Epileptic Seizure Detection in Clinical EEGs Using an XGboost-based Method

Lan Wei
Catherine Mooney

future neuro
RCSI
UCD DUBLIN University College Dublin

Outline :

- Background
- Data
- Method
- Results
- Conclusions
- Future Work

Background

Background

- Epilepsy is one of the most common serious disorders of the brain, affecting about 50 million people worldwide.
- Electroencephalography (EEG) is an electrophysiological monitoring method which is used to measure tiny electrical changes of the brain, and it is frequently used to diagnose epilepsy.
- The visual annotation of EEG traces is time-consuming and typically requires experienced experts.
- Automatic seizure detection can help to reduce the time required to annotate EEGs.
- To date, studies on the automatic detection of seizures in clinical EEGs have been limited.

Data

Data

Temple University Hospital (TUH) EEG Corpus: A big data resource for automated EEG interpretation

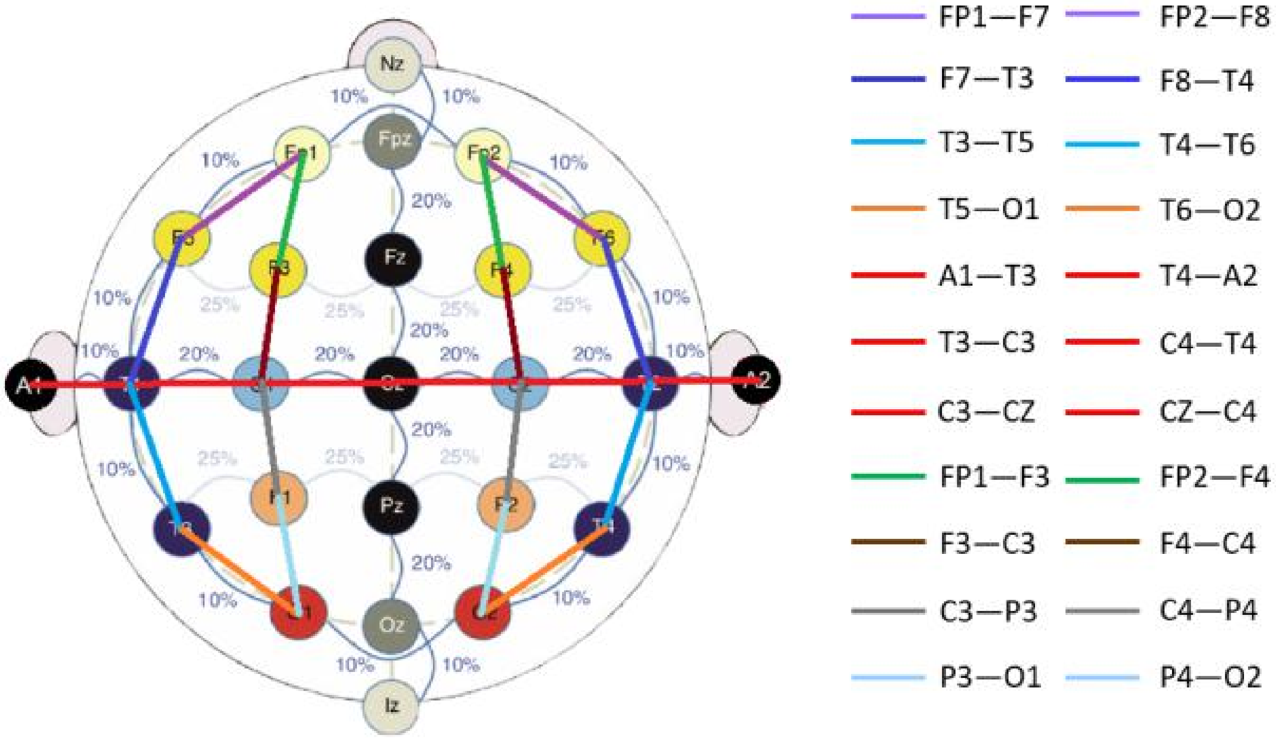
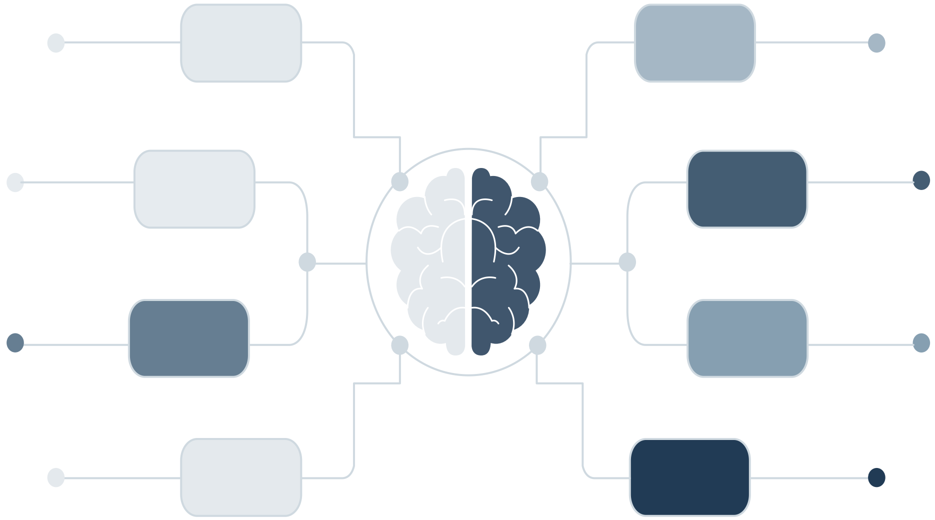


Table 1. Number an duration of seizures used to train, validate and test the method.

