# Conditional Diffusion for EEG Data Synthesis

Semester Project for Advanced Machine Learning 2024
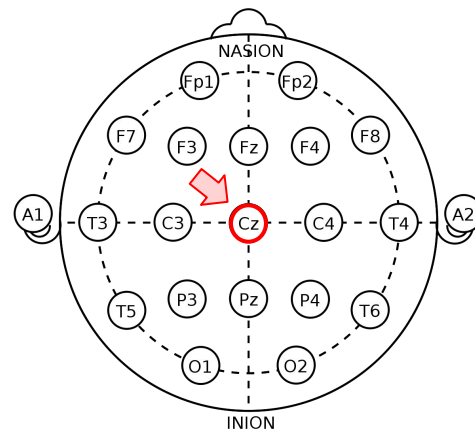by
Group 18

Introduction

- EEG: the brain activity of different areas of the cerebral <u>outer cortices</u>
- Dataset (TUH EEG) with labels for different kinds of seizure activity
- Generally, seizures are slow waves between 2 - 10 Hz, hyper-synchronous in time
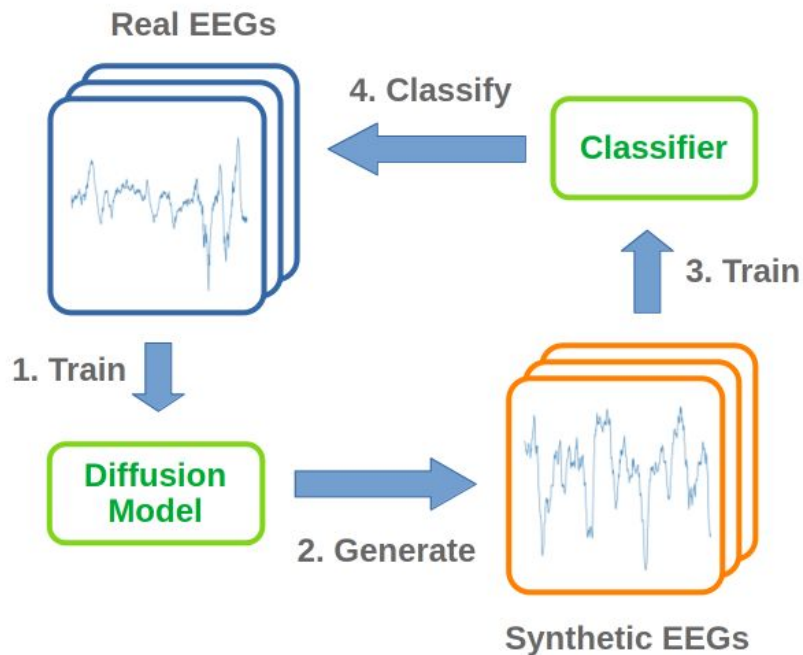- Idea: Generate synthetic EEG with seizure and non-seizure activity
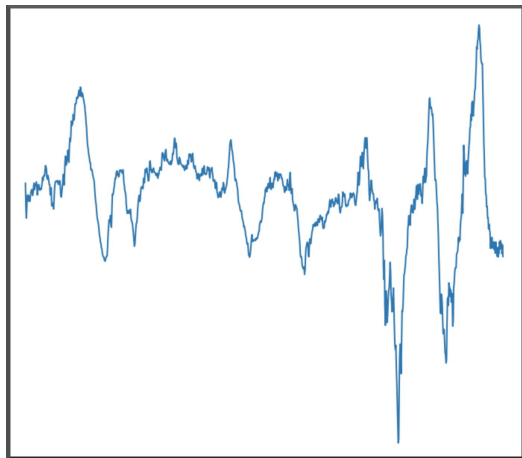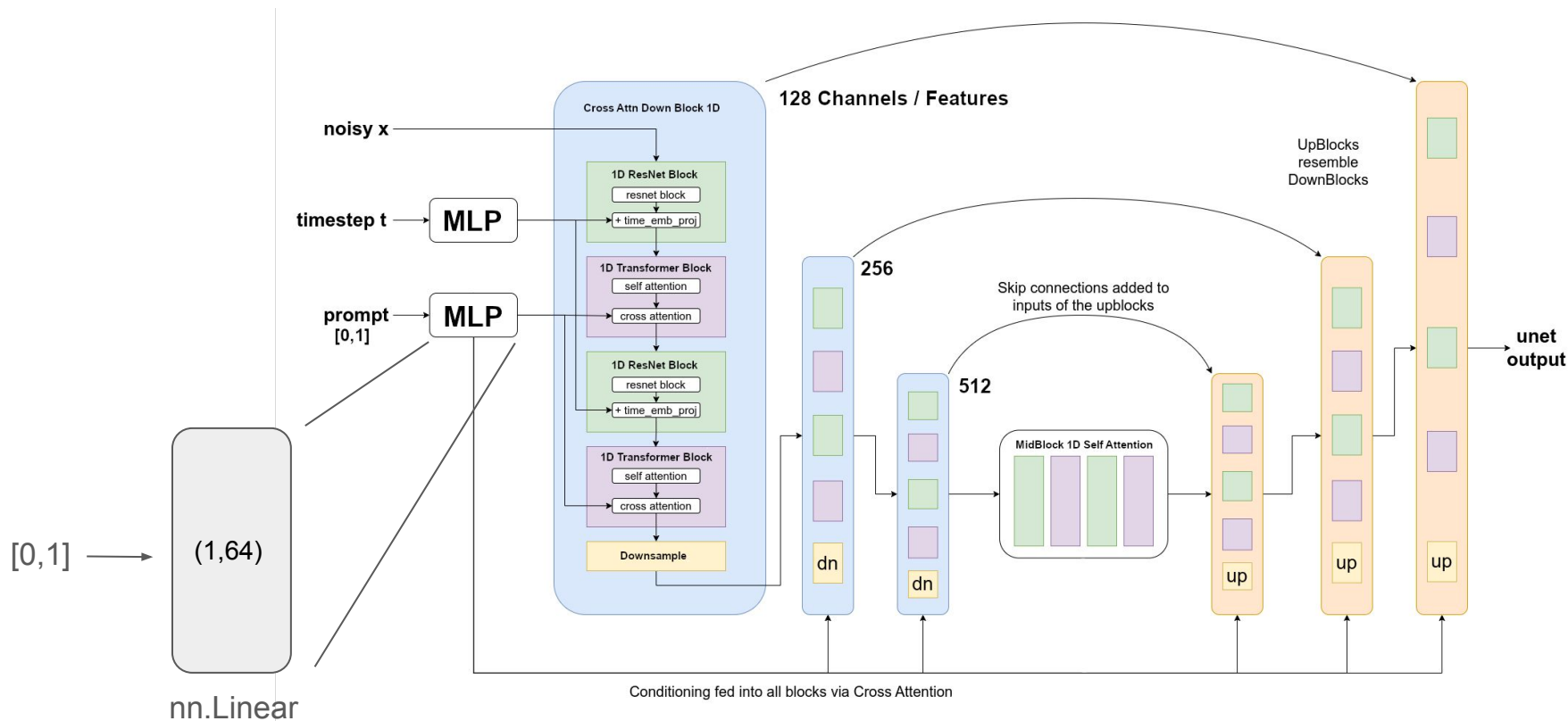

