

## Original Article

# Quantitative and Qualitative Analyses of Invasive EEG for Epileptic Seizure Prediction

Hussein R<sup>1,2\*</sup> and Ward R<sup>1</sup><sup>1</sup>University of British Columbia, Electrical and Computer Engineering, Canada<sup>2</sup>Stanford University, Center for Artificial Intelligence in Medicine & Imaging, USA

**\*Corresponding author:** Ramy Hussein, Electrical and Computer Engineering, University of British Columbia, Vancouver, BC V6T 1Z4, Canada; Center for Artificial Intelligence in Medicine & Imaging, Stanford University, Stanford, CA 94305, USA

**Received:** February 09, 2021; **Accepted:** March 06, 2021; **Published:** March 13, 2021

**Abstract**

This study aims at explaining why existing machine learning methods have achieved limited performance when applied to the problems of seizure prediction using human invasive EEG (iEEG) data. We provide quantitative and qualitative analyses of iEEG data, as this data has commonly been used for seizure prediction tasks. Analyzing and understanding the iEEG signals provide insights into the characteristics of the preictal and interictal brain signals (i.e., the signals preceding and between epileptic seizure attacks). Experimental results show that: (1) the iEEG data varies from one patient to another. Therefore, iEEG features collected from a patient over a specific time period may not achieve reliable seizure prediction performance for other patient neither for the same patient at some future times, (2) for iEEG dimensionality reduction, the exclusion of any individual iEEG channel may result in losing spatial information needed for accurate seizure prediction, as the different iEEG channels contain complementary information. Also, Principal Component Analysis is found to be efficient for iEEG dimensionality reduction in seizure prediction tasks, and (3) For a particular patient, the distribution of the preictal and of the interictal iEEG data vary over time, and thus negatively affect the predictive ability of pre-trained seizure prediction methods.

**Keywords:** Seizure prediction; EEG; PCA; Channel selection; Data mismatch

**Introduction**

Epilepsy is a neurological disorder that affects around 70 million people worldwide [1]. It is characterized by recurrent seizures that strike without warning. Symptoms may range from a brief suspension of awareness to violent convulsions and sometimes loss of consciousness [2]. Currently, anti-epileptic drugs are given to epileptic patients in sufficiently high dosages. These drugs could result in undesirable side effects such as tiredness, stomach discomfort, dizziness, and also blurred vision. Further, the patient's quality of life is severely affected by the anxiety associated with the unpredictable nature of seizures and the consequences therefrom. This motivated researchers to develop automatic seizure prediction systems [3]. The ability to predict seizures with high accuracy could make individualized epilepsy treatment possible (e.g., tailored therapies with fewer side effects). By having warnings of impending seizures, the patients can take their precautions and avoid any probable injuries. This vision inspired the proposed research.

Even though epileptic seizures seem unpredictable and often occur without warning, recent investigations have demonstrated that seizures do not strike at random [4,5]. Most existing seizure prediction methods, however, achieve limited performance [6-11]. Only patient-specific solutions that are tailored to individual patients have given good results [12,13], however, these methods have poor ability to adapt to new unseen data gathered from other patients. Intracranial Electroencephalogram (iEEG) is a common tool used for seizure prediction. The iEEG data that directly precede seizures is analyzed to identify the biomarkers that indicate upcoming seizures.

In 2013, a big dataset of long-term iEEG recordings has been recorded (for 374-559 days) from three patients with drug-resistant epilepsy [8]. A subset of this dataset was made publicly available and used in the Melbourne University Seizure Prediction Competition organized in November 2016 on Kaggle.com. In [12], Kuhlmann et al. describe the human iEEG dataset used in this contest and the results achieved by the top eight seizure prediction solutions. Most of these solutions, however, are patient-specific and have a lower chance of being able to generalize beyond the statistical patterns of the training examples.

Neuroscientists have found that the temporal dynamics of the brain activity of epileptic patients can often be categorized into four different states: preictal (directly prior to a seizure), ictal (during a seizure), postictal (directly after a seizure), and interictal (between two consecutive seizures). Accurate seizure prediction solutions necessitate that prediction methods are able to differentiate between the preictal and interictal brain activities with high levels of accuracy. In this work, we provide extensive analyses of the human iEEG data preceding seizures (i.e., preictal) and between consecutive seizures (i.e., interictal). These analyses present some insights into the brain data and help us to better understand the unpredictable nature of epileptic seizures that we have been attempting to quantify.