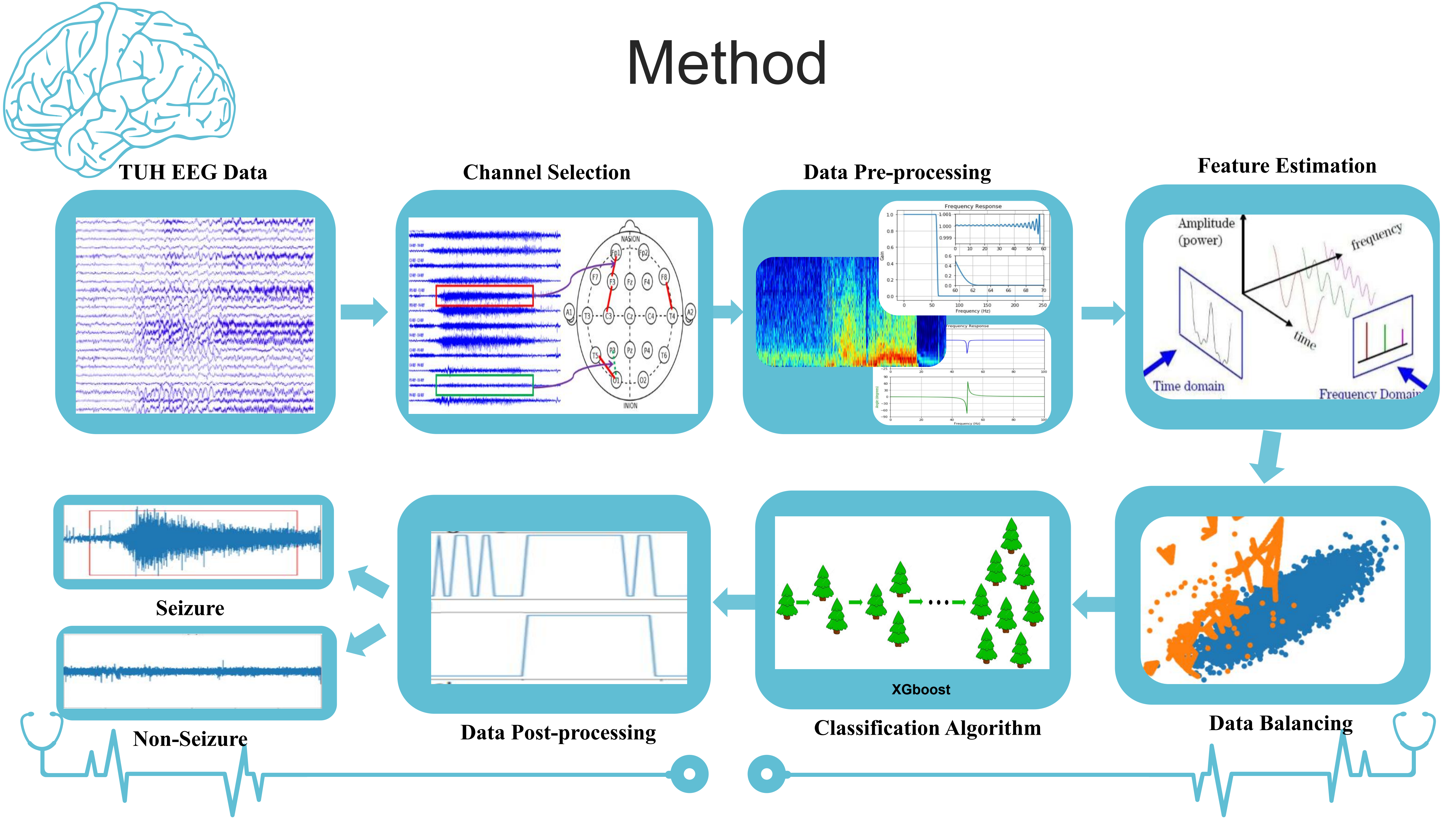
	Train	Validation	Test
Number of EEGs	4,597	1,013	1,026
Number of patients	592	50	Unknown
Number of seizures	2,370	673	Unknown
Seizure duration(s)	168,139.23	58,445.11	Unknown
Non-seizure duration(s)	2,540,144.77	554,786.89	Unknown