Healthy EEG
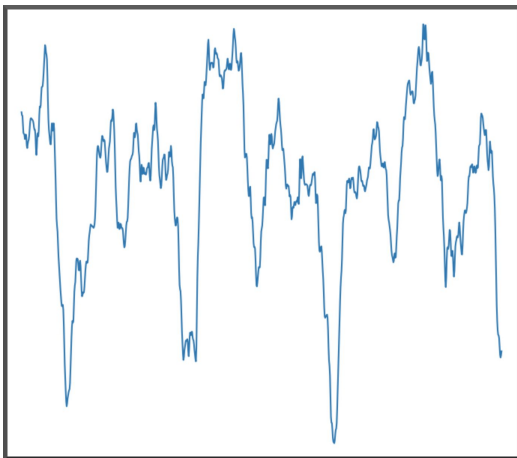
Seizure EEG

Overview

1. Train a diffusion model using real EEGs with two targets (seizure, non-seizure)

2. Use diffusion model to sample synthetic EEGs (conditioned by target)

3. Use synthetic EEGs to train a (simple) classifier model
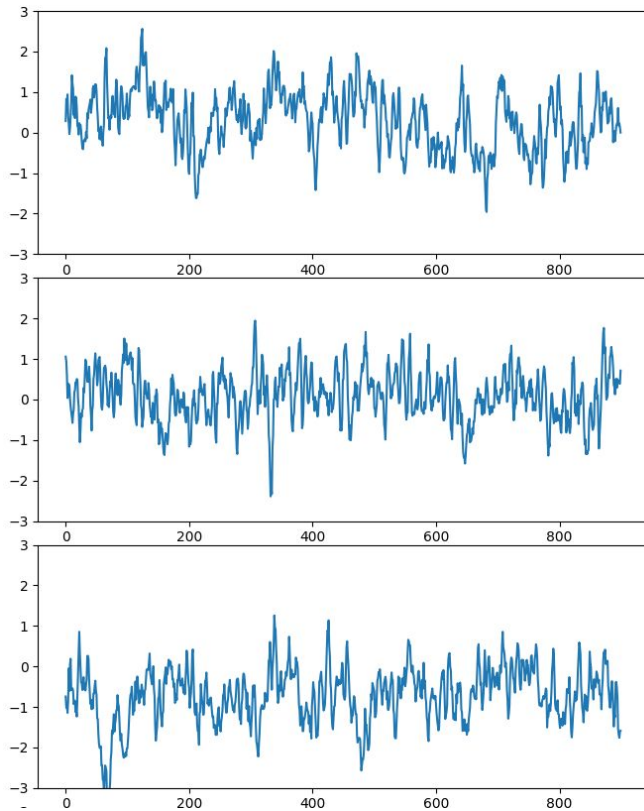
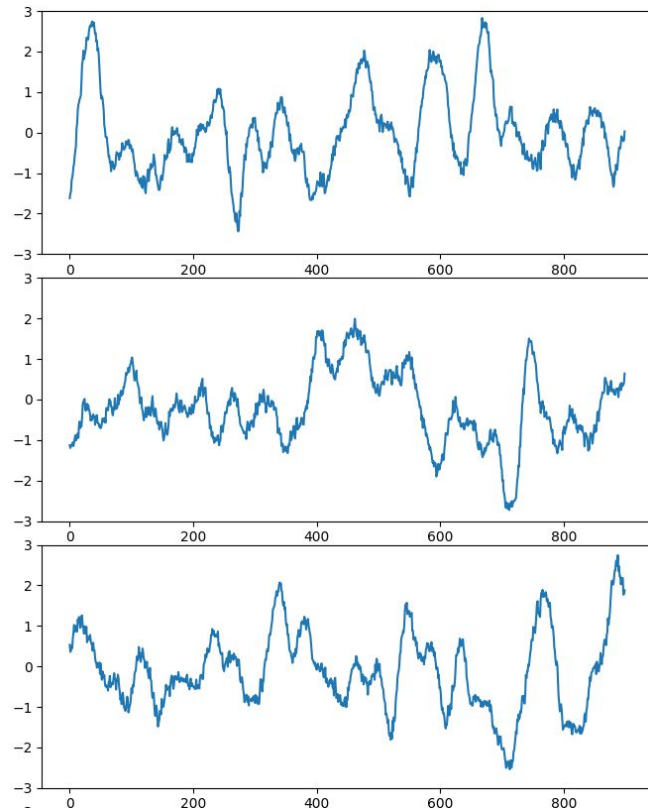4. Evaluate classifier model on real EEGs

Model

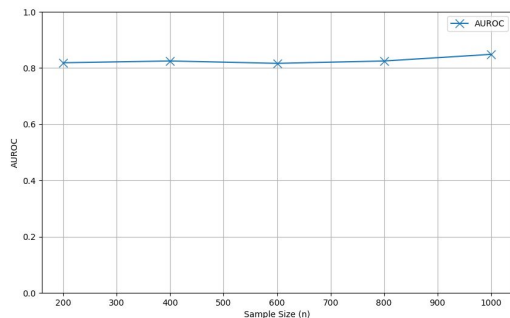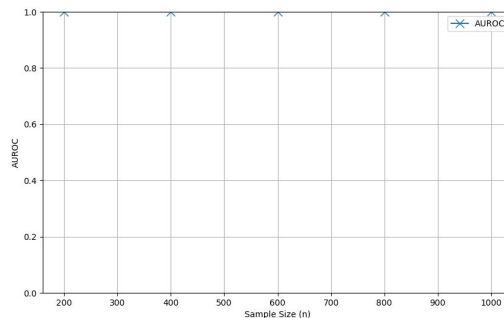# Results

non seizure

(0)

fnsz

(1)

