# Learning EEG: Identification of novel electroencephalogram classifications and variability of baseline features in a large clinical dataset

David O. Nahmias, Kimberly L. Kontson, and Eugene F. Civillico

***Abstract*— Despite its increasing use and public health importance, very little is known about the consistency and variability of baseline electroencephalogram (EEG) measurements in healthy individuals and in patient populations. This research aims to investigate population variability of EEG features and their ability to classify patients based on patient characteristics. We first propose a data-driven method of evaluating consistency and variability of commonly used EEG features. Using a Kruskal-Wallis test we find normalized features compared across different time intervals within a given recording to be consistent. Further, we find certain features to have better intra-subject consistency than others, notably spectral entropy, skew, and kurtosis. Using these results we apply machine learning techniques to classify patients based on sex, age, medications taken, and clinical impressions. We find statistically significant classifications for age, medications taken, and clinical impressions, achieving accuracies above 86%, 72%, and 74%, respectively.**

## I. Background and Motivation

MEDICAL devices that interface with the nervous system for diagnostic, therapeutic, or rehabilitative purposes are a major area of innovation in the medical products industry. Electroencephalography (EEG), which measures brain electrical signals from the scalp, is a common neuro-monitoring technique used in both clinical and research settings. EEG is typically non-invasive, relatively inexpensive, and has been shown to contain biomarkers for neurological disorders and for behavioral states, including specific movement intent. EEG signals can also be used for brain-machine interfaces to control external robotic devices, such as exoskeletons and prostheses. However, EEG has only recently begun to move into quantitative applications beyond visual inspection by trained professionals. Further, despite its increasing use and public health importance, very little is known about the consistency and variability of baseline EEG measurements in healthy individuals and in patient populations. Understanding population variability and individual variation in EEG could affect the safety and efficacy of therapeutic, diagnostic, and rehabilitative medical devices.

This work was funded by CDRH OSEL Division of Biomedical Physics and DARPA-FDA IAA 224-14-6009, DARPA Biological Technology Office through the Hand Proprioception and Touch Interfaces (HAPTIX) program under Dr. Doug Weber, Program Manager.

DN is with the Electrical and Computer Engineering Department at the University of Maryland-College Park (Corresponding author: David.Nahmias@fda.hhs.gov). DN and KK are with the US Food and Drug Administration, Center for Devices and Radiological Health, Office of Science and Engineering Laboratories, Division of Biomedical Physics. EC is with the Department of Health and Human Services.

## II. Methods

### A. Dataset

The Neural Engineering Data Consortium (NEDC) EEG Corpus from Temple University Health, version 0.6, has 10,535 subjects with 16,377 EEG recording sessions (~500GB) [1]. Each recording session is accompanied by a clinical note containing clinical impressions and patient characteristics. Further, 2,999 of the sessions are labeled by the NEDC as 'Normal', n=1,509, or 'Abnormal', n=1,490.

### B. Extracting patient information and common EEG

Patient information and demographics were obtained through textual analysis of the clinical notes, an example of which can be seen in Fig. 1. Through custom Python code, the following patient information was extracted:

- Sex: Male (n=4,673), Female (n=5,039)
- Age: Mean=47.97, SD=22.08, IQR=32 (n=10,233)
- Medication Lists (e.g. Dilantin, Keppra)
- Neurological Disorders (e.g. Epilepsy, Seizures)
- Expanded 'Normal' (n=4,164), 'Abnormal' (n=9,912)

MEDICATIONS: Metoprolol, others.
CLINICAL HISTORY: A 60-year-old male with bilateral hand tremor. Patient reports that the tremor waxes and wanes. He had difficulty providing the history.
INTRODUCTION: Digital video EEG is performed in the lab using standard 10-20 system of electrode placement with one channel of EKG. Hyperventilation and photic stimulation were completed.
TECHNICAL: No significant issues.
DESCRIPTION OF THE RECORD: In wakefulness, the background includes a well organized pattern with a 9 Hz alpha rhythm and a small amount of low voltage, frontocentral beta. Drowsiness is characterized by slow rolling eye movements why deeper stages of sleep were not sustained. Photic stimulation did not significantly activate the record. Hyperventilation did not activate the record significantly. Heart rate 72 beats per minute.
IMPRESSION: Normal EEG in wakefulness and brief drowsiness.
CLINICAL CORRELATION: No focal or epileptiform features were identified in this EEG. A normal EEG does not exclude a diagnosis of epilepsy. If seizures are an important consideration, a follow up tracing capture in deeper stages of sleep is suggested.

Fig. 1. Sample clinical report for single EEG recording session.

Further, since the data were acquired across a variety of systems, a standard 19 channel 10-20 layout was used for all subjects. The subset of channels was: Fp1, Fp2, F3, F4, C3, C4, P3, P4, O1, O2, F7, F8, T3, T4. T5, T6, Fz, Cz, Pz.