https://www.isip.piconepress.com/projects/tuh_eeg/

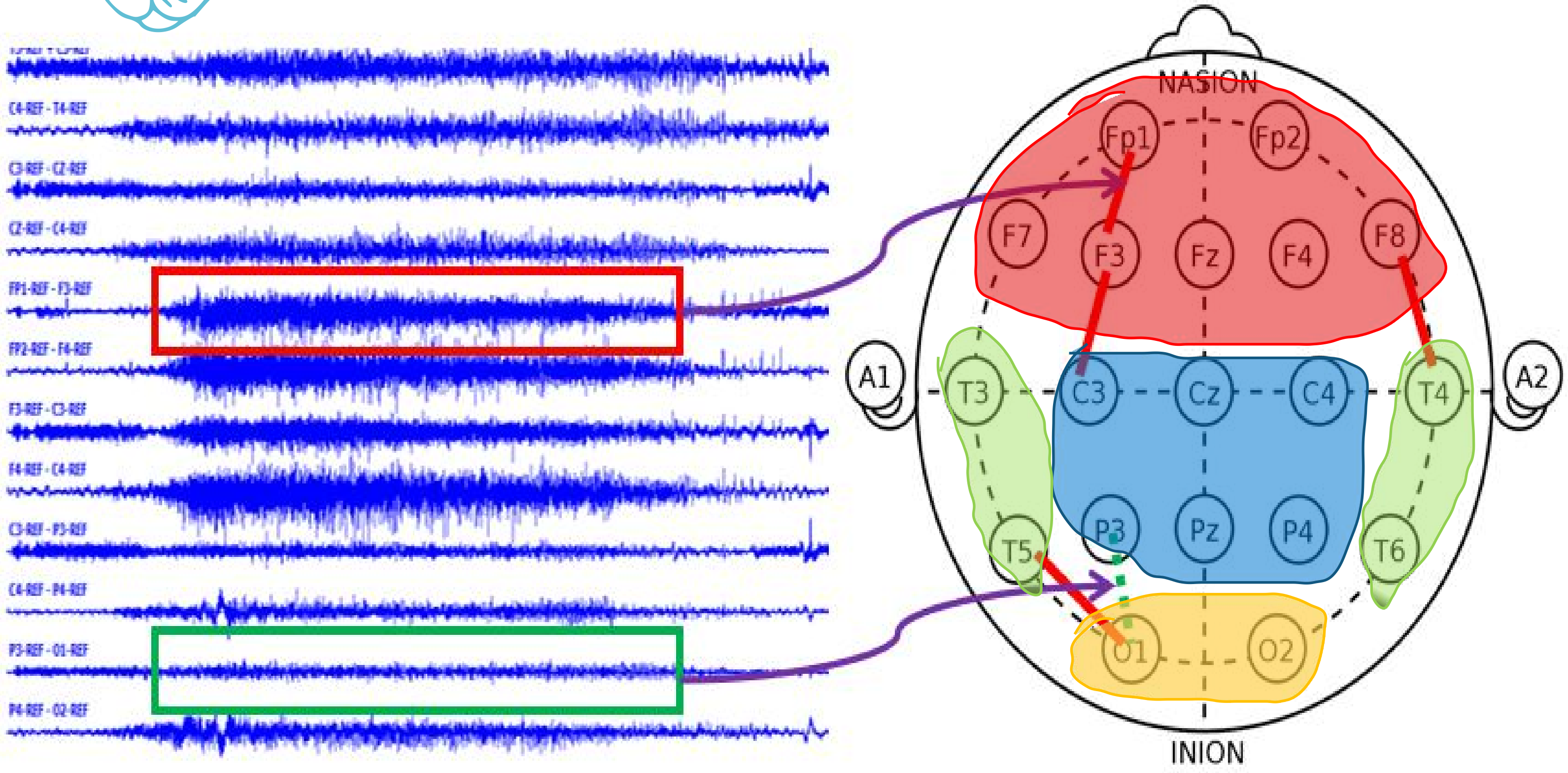


Method

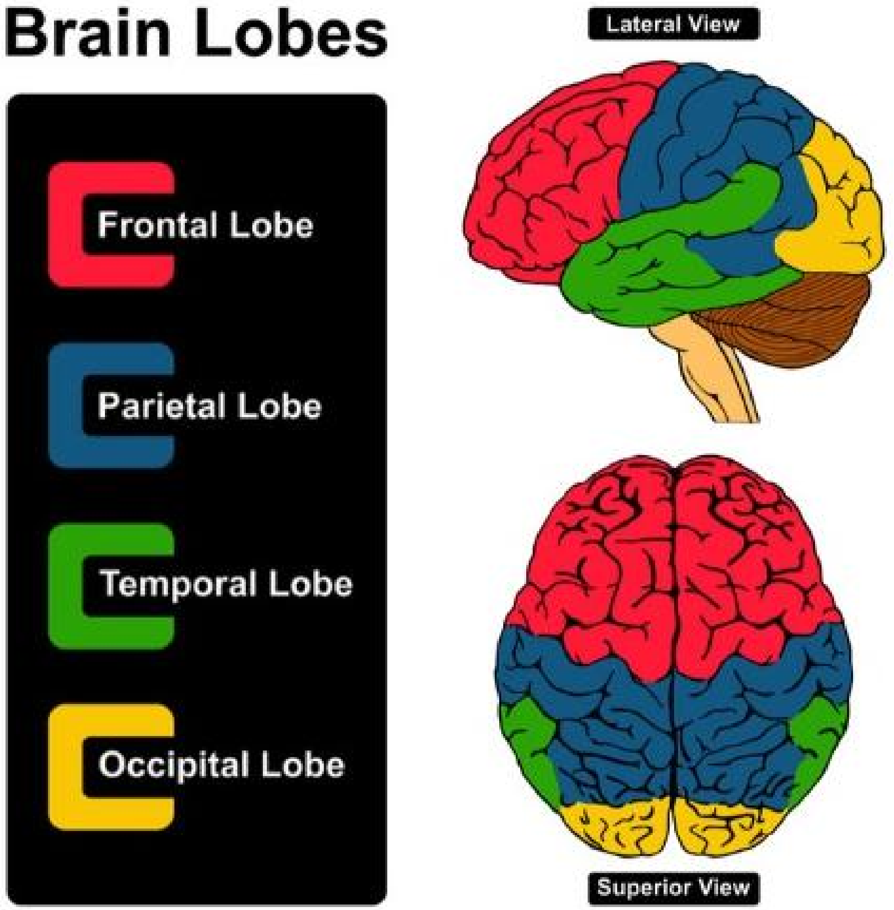
Method


TUH EEG Data
Channel Selection
NASION
INION
Data Pre-processing
Frequency Response
Feature Estimation
Amplitude (power)
frequency
time
Time domain
Frequency Domain
Data Balancing
XGboost
Classification Algorithm
Data Post-processing
Seizure
Non-Seizure