Validation

(w/ perfect class balance)

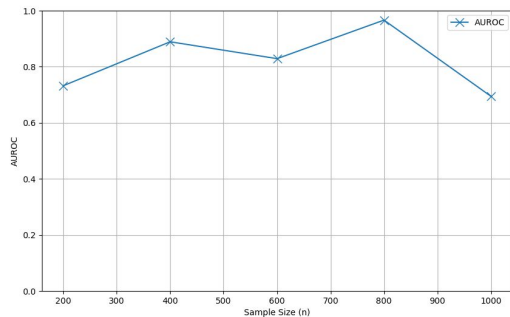Accuracy: 0.80
TPR: 0.89
TNR: 0.71
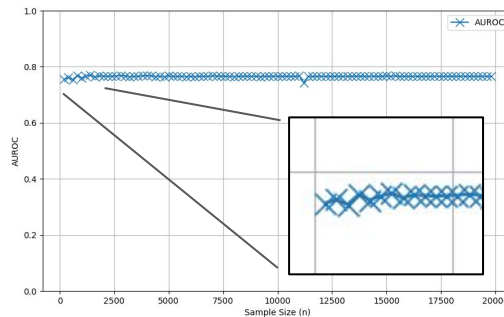F1 score: 0.81
AUROC: 0.85

a) Train: Real - Test: Real

Accuracy: 1.0
TPR: 1.0
TNR: 1.0
F1 score: 1.0
AUROC: 1.0 😎

b) Train: Synth - Test: Synth

Accuracy: 0.5
TPR: 1.0
TNR: 0.0
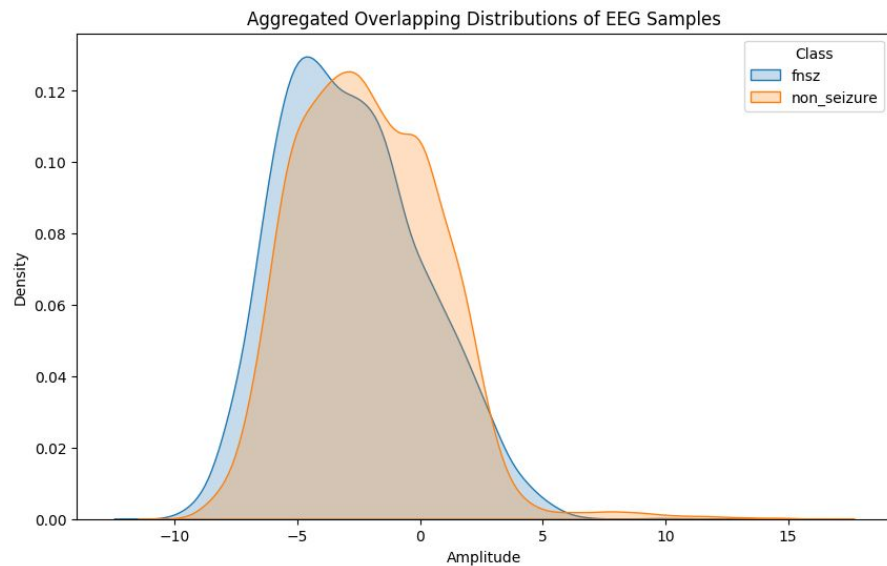F1 score: 0.66
AUROC: 0.70

