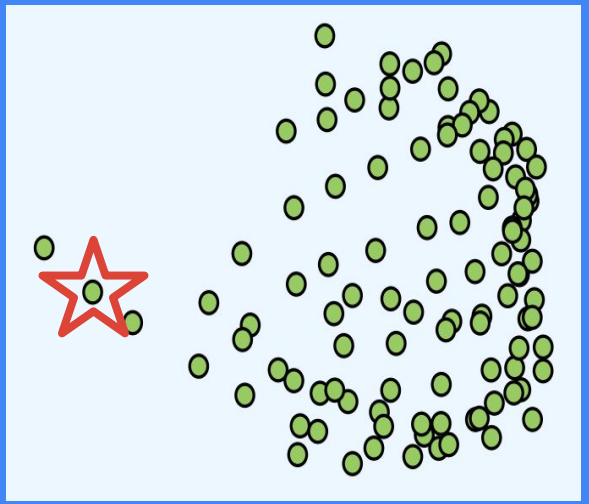


Talukder et al. 2022

Visualize Your Data



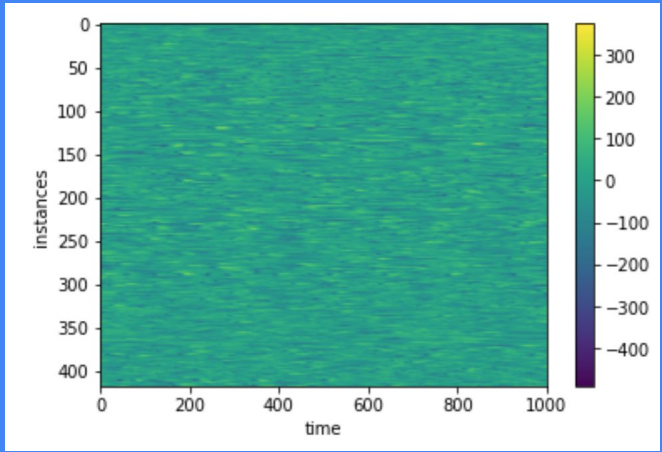




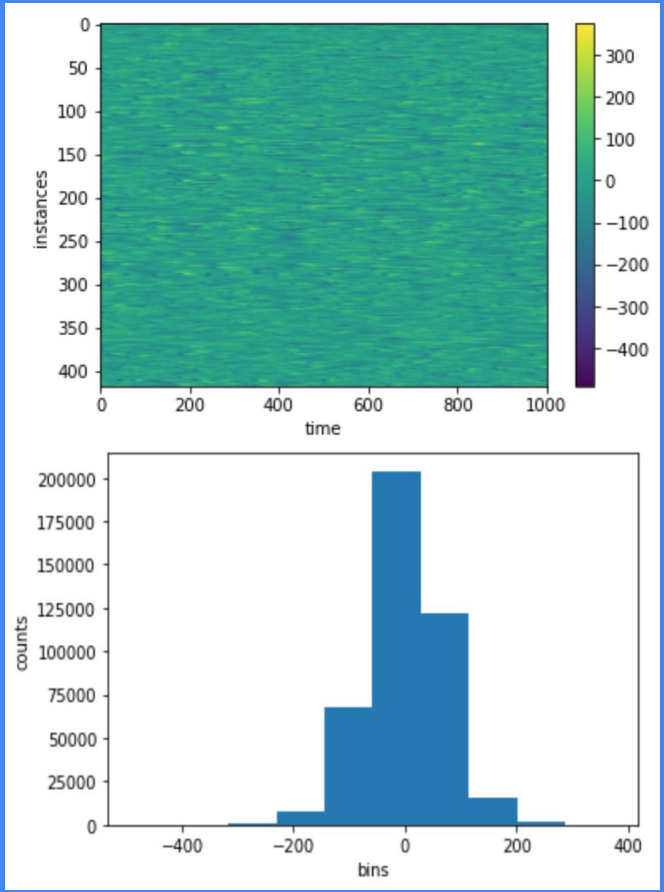
Talukder et al. 2022

Visualize Your Data

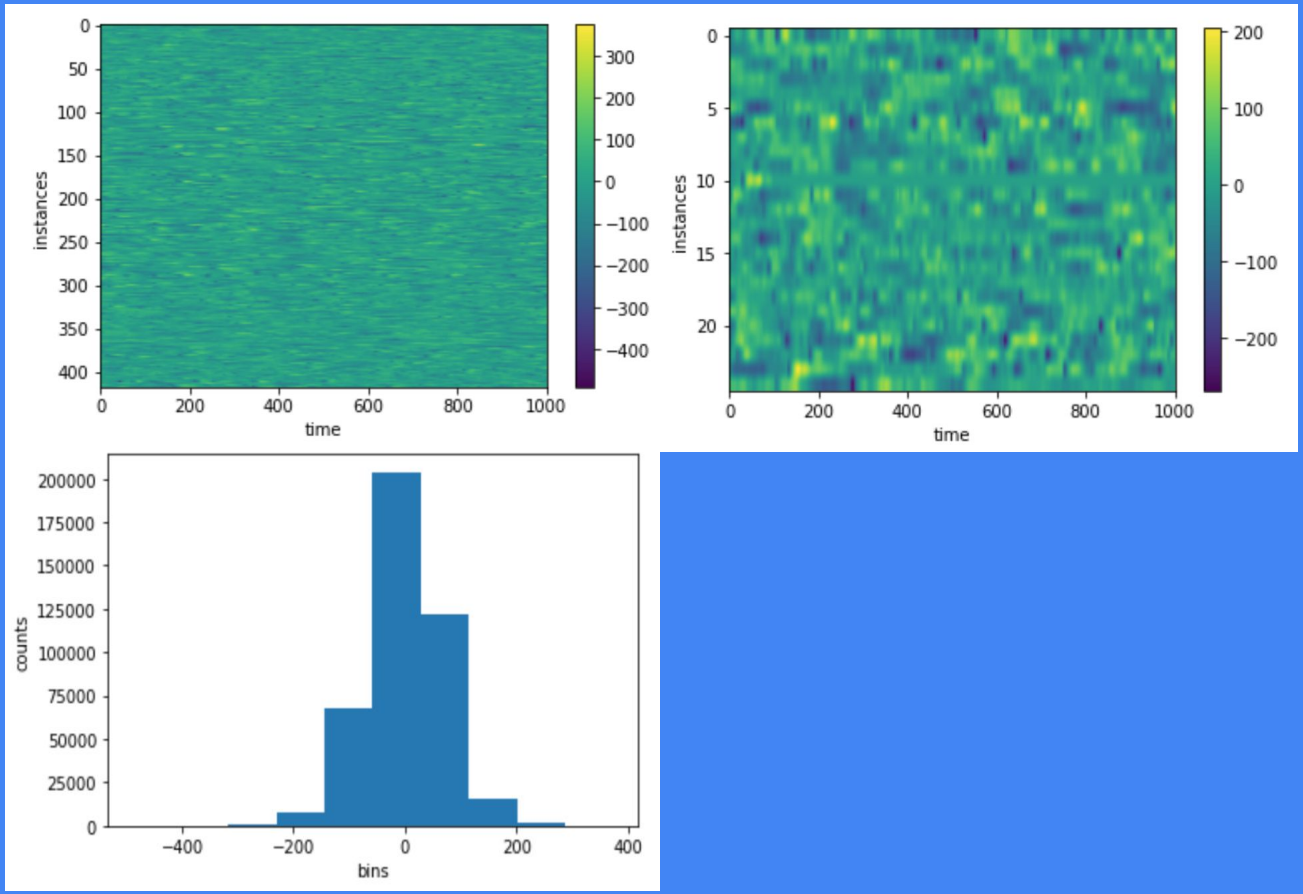
Visualize Your Data



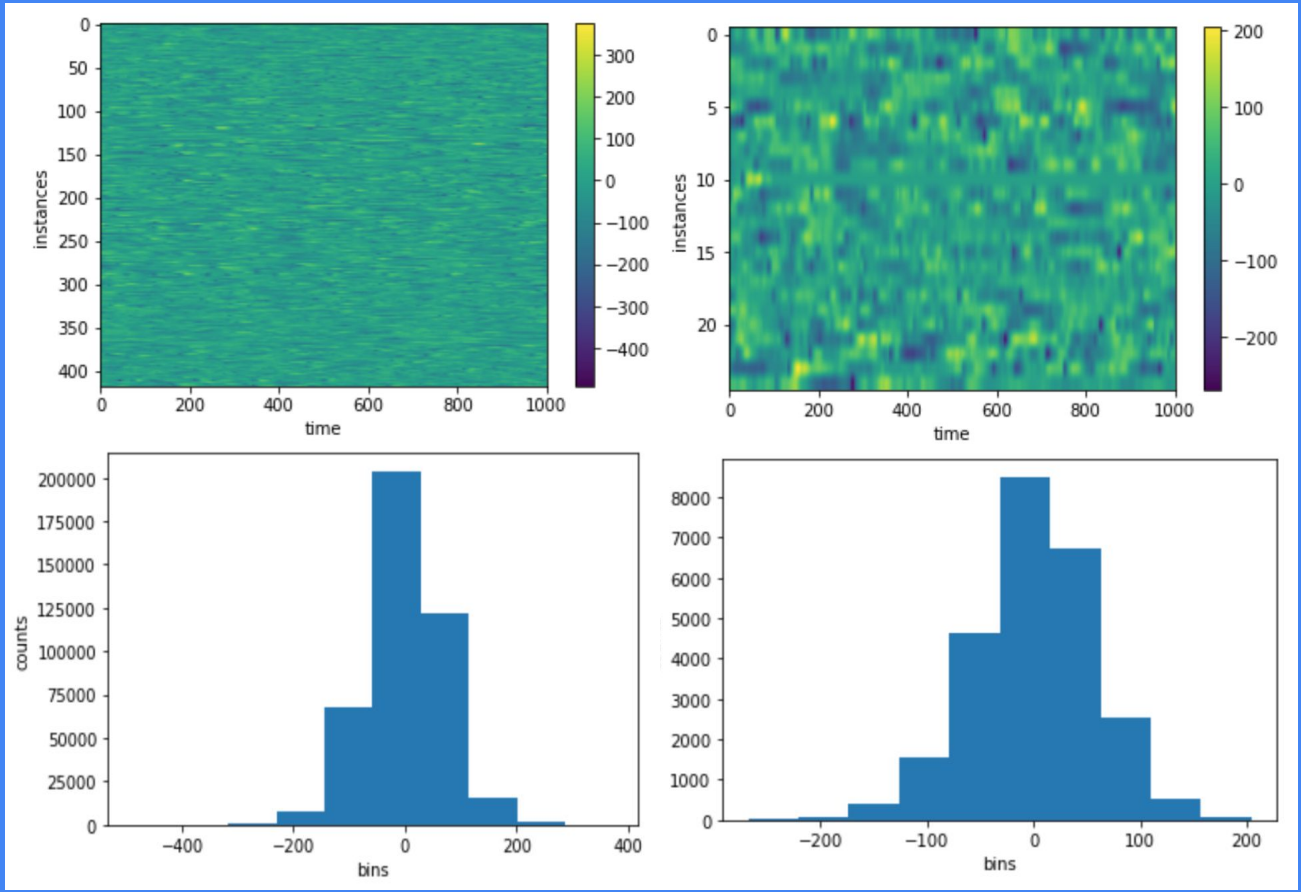
Visualize Your Data



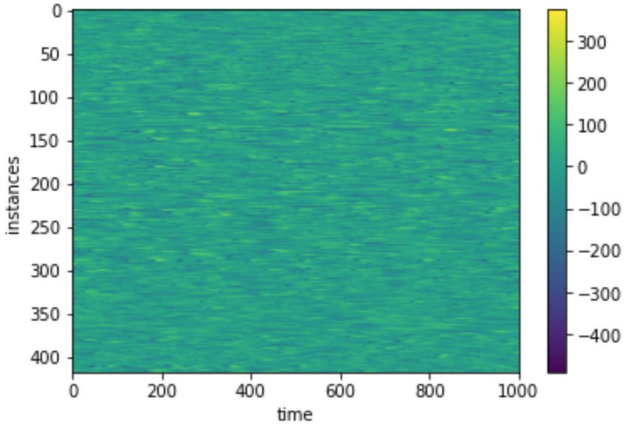
Visualize Your Data

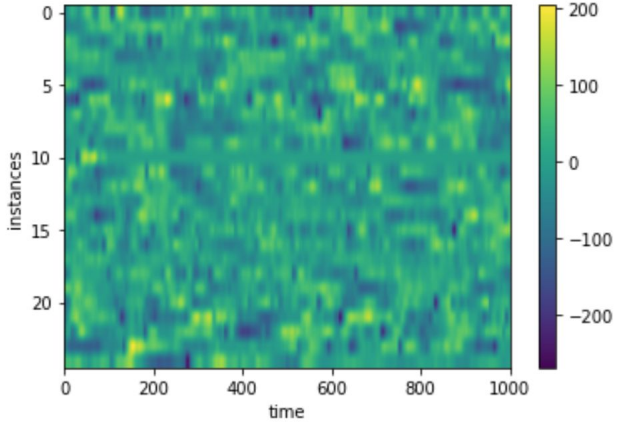


