Effects of Varying Time Windows of EEG Recordings on Accuracy of ML Identification of Abnormal EEG Signals

Olga Soldatenko & Benjamin Kaplan
097248: Machine Learning in Medicine, Final Project

Summary

While many researchers have attempted to use machine learning methods to identify abnormal EEG readings, there seems to be no consensus in the literature regarding what length of time the EEG signal should be to maximize identification accuracy. We compared accuracies obtained using three different machine learning architectures and a variety of time windows to identify abnormal EEG readings from the TUH Abnormal EEG Corpus.

Introduction

- Electroencephalograms (EEGs) are noninvasive measurements of electrical activity on the surface of the scalp which is representative of brain activity
- Abnormal EEG readings can be indicative of a variety of potential neurological disorders
- Machine learning (ML) is a promising avenue for allowing easier identification of abnormal EEG readings in clinical settings
- However, there appears to be no common consensus of how much time of each patient's recording should be used for analysis
- While some sources indicate as little as 1 minute of each readings can be used, other sources identify 11 minutes as the optimal time window, while others use up to 20 minutes

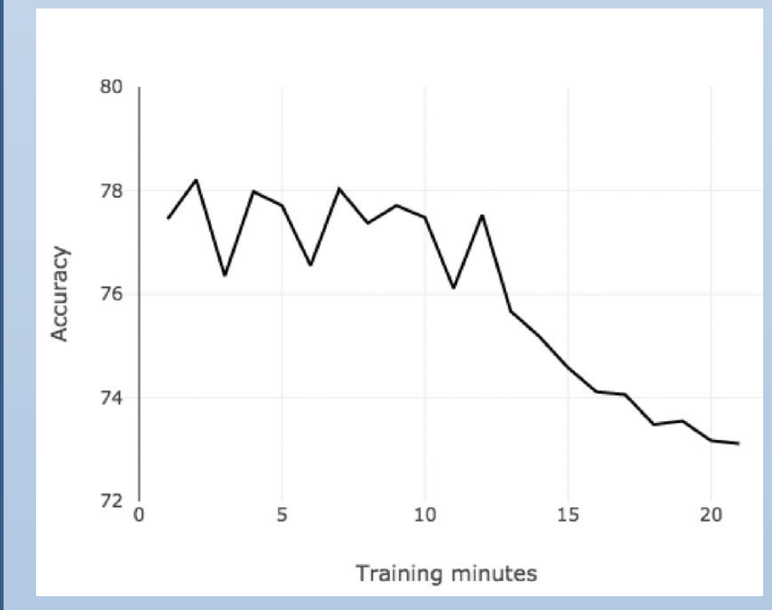

Figure 1. Figure from Roy, et al. showing their finding on a simplified dataset and sorting algorithm that 11 minutes is the maximum time window that can be used while maintaining accuracy

- Additionally, some researchers discarded the first 60 seconds of each EEG recording since they found high amounts of noise during that time, while others did not.