c) Train: Real - Test: Synth

Accuracy: 0.57
TPR: 0.30
TNR: 0.84
F1 score: 0.41
AUROC: 0.77

c) Train: Synth - Test: Real
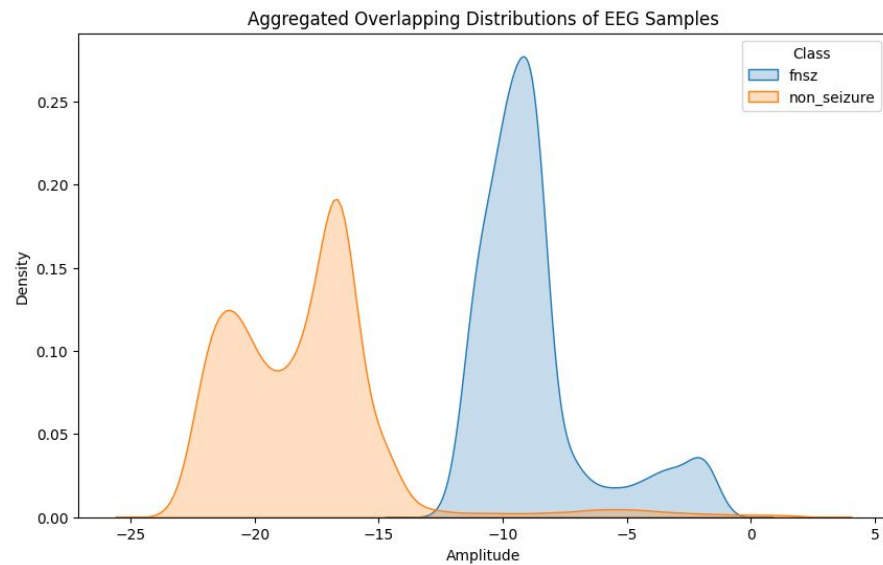
Thank you for your attention

Questions?

Appendix A
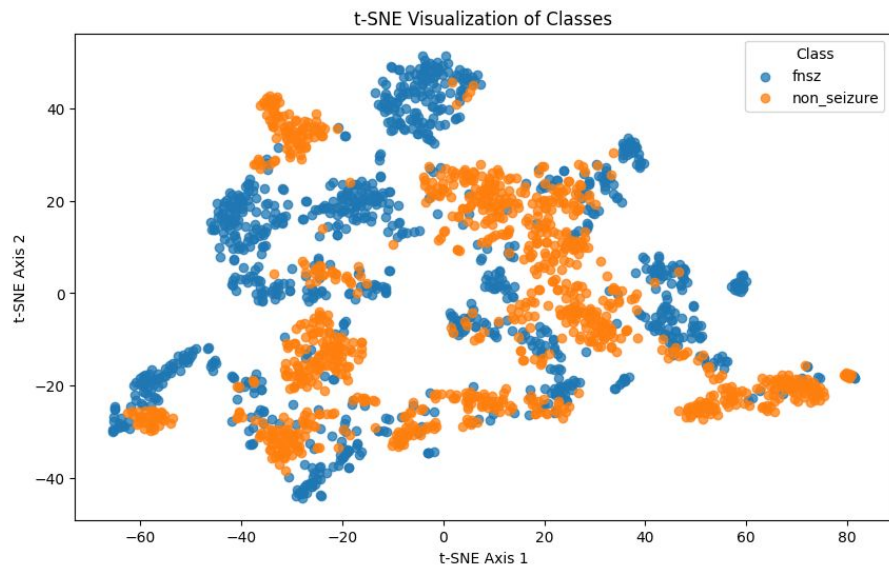

Aggregated Overlapping Distributions of EEG Samples

a) Real samples

b) Synthetic samples

Frequency features
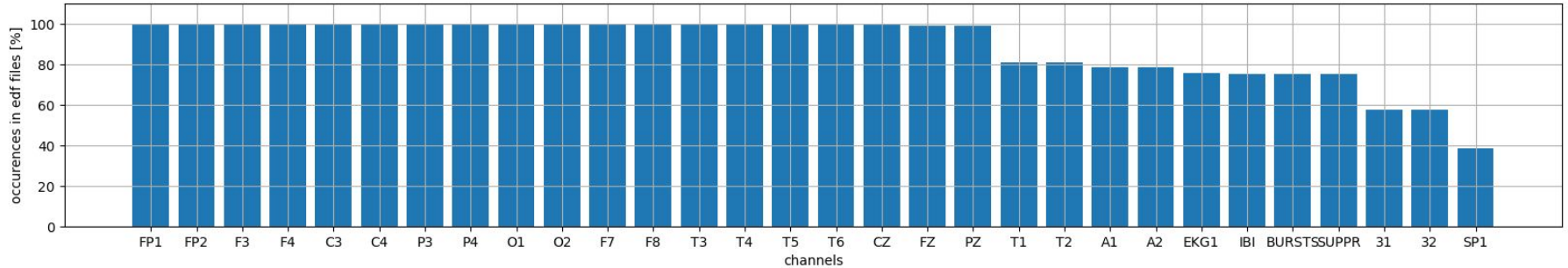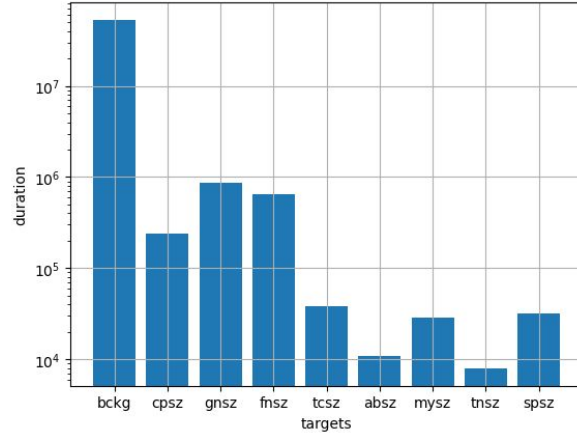
Appendix B