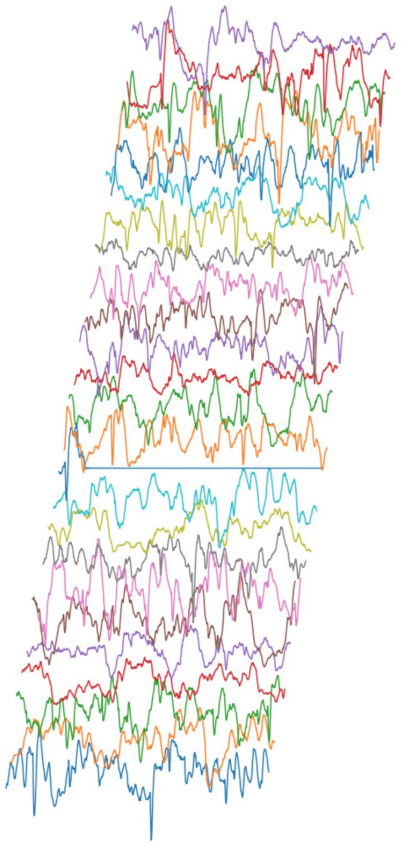
Visualize Your Data

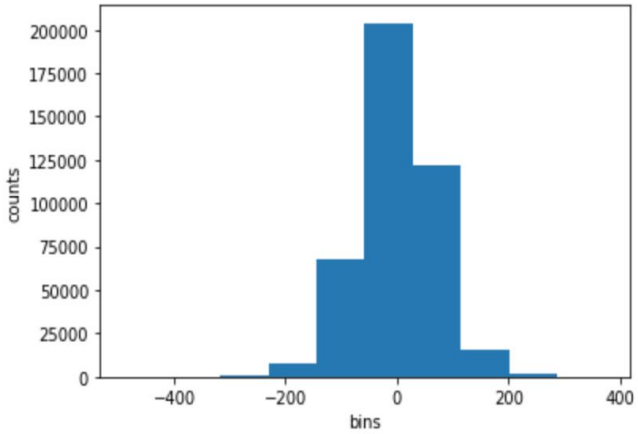


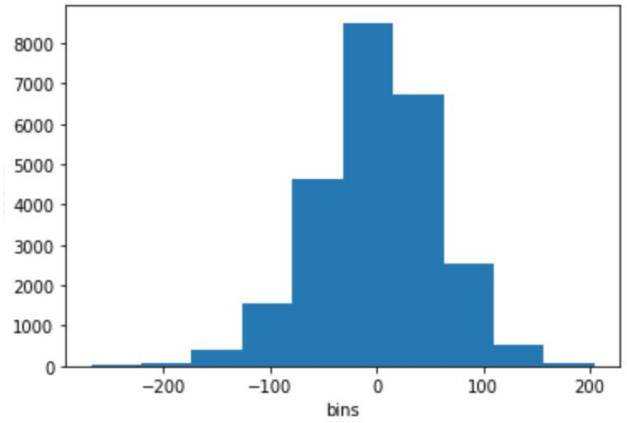
Visualize Your Data











Signal Extraction

Signal Extraction

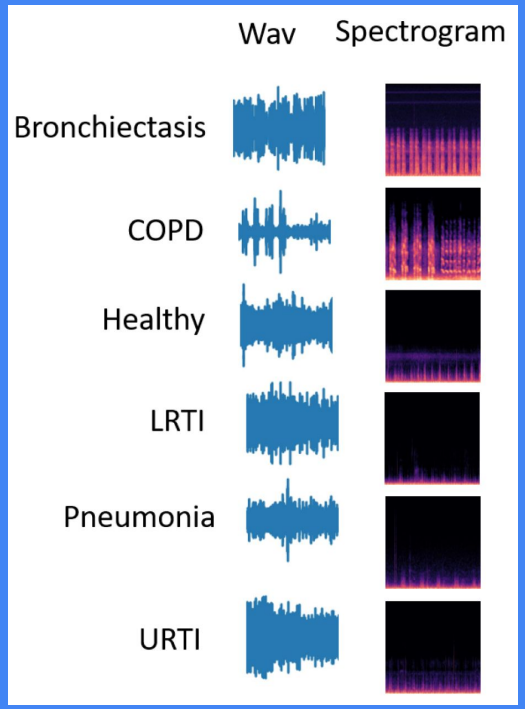
What types of signals might you have in your data?

Signal Extraction

- Lung Artifacts
- Heart Artifacts
- Stimulation Artifacts
- Movement Artifacts
- Electrical Noise
- …