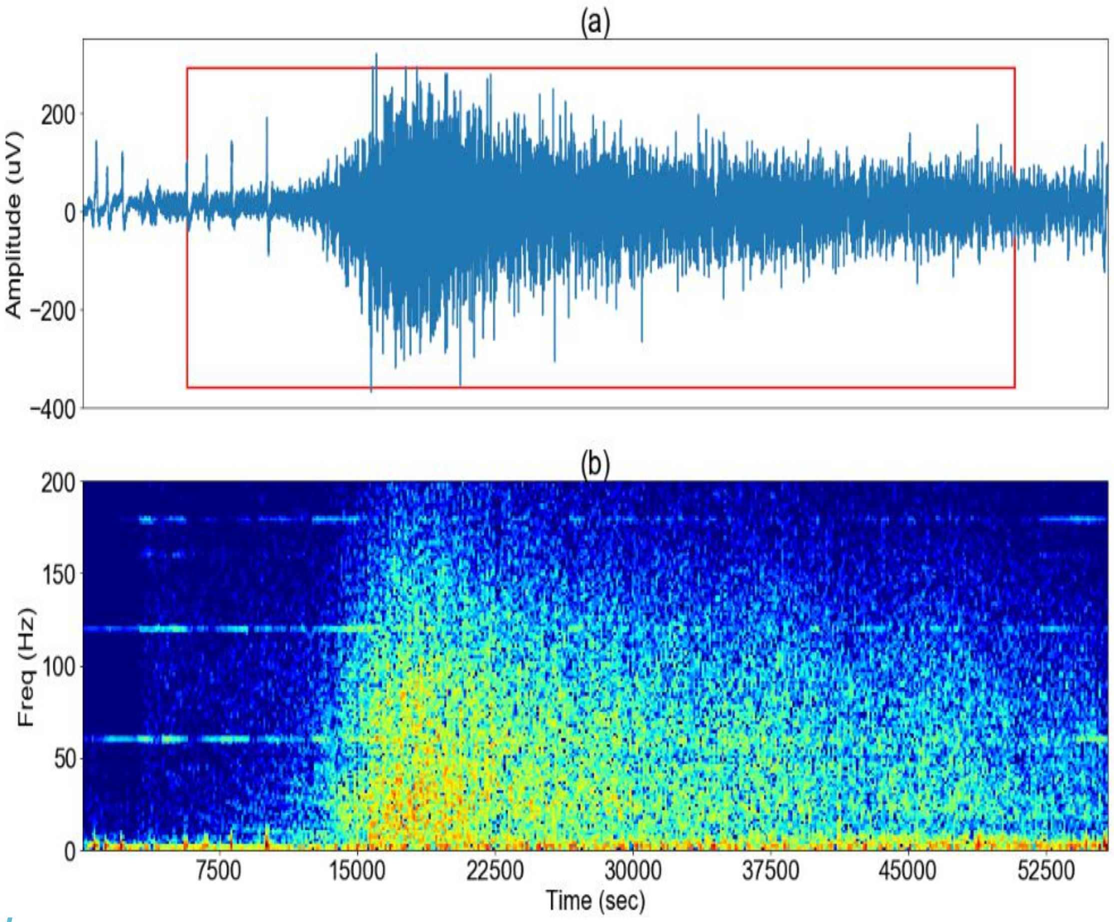
Channel Selection



Channel Fp1-F3
Channel F3-C3
Channel F8-T4
Channel T5-O1

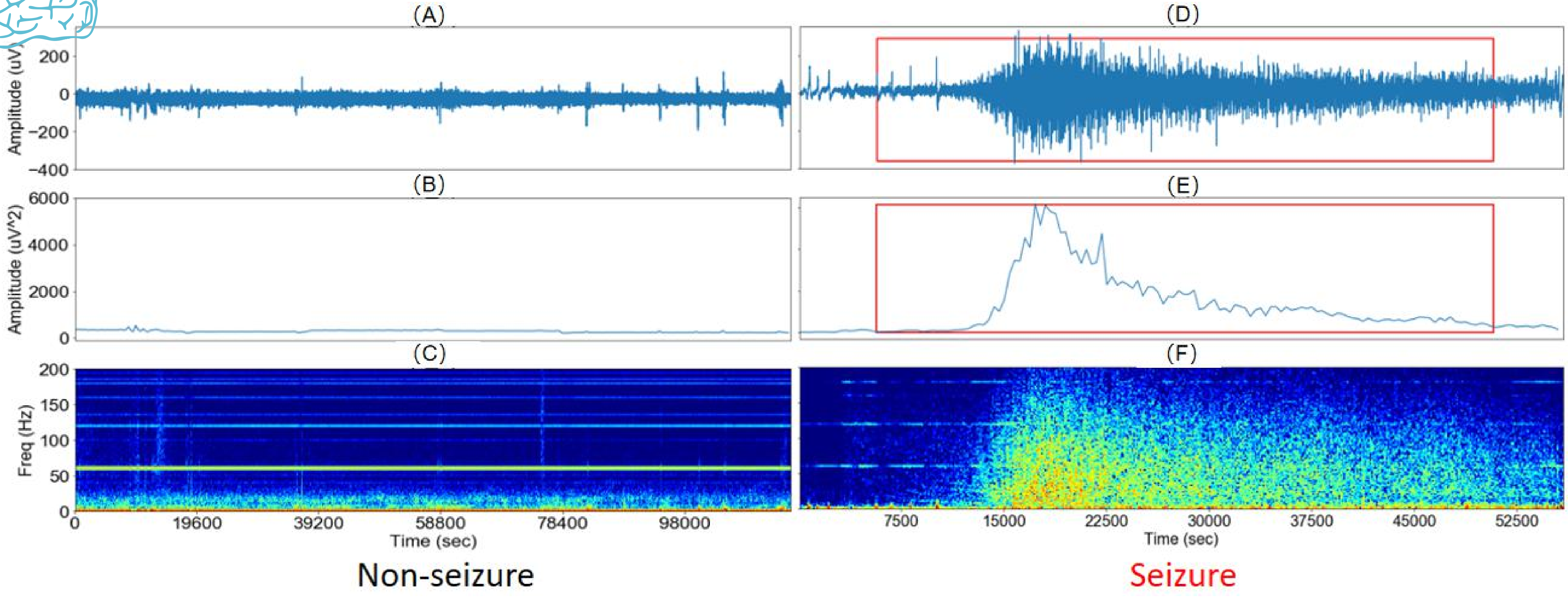


Data Pre-processing



- EEG recordings from the TUH-EEG dataset were sampled at different sample frequencies of 250Hz, 256Hz, 400Hz, and 1000Hz.
- Re-sample the signal to 250 Hz.