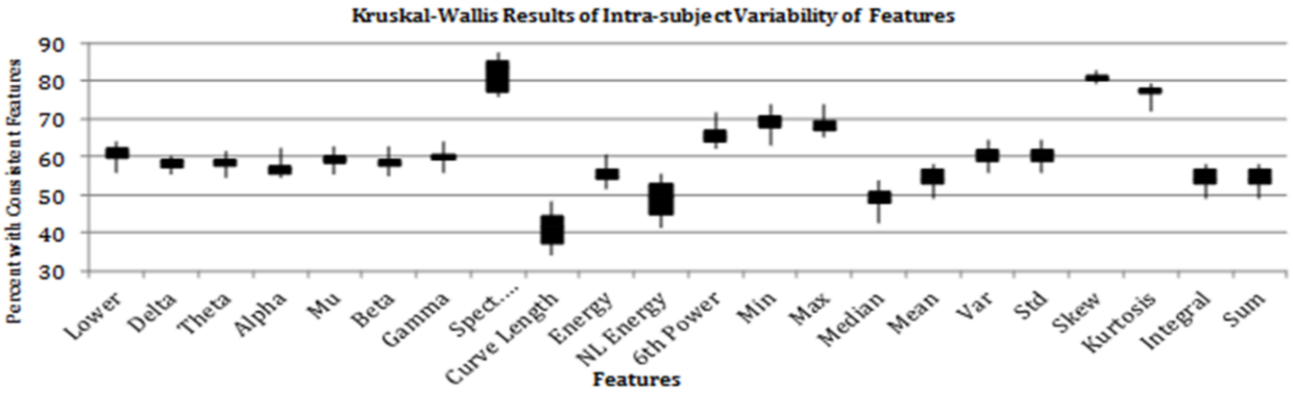


Fig. 2. Results of intra-subject Kruskal-Wallis test showing percent of intra-subject tests that failed to reject the null hypothesis, $p > 0.05$. Notably, spectral entropy, skew, and kurtosis have high intra-subject consistency (n=2,257).

### C. *Preprocessing raw EEG*

From each session, the first 16 minutes of an EEG recording across the same 19 channels was used. Sessions without at least one recording with 16 minutes were ignored. If necessary, recordings were down-sampled to 250Hz. The data were then filtered using a forward-backward 6$^{th}$ order Butterworth bandpass filter from 1Hz to 50Hz, and re-referenced using a common average.

### D. *EEG feature validation and variability analysis*

Frequency and time domain features were computed and analyzed. The frequency domain features analyzed were Power Spectrum Density (PSD) band ratios (<1Hz [lower], 1-4Hz [Delta], 4-8Hz [Theta], 8-12Hz [Alpha], 12-16Hz [Mu], 16-25Hz [Beta], 25-40Hz [Gamma], where the band power of each range is divided by the total power, 0-40Hz), and spectral entropy. The time domain features analyzed were curve length, energy, non-linear energy, sixth power, minimum value, maximum value, median, mean, variance, standard deviation, skew, kurtosis, integral, summation [2]. Features were computed on varying lengths of time, from 30 seconds to 4 minutes, on the same EEG recording. Stationarity of these features was determined by applying a non-parametric Kruskal-Wallis (K-W) test. Further, intra-subject variability of these features was computed by applying a Kruskal-Wallis test on features from the same subject across different sessions. Finally, for the same time lengths, Intraclass Correlation Coefficients (ICC) were computed for PSD band ratios [3].

### E. *EEG Classification*

Using a kernel Support Vector Machine (k-SVM), with 418 (22 features x 19 channels) possible features, we attempt to classify sex, age groups, medication taken, and 'normal' and 'abnormal' impressions, the labels of which were extracted through text analysis of the clinical notes.

## III. Results

For normalized features (those with a factor 1/N where N is the number of samples) compared across different time intervals within a given recording, the K-W test indicates greater stationarity compared to unnormalized features. (n=13,970). Assessment of intra-subject variability suggests spectral entropy, skew, and kurtosis features are highly consistent (Fig. 2). Lastly, analysis of the ICC showed that Delta, Theta, Alpha, and Beta band ratios had higher agreement than band ratios outside this range. These initial results were used to guide feature selection for classification. Cross-validation was used to train the SVM on different kernels with varying hyperparameters on 90% of the total samples available. Selected binary classification results, using no more than 100 features, on 10% untrained testing data are shown in Table I.

TABLE I
Classification Results

| Classes | Total Samples | Test Accuracy |
|---|---|---|
| Male vs. Female | n=12,222 | ~50.00% |
| Age < 20 vs. others | n=754 | 77.03% |
| 20 < Age < 50 vs. others | n=9,220 | 64.75% |
| Age > 50 vs. others | n=8,422 | 67.46% |
| Age < 10 vs. Age > 60 | n=738 | 86.11% |
| Taking Dilantin vs. none | n=982 | 66.32% |
| Taking Keppra vs. none | n=802 | 72.50% |
| NEDC 'Normal' vs. 'Abnormal' | n=2,920 | 74.35% |
| All 'Normal' vs. 'Abnormal | n=7,681 | 71.62% |

## IV. Conclusion

This research proposes novel data-driven metrics for evaluating consistency of EEG features. Further, significant classifications are found on certain patient characteristics and clinical EEG impressions using a limited number of features. Future work may include testing additional features and classifications.

## References

[1] A. Harati, S. Lopez, I. Obeid, M. Jacobs, S. Tobochink, and J. Picone. "The TUH EEG CORPUS: A Big Data Resource for Automated EEG Interpretation." *Proceedings of the IEEE Signal Processing in Medicine and Biology Symposium*, pp. 1-5, 2014.

[2] M. D'Allessandro, R. Esteller, G. Vachtesevanos, A. Hinson, J. Echauz, and B. Litt. "Epileptic Seizure Prediction Using Hybrid Feature Selection Over Multiple Intracranial EEG Electrode Contacts: A Report on Four Patients." *IEEE Trans on Biomedical Engineering*, 50(5), 2003.

[3] S. Gudmundsson, T. Runarsson, S. Sigurdsson, G. Eiriksdottir, and K. Johnsen. "Reliability of Quantitative EEG Features." *Clinical Neurophysiology*, 118, pp. 2162-2171, 2007.