Signal Extraction

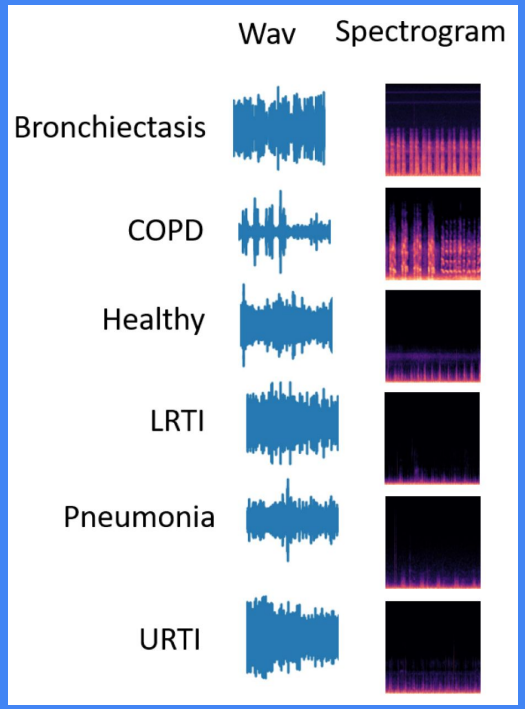
Lung Artifacts



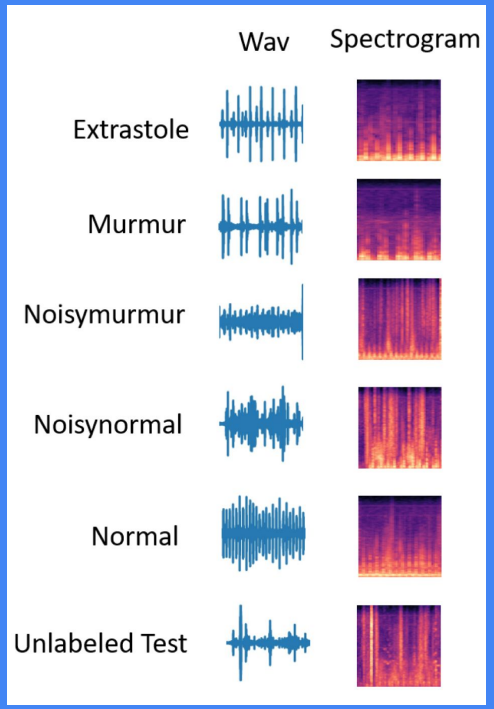
Tariq et al. 2022

Signal Extraction

Lung Artifacts

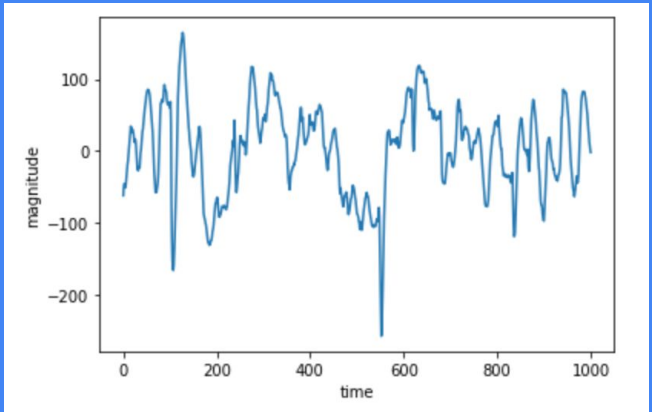


Heart Artifacts

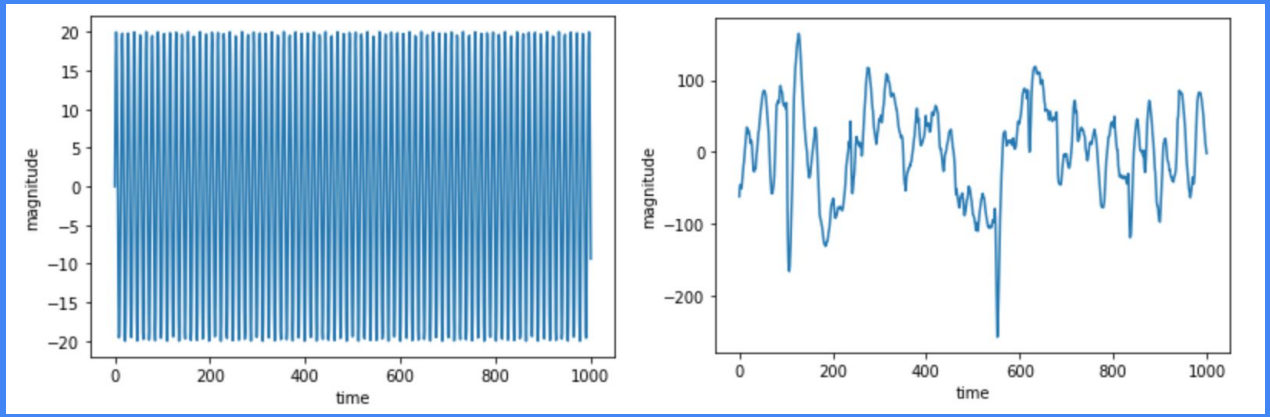


Tariq et al. 2022

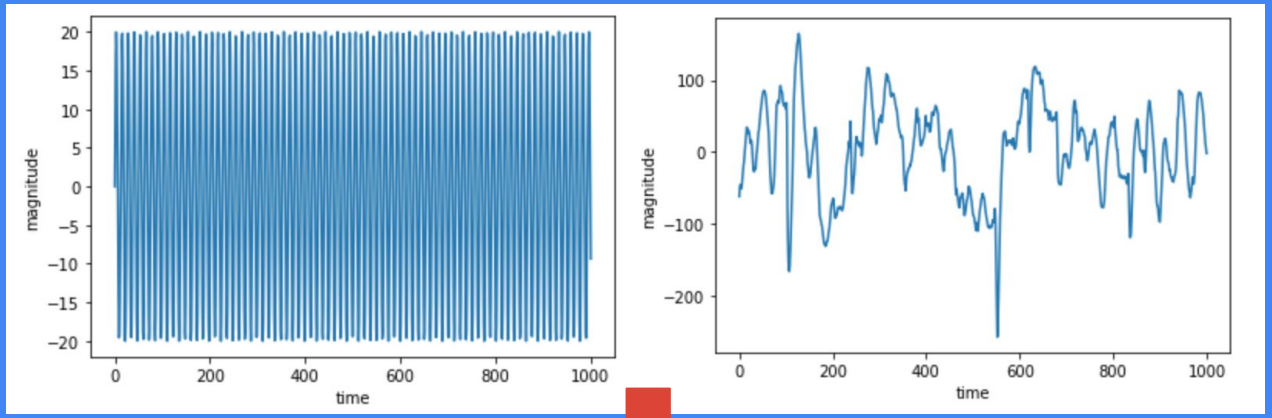
Signal Extraction

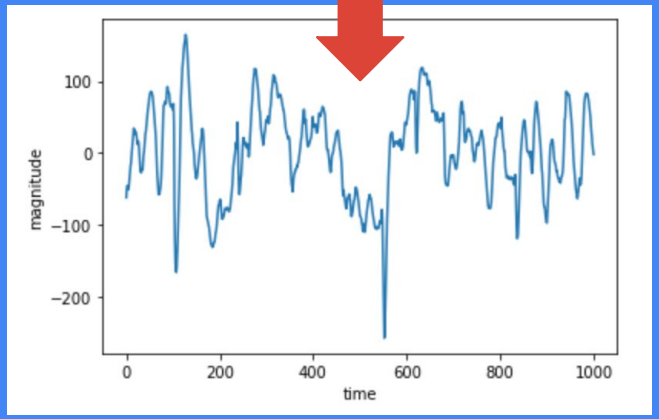


Signal Extraction

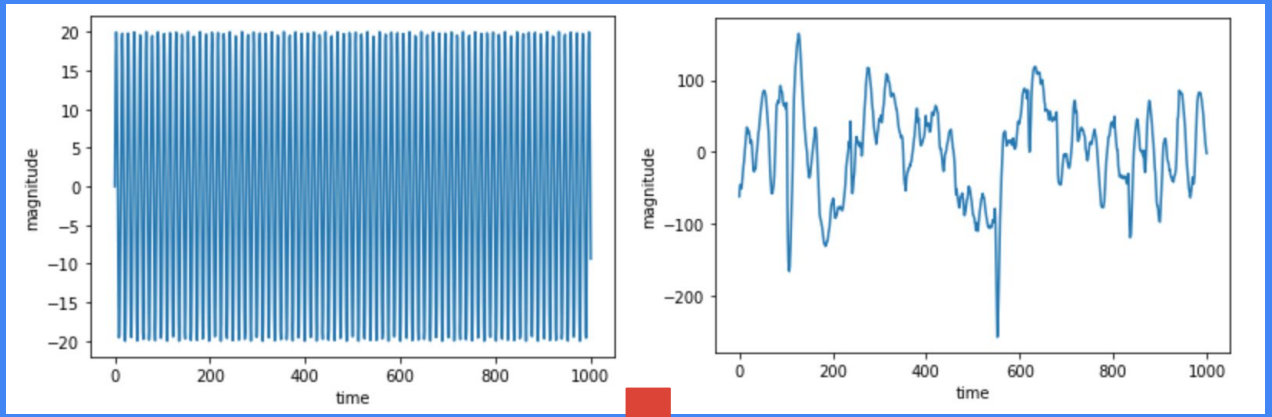


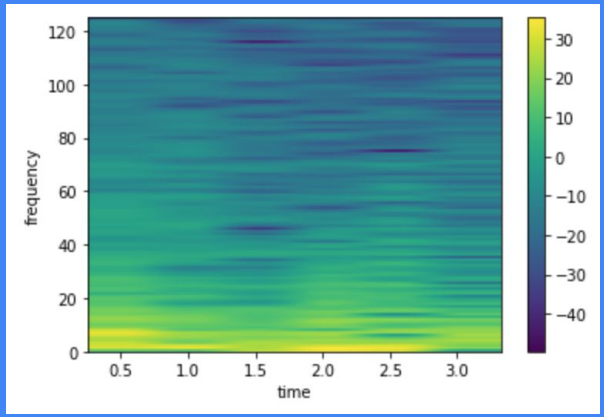
Signal Extraction

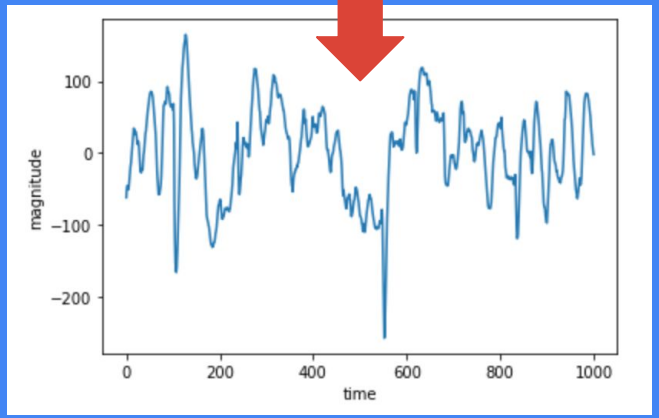




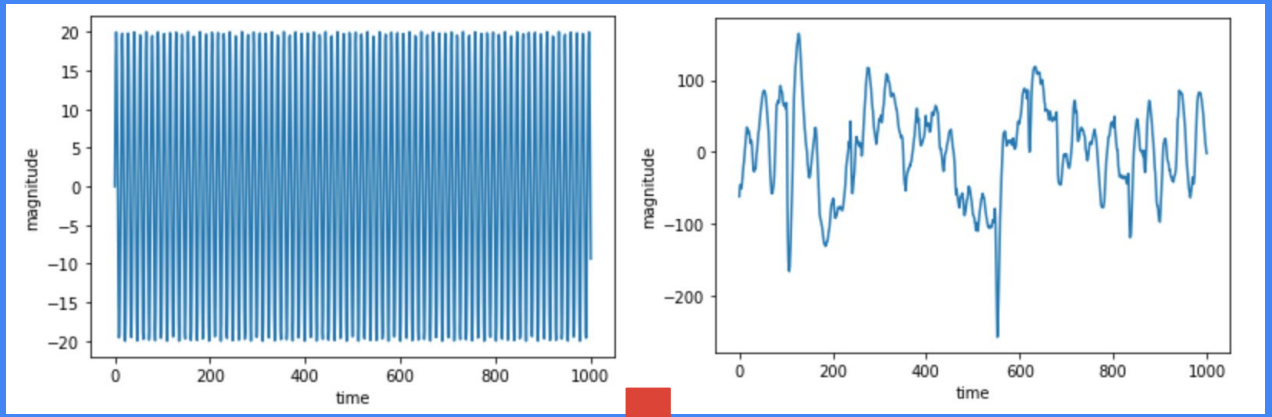
Signal Extraction

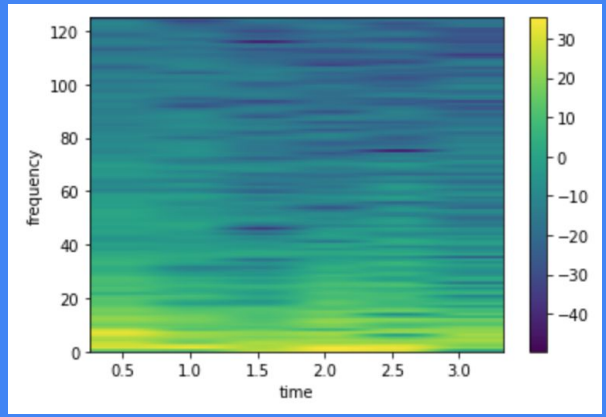


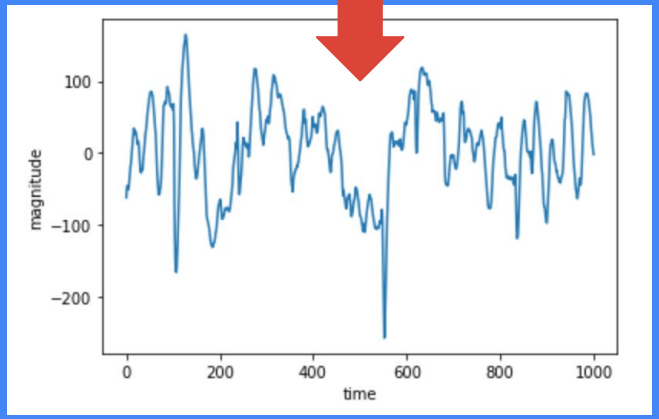


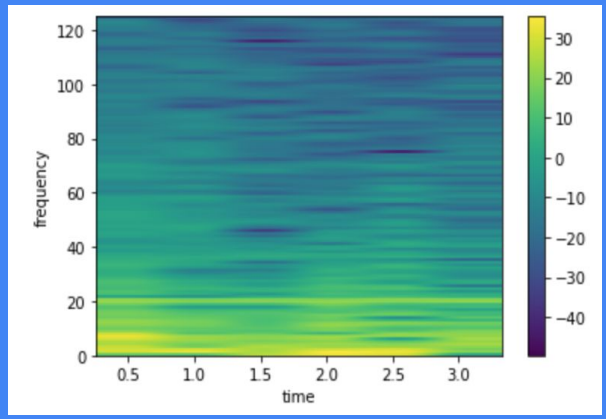


Signal Extraction









Dataset Augmentation

Dataset Augmentation