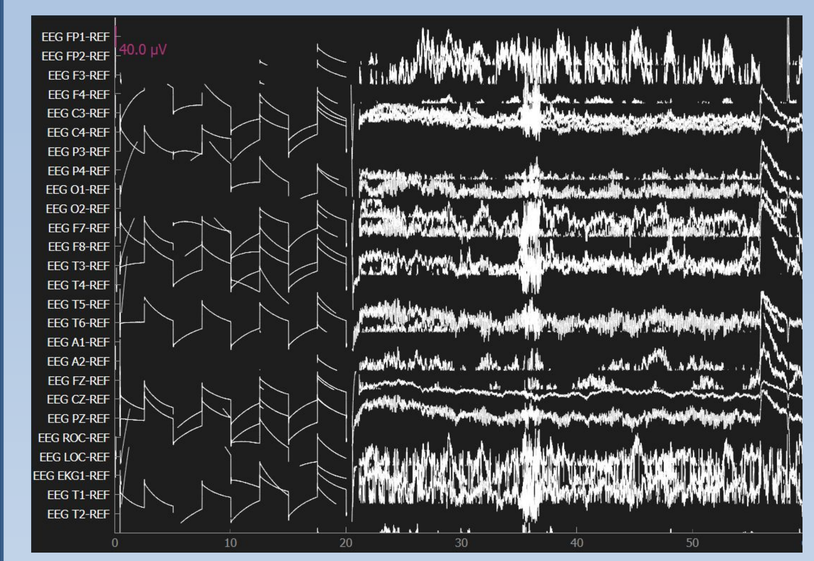

Figure 2. Example of noise observed during first 60 seconds of some recordings in the TUAB EEG Corpus

Methods

Description	Patients			Sessions		
	Normal	Abnormal	Total	Normal	Abnormal	Total
Train	1237	893	2130	1371	1346	2717
Evaluation	148	105	253	150	126	276
Total	1385	998	2383	1521	1472	2993

Figure 3. TUH abnormal EEG corpus by Temple University: data description

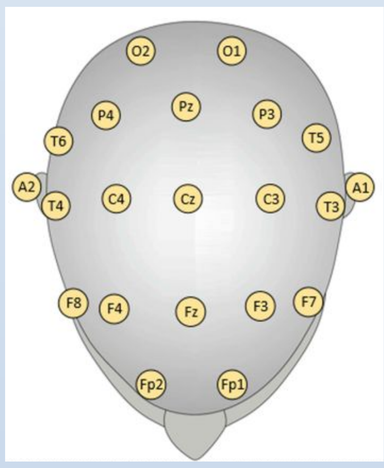

Figure 4. Electrodes' placement

Models used: BD-Deep4, BD-Shallow, EEGNet

Hyperparameter	BD-Deep4	BD-Shallow	EEGNet
Batch size	16	16	16
Optimizer	Adam	Adam	Adam
Max epochs	35	35	35
Number of channels	21	21	21
Learning rate	1 * 0.01	0.06 * 0.01	0.11 * 0.01
Weight decay	0.5 * 0.001	0	5.8 e-07
Dropout	0.5	0.5	0.25

Figure 4. Training hyperparameters

- Our goal was to estimate accuracies of these models on different time windows of the EEG recordings.
- We have examined the windows of length: 1, 5, 11 and 20 minutes, those that include the first minute of the recording and those that don't

Results

Time windows in min	Shallow	Deep	EEGNet
0+1: [0, 1]	78.5	79.8	76.3
1+1: [1, 2]	82.8	82.5	80.4
0+5: [0, 5]	81.5	82.3	80.8
1+5: [1, 6]	82.2	83.1	80.8
0+11: [0, 11]	82.2	82.8	80.2
1+11: [1, 12]	82.8	83.3	81.5
1+20: [1, 21]	X	82.4	X

Figure 5. Models' accuracies

Discussion

- While it does seem that 11 minutes still remains a more ideal time window than 20 minutes, researchers using BD-Deep and other advanced ML models should be unconcerned about significant losses in accuracy resulting from using larger time windows
- Based on our results, a time window of one minute should be enough time to obtain accuracies comparable to larger time windows
- Our results validate Gemein, et al.'s assertion that the noise in the first minute of the EEG readings causes lower overall sorting accuracies across all three architectures
- While this drop in accuracy is particularly notable when using a time window of one minute, dropping the first 60 seconds of each recording yields higher accuracies on average for larger time windows as well
- Enlarging the overall time window effects only a minor accuracy increase for the Shallow and Deep architectures, though the EEGNet seems to respond better to larger time windows
- While we did observe a slight drop in accuracy in the 20-minute time window as compared to the 11-minute window, it was not nearly as dramatic as what Roy, et al. observed
- The practice of dropping the first minute should be recommended as standard for anyone using the TUAB database