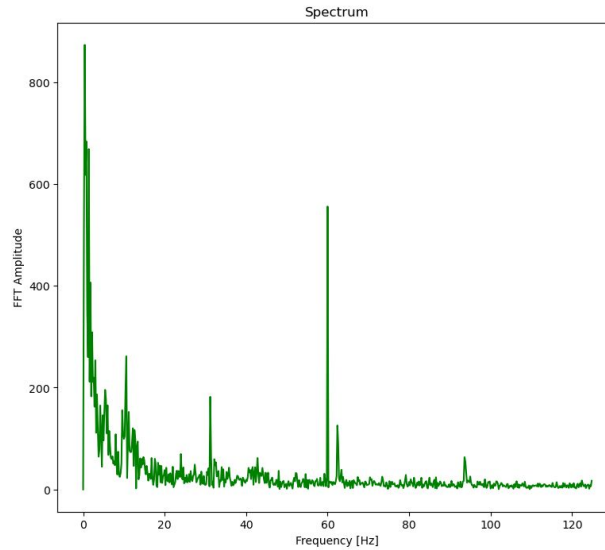t-SNE Visualization of Classes

a) Real samples
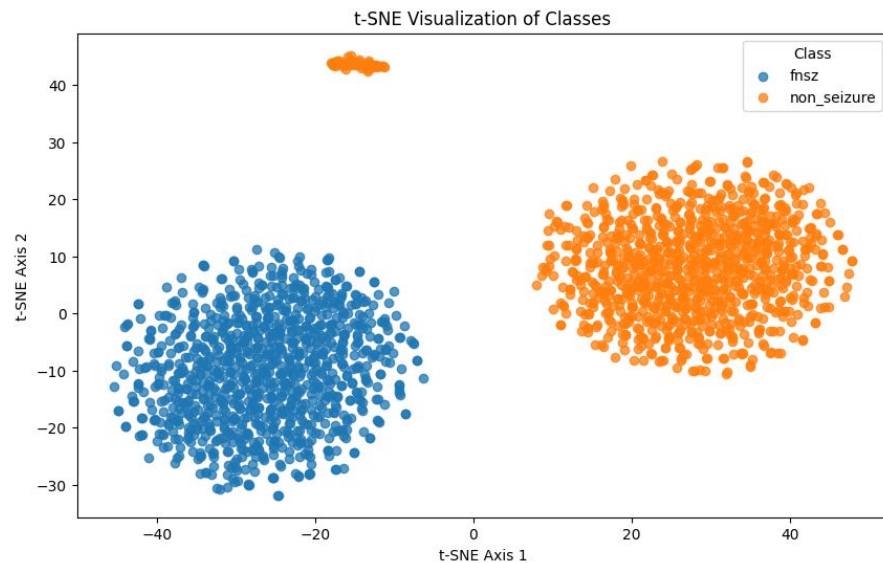
b) Synthetic samples

Frequency features

Appendix C: Channels and classes

Appendix D: Original signal and spectrum
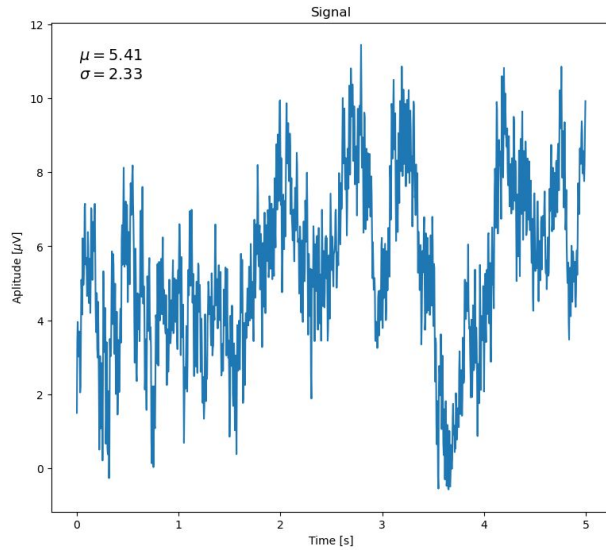
Appendix E: Choice of classifier