Data augmentation in data analysis are techniques used to increase the amount of data by adding slightly modified copies of already existing data or newly created synthetic data from existing data.

https://en.wikipedia.org/wiki/Data_augmentation

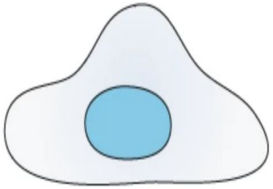
Dataset Augmentation

Data augmentation in data analysis are techniques used to increase the amount of data by adding slightly modified copies of already existing data or newly created synthetic data from existing data.

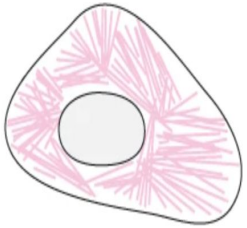
It acts as a regularizer and helps reduce overfitting when training a machine learning model.

https://en.wikipedia.org/wiki/Data_augmentation

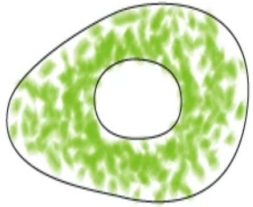
Dataset Augmentation



→ Nuclear



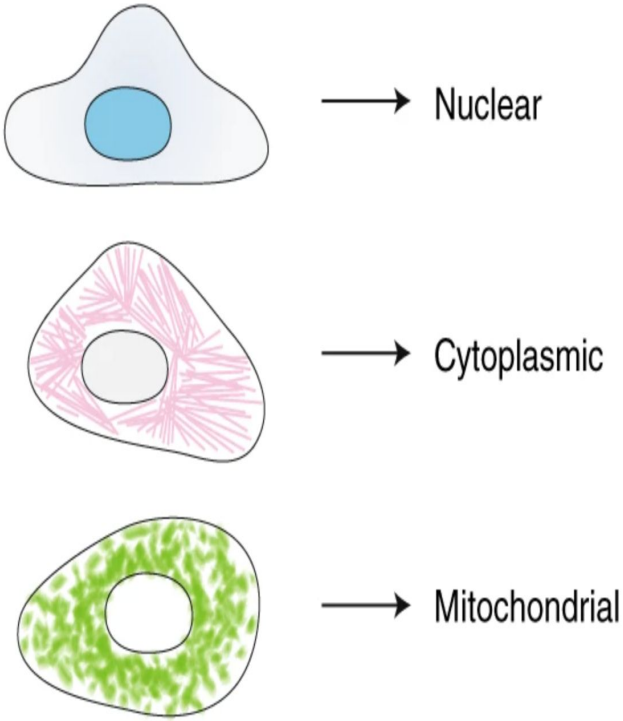