- A 60 Hz notch filter was applied to remove power line interference.
- DC offset was removed from the EEG recordings.
- The EEG signal was divided into epochs (1s with 0.5s overlap) with each epoch corresponding to seizure events or non-seizure events.

Feature Estimation



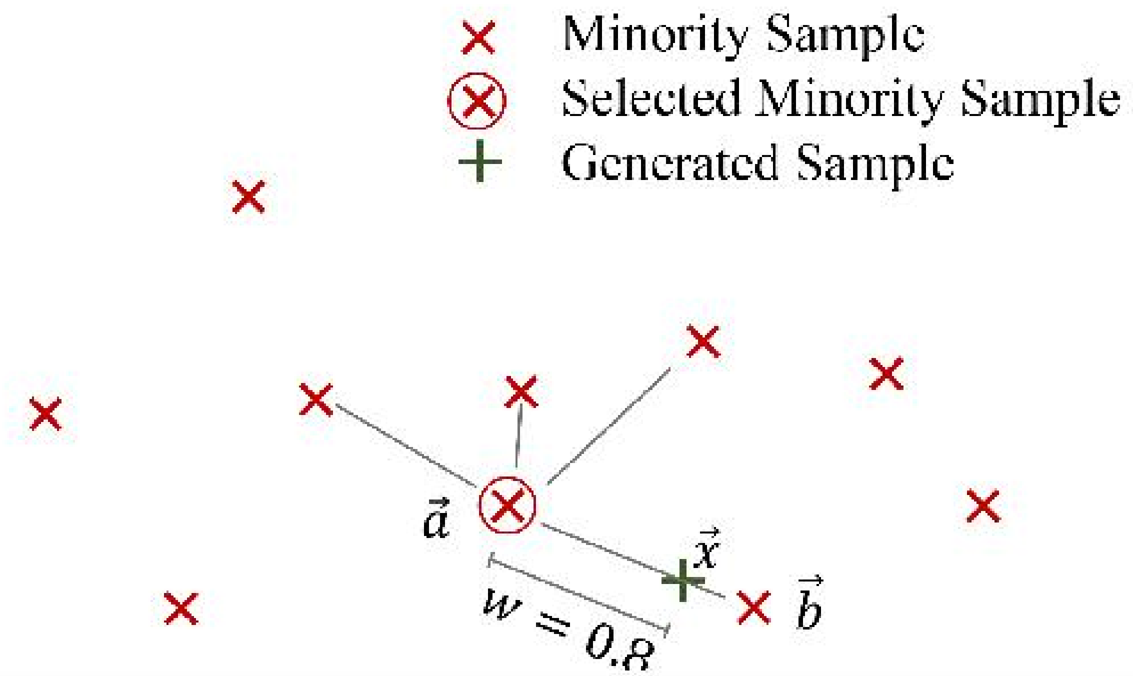
Standard features (9):

- TKEO
- Standard deviation
- Signal envelope
- etc.