- Try out different classifier models
- Train on real data, test on real data
- Features:
    - FFT
    - Discrete wavelet transform
    - Mean, Standard deviations, Percentiles
    - Welch's PSD
- Models:
    - Gradient boosting
    - SVC
    - Multilayer perceptron

- Model chosen: FFT + SVC

Appendix F

logloss

— comb long 2 class Unet1D attn mlp - pred_x0 — long 2 class Unet1D time mlp - pred_x0
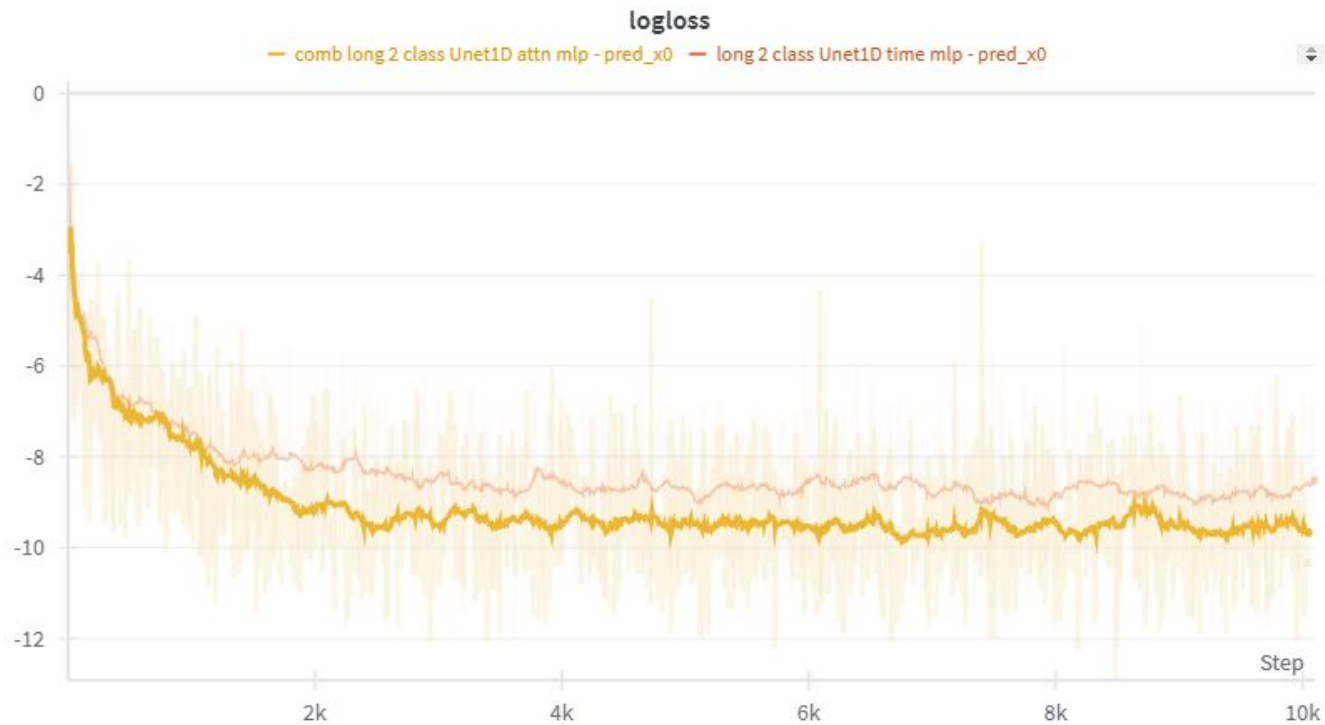
Appendix G: Dataset

- https://isip.piconepress.com/projects/tuh_eeg/html/downloads.shtml
- Access on request
- 26 846 clinical EEG recordings collected at Temple University Hospital (TUH) from 2002 - 2017
- EEGs stored as EDF files (EDF = European Data Format)
- Annotations stored as CSV (channels, start, stop, class [e.g. fnsz])