→ Cytoplasmic

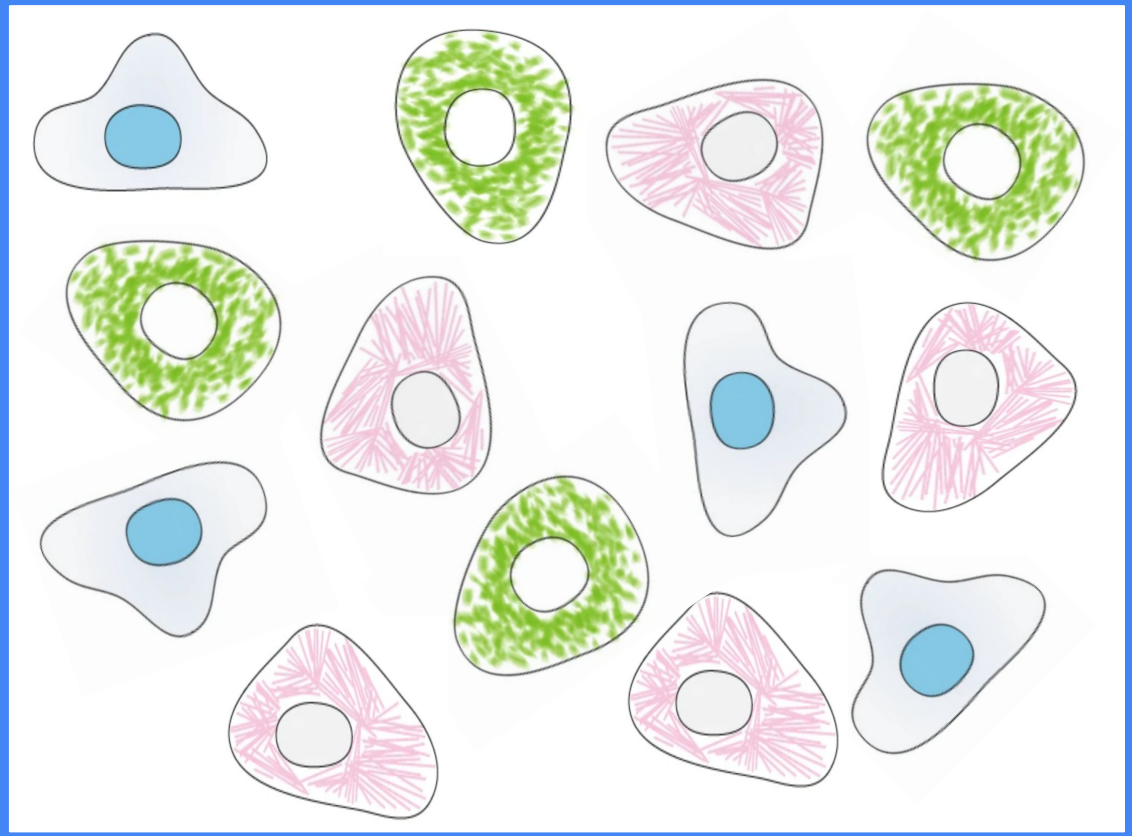


→ Mitochondrial

Moen et al. 2019

Dataset Augmentation





Moen et al. 2019

Normalization / Standardization

Normalization / Standardization

Normalization: Is adjusting values measured on different scales to a notionally common scale… [or] more sophisticated adjustments where the intention is to bring the entire probability distributions of adjusted values into alignment.

https://en.wikipedia.org/wiki/Normalization_(statistics)
https://en.wikipedia.org/wiki/Standard_score

Normalization / Standardization

Normalization: Is adjusting values measured on different scales to a notionally common scale… [or] more sophisticated adjustments where the intention is to bring the entire probability distributions of adjusted values into alignment.