Features in sub-frequency bands (7):

- The relative and absolute band power in
  - delta (0-4 Hz); theta (4-8 Hz); alpha (8-16 Hz); beta (16-32 Hz); gamma (32-64 Hz);
- Absolute band power of the EEG amplitude
- etc.

Data Balancing

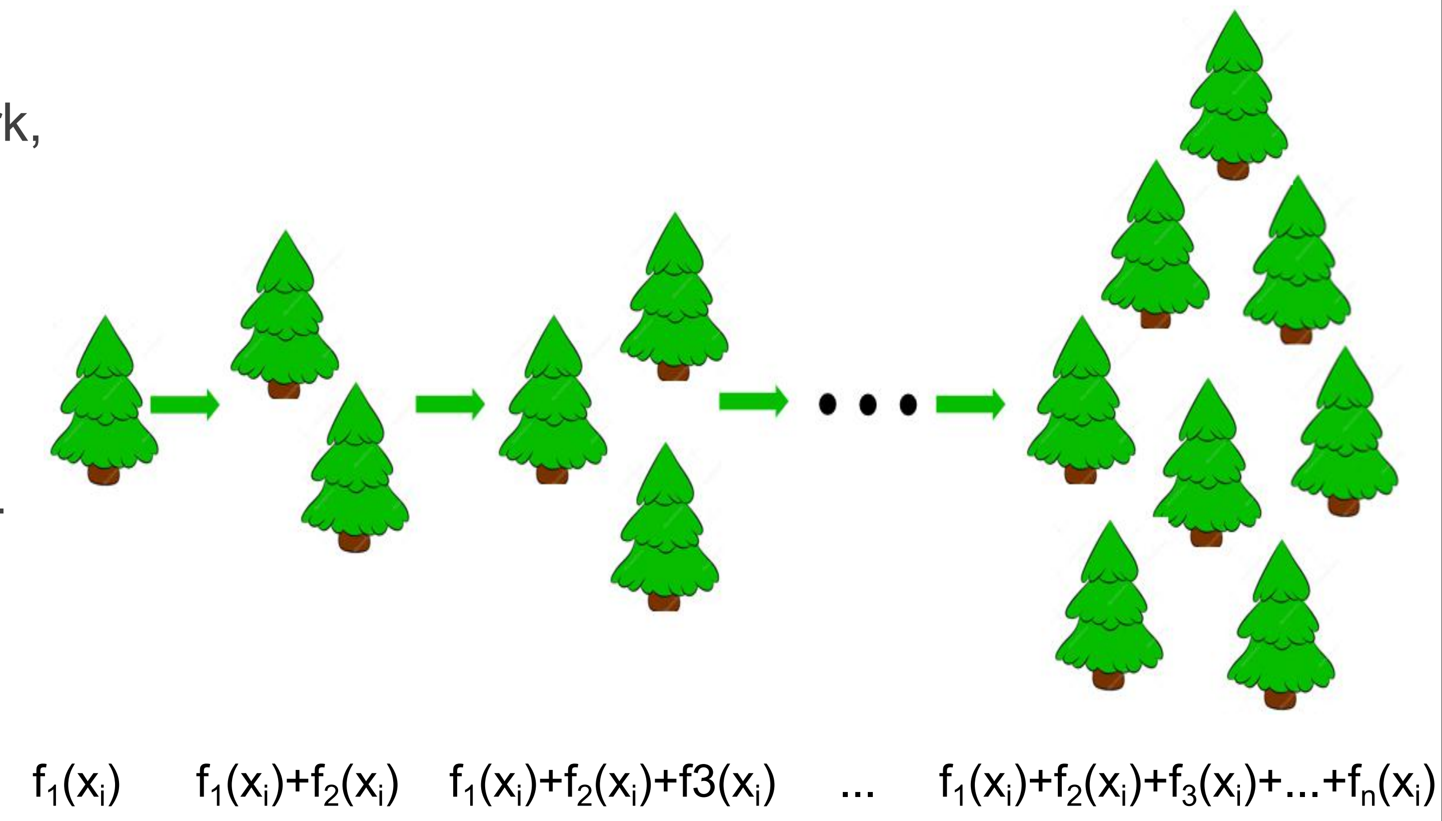


- Synthetic Minority Over-sampling Technique (SMOTE) was used to balance the data in the training set.
- SMOTE works by selecting examples that are close in the feature space, drawing a line between the examples in the feature space and drawing a new sample at a point along that line.
- It starts by randomly selecting an example from the minority class and then finds the nearest neighbour k to this example (k=5 in this study).