Standardization: Is when we subtract the population mean from an individual raw score and then dividing the difference by the population standard deviation (aka Z-scoring).

https://en.wikipedia.org/wiki/Normalization_(statistics)
https://en.wikipedia.org/wiki/Standard_score

Normalization / Standardization

Normalization: Is adjusting values measured on different scales to a notionally common scale… [or] more sophisticated adjustments where the intention is to bring the entire probability distributions of adjusted values into alignment.

Standardization: Is when we subtract the population mean from an individual raw score and then dividing the difference by the population standard deviation (aka Z-scoring).

$$z = \frac{x - \mu}{\sigma}$$

https://en.wikipedia.org/wiki/Normalization_(statistics)
https://en.wikipedia.org/wiki/Standard_score

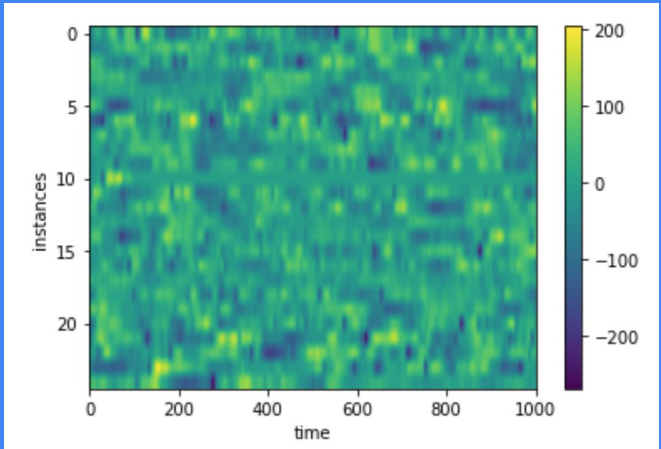
Normalization / Standardization

What does this help correct for in our data?

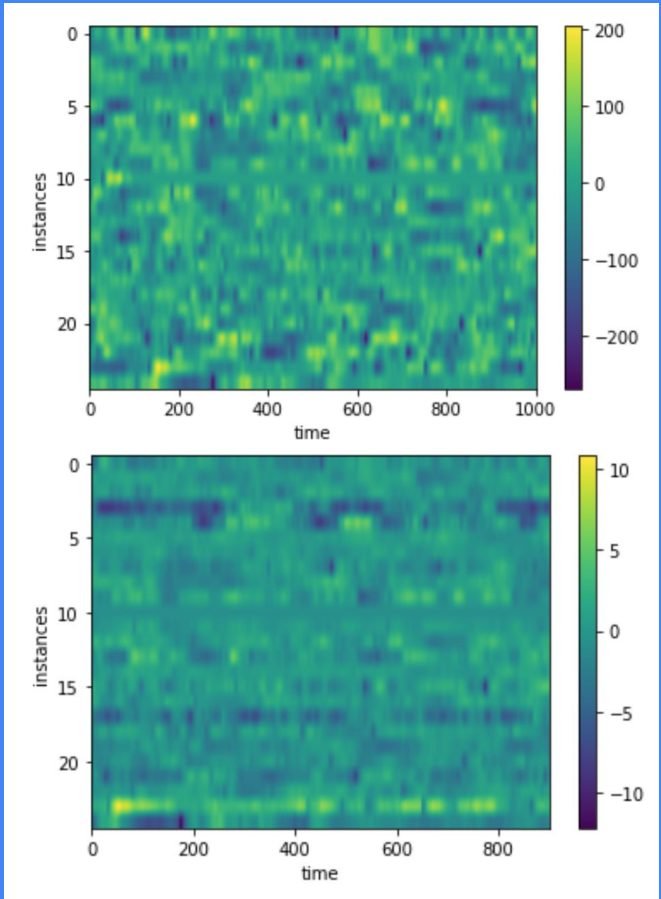
Normalization / Standardization

- Neural population drift across time.
- Electrode shift across days.
- Different dynamic ranges across patients.
- Habituation to stimuli.
- …

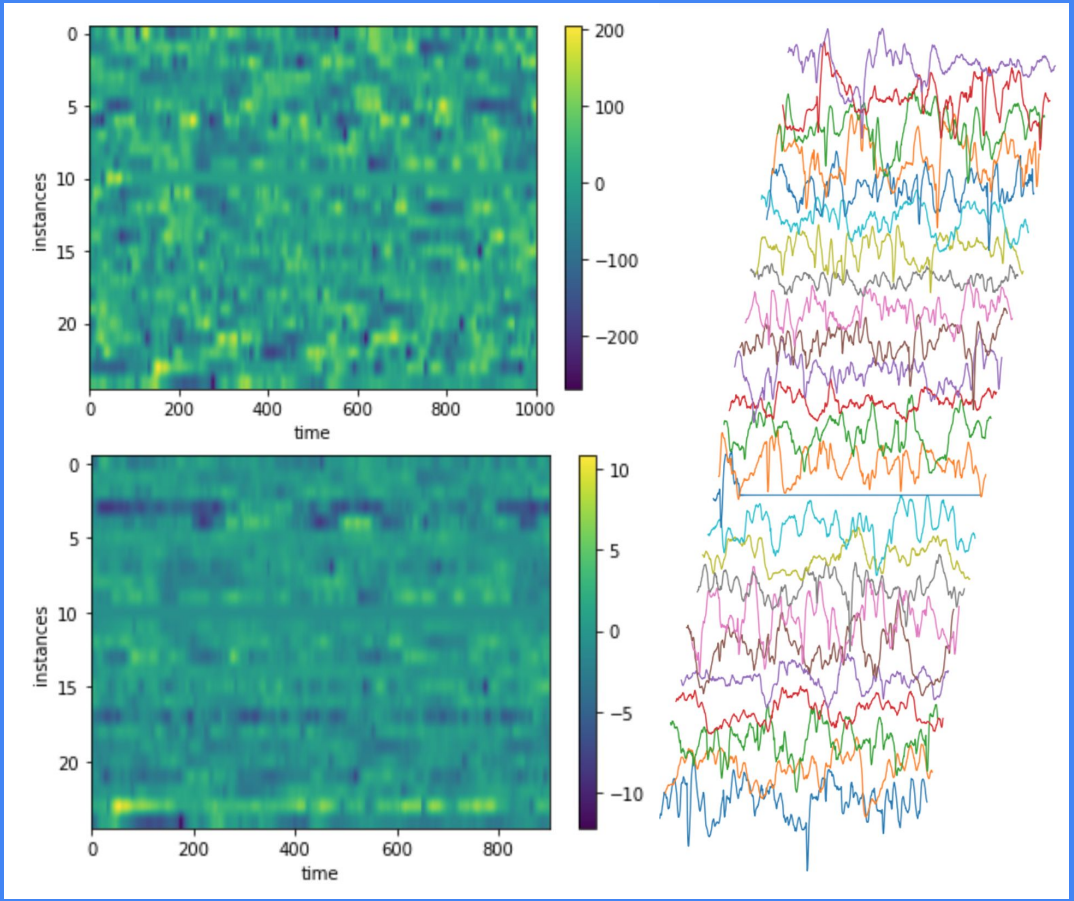
Normalization / Standardization



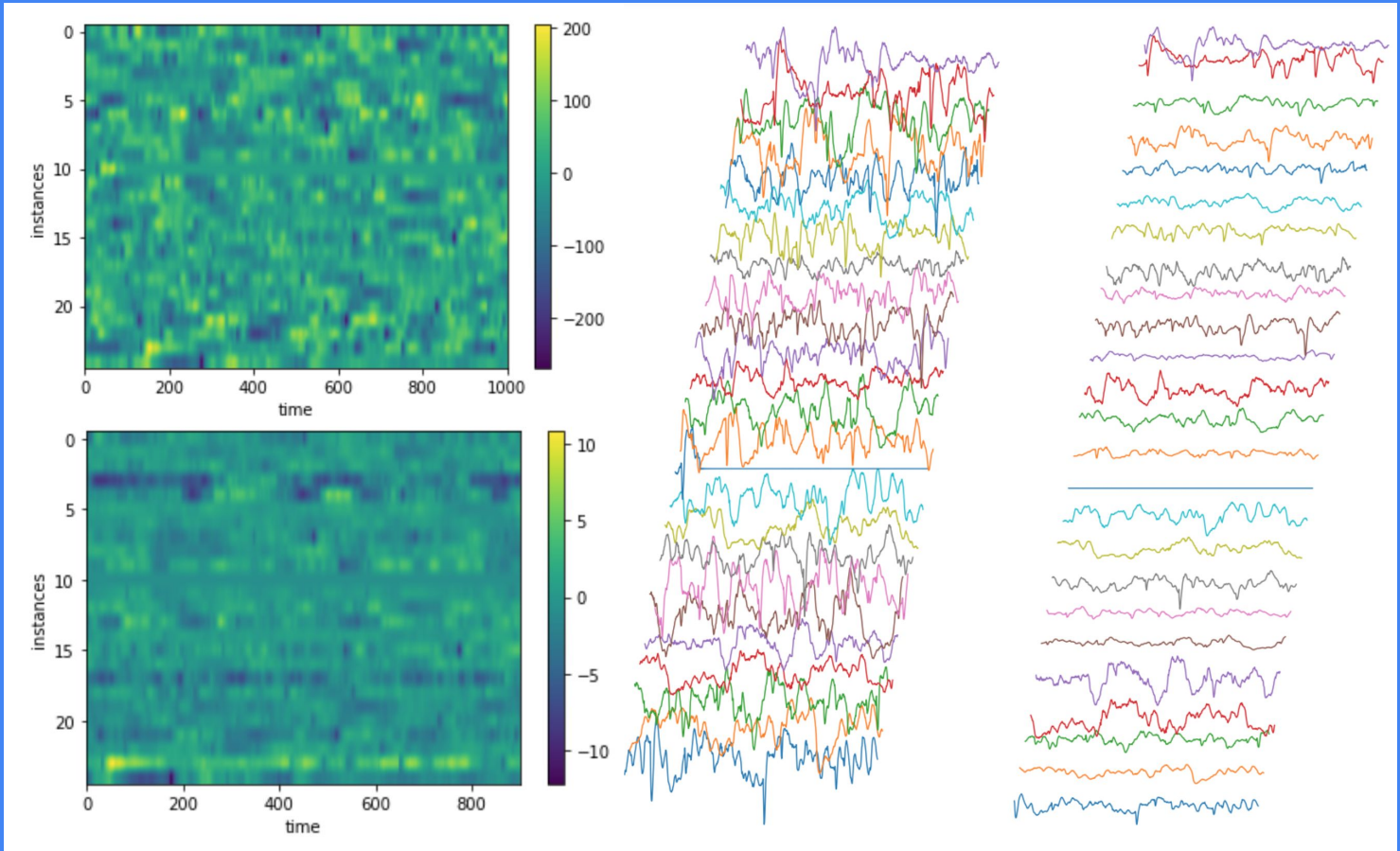
Normalization / Standardization



Normalization / Standardization

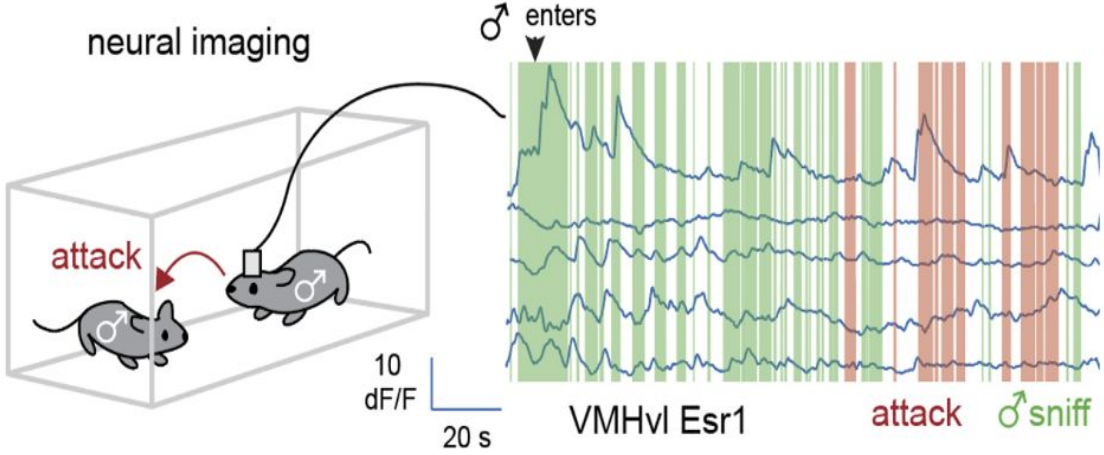


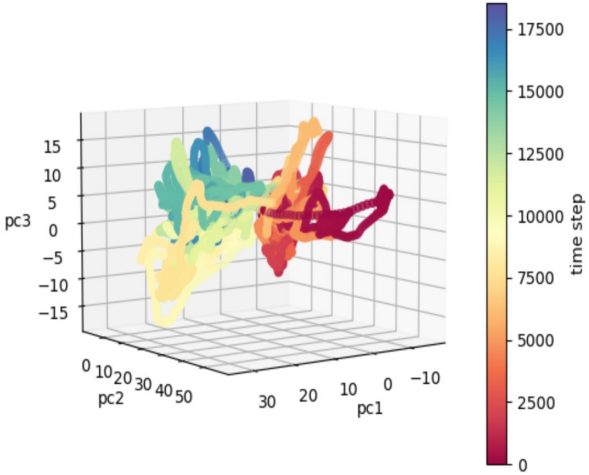
Normalization / Standardization



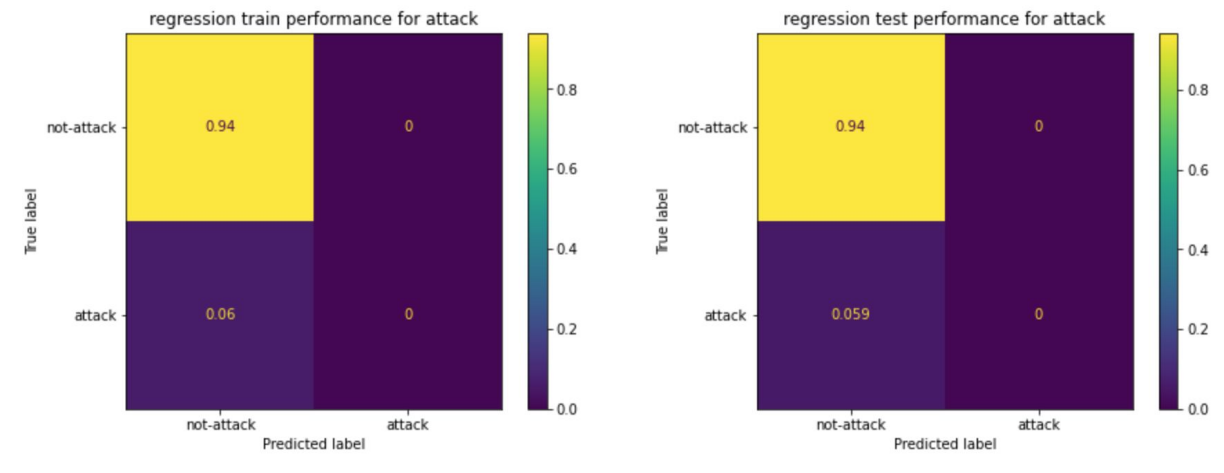
Train / Validation / Test Splits

Train / Validation / Test Splits

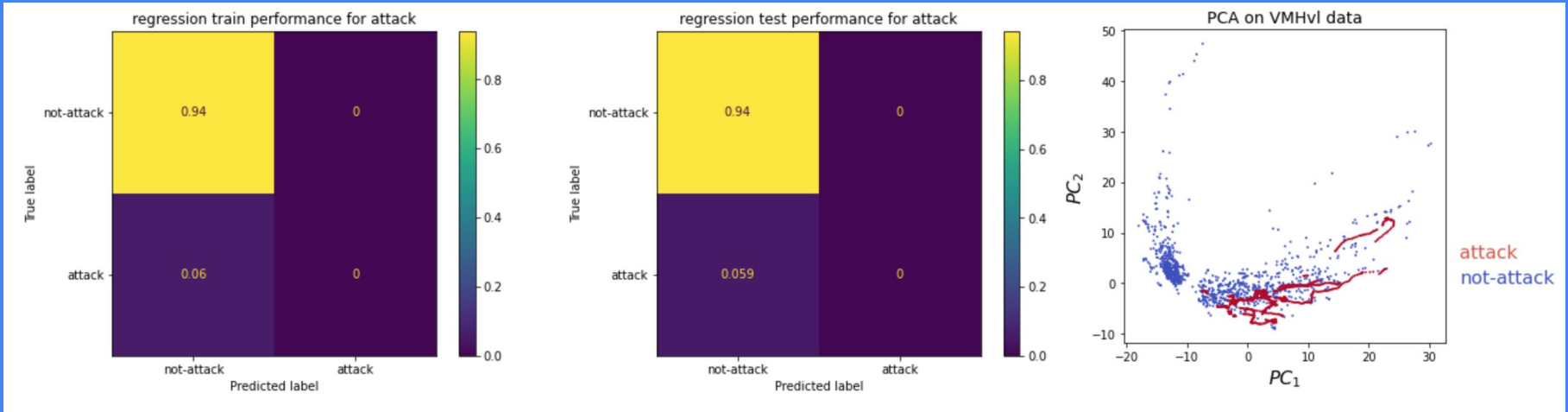




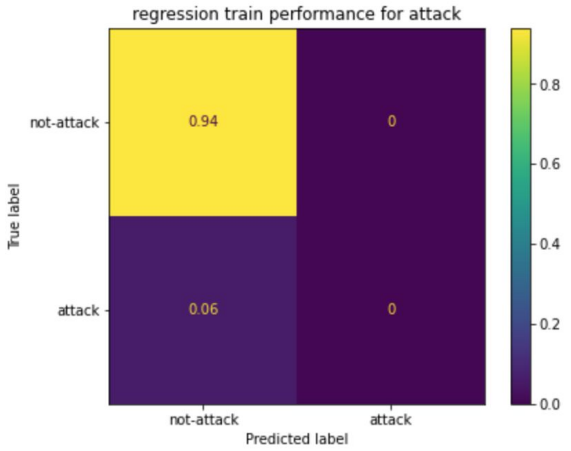