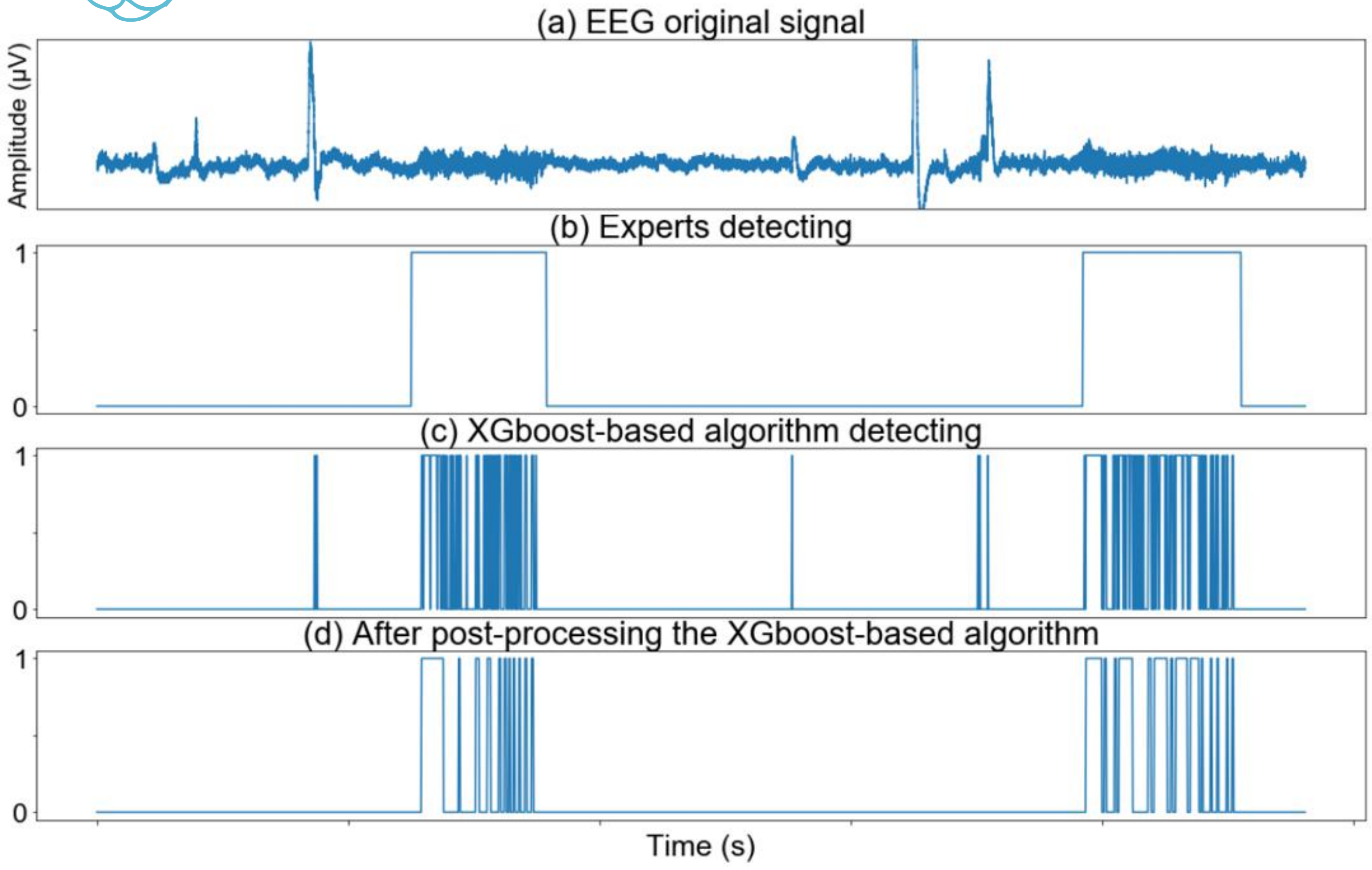
Classification Algorithms

- XGBoost is a decision-tree-based ensemble machine learning algorithm that uses a gradient boosting framework, which integrates many weak classifiers to form a robust classifier.

- XGBoost algorithm was implemented within the Python 3 environment, scikit-learn.
    - n-estimators = 100
    - gamma = 0.015
    - max-depth = 10
    - learning rate = 0.01



Data Post-processing



Figures are generated from:
00005625_s002_t001.edf Channel T5_O1

- The seizures detected by the XGBoost-based method with an interval less than 2s were grouped together, and their duration was extended from the start time of the first component to the end time of the last component.

- The duration of the seizure is typically greater than15s; therefore, if the duration of seizure detected by the XGBoost-based method was not greater than 15s, they were relabeled as a non-seizure. Otherwise, the seizure was defined as the final seizure event.