Train / Validation / Test Splits

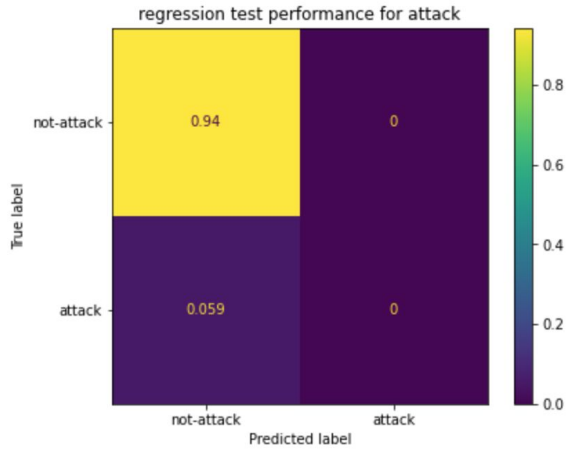


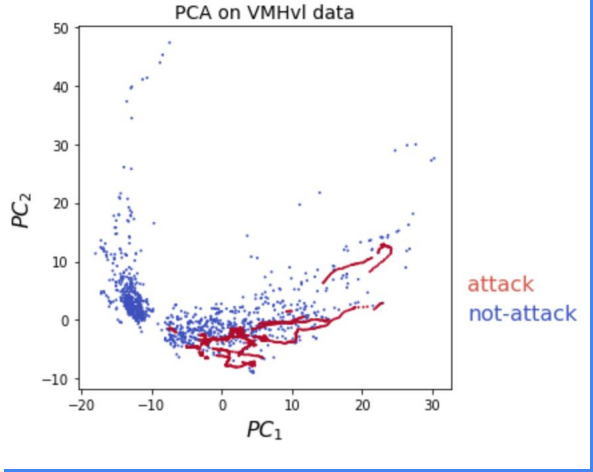
Train / Validation / Test Splits

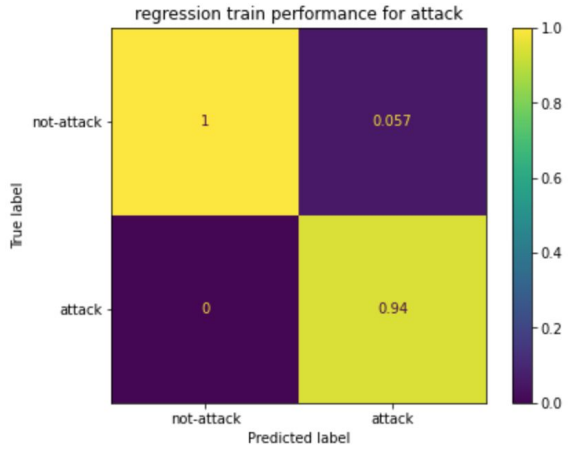


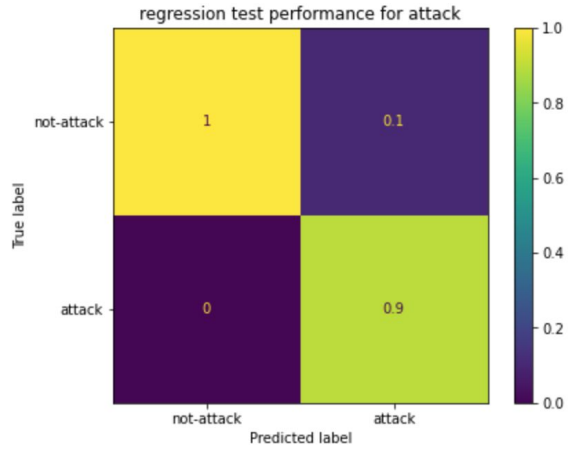
Train / Validation / Test Splits











Train / Validation / Test Splits

https://github.com/SaberaTalukder/Chen_Institute_DataSAI_for_Neuroscience/blob/main/07_05_22_day1_overview/code/diy_notebooks/dataset_engineering.ipynb