Results

Results

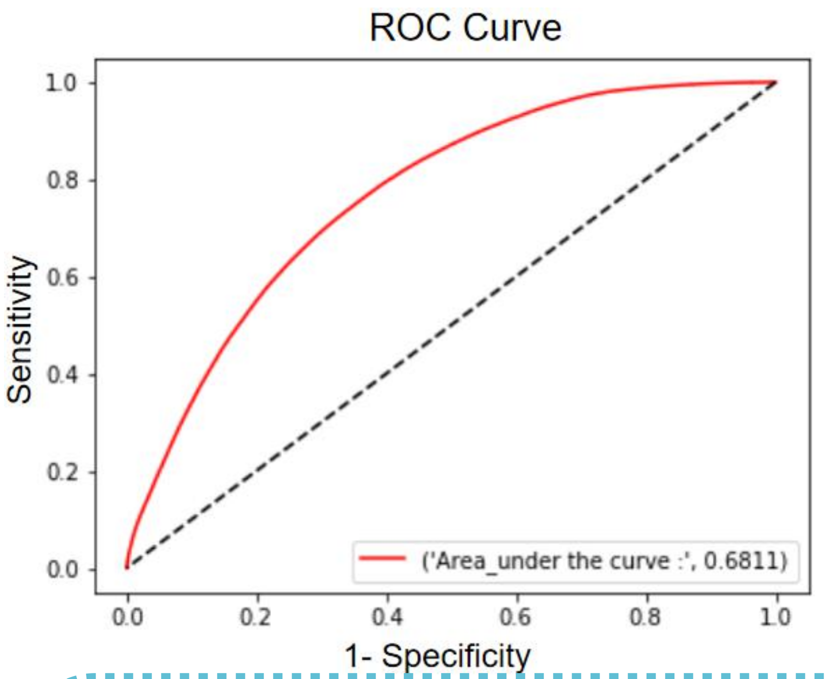
Performance evaluation:

$$\mathrm{Sens} = \frac{TP}{TP + FN}$$

$$\mathrm{Acc} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$F1 = \frac{TP}{TP + 1/2(FP + FN)}$$

True Positives (TP): Seizures event predicted as seizure event
False Positives (FP): Non-seizures event predicted as seizure event
True Negatives (TN): Non-seizure event predicted as non-seizure event
False Negatives (FN): Seizure event predicted as non-seizure event



AUROC: The area under the receiver operating characteristics curve

FA/24hr: False alarm rate per 24 hours

Table 1. Performance of the XGBoost-based seizure detection method.

	Sens(%)	Acc(%)	F1	AUROC	FA/24hr
Train (N=4,597)	59.80	67.01	0.5881	0.8030	55.69
Val (N=1,013)	51.90	58.85	0.4935	0.6811	81.35
Test (N=1,026)	20.00	-	-	-	15.59

Train: training set; Val: validation set; Test: testing set.

Results

Table 1. Comparison of the performance of our seizure detection method with other published work.

Method	Channel	Sens (%)	FA/24hr
2D CNN-L3 (Val)	22	39.15	22.83
2D CNN-L3 (Val)	20	34.54	49.25
2D CNN-L3 (Val)	16	36.54	53.99
2D CNN-L3 (Val)	8	33.44	38.19
2D CNN-L3 (Val)	4	33.11	325.54
2D CNN-L2 (Val)	8	30.66	28.57
2D CNN-L1 (Val)	4	34.09	332.15
2D CNN-L3 (Val)	2	31.15	308.74
HMM/Sda (Val)	22	17.29	82.00
HMM/LSTM (Val)	22	22.84	68.00
IPCA/LSTM (Val)	22	22.12	83.00
CNN/MLP (Val)	22	31.58	91.00
CNN/LSTM (Val)	22	12.48	8.00
XGBoost (Val)	4	51.90	81.35

Table 1. Comparison of the performance of other work in the Neureka 2020 Epilepsy Challenge.

Po	Team/Ind.	Sen (%)	FAs/24hr	Chan	Score
1	Biomed Irregulars	12.37	1.44	16	2.46
2	NeuroSyd	2.04	0.17	2	0.82
3	USTC-EEG	8.93	0.71	17	0.45
4	RocketShoes	5.98	3.36	3	-3.60
5	**Lan Wei (Ind.)**	**20.00**	**15.59**	**4**	**-20.56**
6	EEG Miners	16.00	16.54	9	-28.89
7	Anonymous (Ind.)	21.65	28.05	4	-50.05
8	James Msonda (Ind.)	11.33	29.27	10	-65.79
9	TABS	9.03	31.21	19	-76.50
10	cpl team	5.66	94.34	1	-230.59
11	DeepAlert	9.86	172.92	10	-426.40
12	Interfaces	26.53	186.63	1	-440.44
13	Neurocomputacion	0.22	758.48	11	-1,900.32
14	TeamPT2	34.75	927.12	19	-2,290.53
15	Last Dance	10.13	1,385.03	1	-3,452.83

Conclusions

Conclusions

- ✓ An XGBoost-based method to detect seizures in the world's largest publicly accessible archive of clinical EEG data set (TUH-EEG).
- ✓ The XGBoost-based method shows a sensitivity of 20.00% and FA/24hr of 15.29 by using four channels in the test set of TUH-EEG data
- ✓ Has the potential to assist researchers in the automated analysis of seizures in these long EEG recordings using fewer EEG channels.

Future Work

Future Work

- ☐ Analyse the difference between the features of the seizure period and the normal part of the EEG

- ☐ Explore features that can distinguish the seizure events and non-seizure events better

- ☐ Explore other algorithms to obtain a more accurate seizure detection method for TUH-EEG data

Thank You

Acknowledgments

This publication has emanated from research supported in part by a research grant from Science Foundation Ireland (SFI) under Grant Number 16/RC/3948 and co-funded under the European Regional Development Fund and by FutureNeuro industry partners. We acknowledge the Research IT HPC Service at University College Dublin for providing computational facilities and support that contributed to the research results reported in this paper.