As the success of seizure prediction solutions rely on their ability to differentiate between the interictal and preictal brain states, we first investigate whether interictal and preictal iEEGs are statistically different. We also study dimensionality reduction of iEEG data, as the huge amount of iEEG data involved in seizure studies could hamper the practical applicability of feature engineering and classification



Figure 1: CT scan of the NeuroVista seizure advisory system implanted in a patient [8].

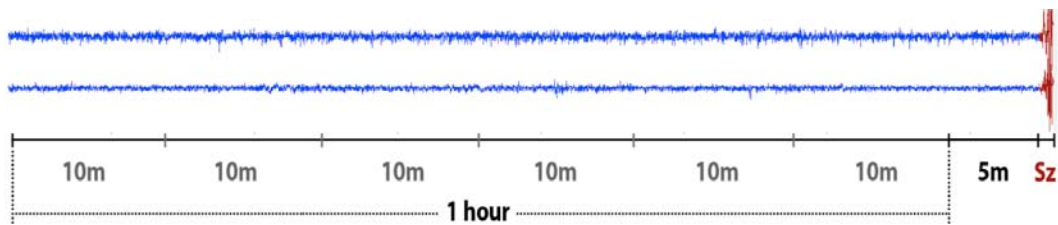


Figure 2: Examples of one-hour preictal iEEG signals with a 5-minute offset before seizures. Sz stands for the seizure onset.

procedures. We study whether the popular Principal Component Analysis (PCA) could be effectively used for reducing the iEEG dimensional space, and whether excluding some iEEG channels (i.e., sensors) could result in reliable dimensionality reduction. Finally, we examine the cross-correlations between the different iEEG sensor readings during the preictal and interictal brain states and how could this knowledge be used. Below we first describe the human iEEG data and the subjects employed in this paper.

## Subjects and Data

The iEEG data used is from the 2016 Kaggle seizure prediction competition and is described in [12]. The data were recorded constantly from humans suffering from refractory (drug-resistant) focal epilepsy using the NeuroVista Seizure Advisory System (described in [8]). Sixteen electrodes (4×4 contact strips) were implanted in all patients, directed to the presumed seizure focus, and connected to a telemetry unit embedded in the subclavicular area. Data were sampled at 400Hz, digitized using a 16-bit analog-to-digital converter, wirelessly transmitted to an external hand-held advisory device, and continuously stored in a removable flash drive. (Figure 1) shows an example of a Computerized Tomography (CT) scan of the head of one of the patients and reveals the locations of the 16 iEEG electrodes on the cortical surface of the brain.

Three drug-resistant patients who also had the least seizure prediction performance in [8] were chosen for our study. The selection of these patients was propelled by the intention to understand why 40% of the epileptic patients do not respond to medications. The large number of seizures recorded per patient (~380) gives us the chance to explore the common signatures or biomarkers within the iEEG recordings preceding seizures. The three patients were females and they all had resective surgery (i.e., removal of a small portion of the brain) before the trials. The first, second, and third patients were 22,

51 and 53 years old at the time of the clinical trial (i.e., in time of iEEG acquisition), but were diagnosed with epilepsy at the age of 16, 10 and 15, respectively.

As stated above, the key challenge in seizure prediction is to differentiate between the preictal and interictal brain states in people with epilepsy. In our study and also in previous seizure prediction work, only the lead seizures of every patient were used. Lead seizures are those that occur at least 4 hours after a previous seizure. The captured iEEG data were labeled and separated into training and testing sets. The testing data clips were recorded a long time (a few months) after recording the training data clips. As shown in (Figure 2), the preictal data clips were extracted from the 60 minutes prior to the lead seizures, with 5 minutes offset from the Seizure onset (Sz). Interictal clips were segmented from 60 minutes iEEG recordings that started at an arbitrarily time that was 4 hours after any seizure and ended at least 3 hours before the consecutive seizure.

## Are Interictal and Preictal iEEGs Statistically Different?

Most seizure prediction problems focus on the difference between the interictal and preictal brain states. Then the following question arises: Are the interictal and preictal iEEGs statistically different? If yes, then can we use their statistical features to distinguish between them? In an effort to address this question, we use boxplot—a popular descriptive statistics tool—to present the interictal and preictal iEEGs distributions across all channels. The boxplot presents the data distribution using the five-number summary: minimum, first quartile, median, third quartile, and maximum. The first quartile is the median of the data whose values are less than the median and the third quartile is the median of the data whose values are above the median. The difference between the maximum and the minimum numbers is referred to as “overall range” and the difference

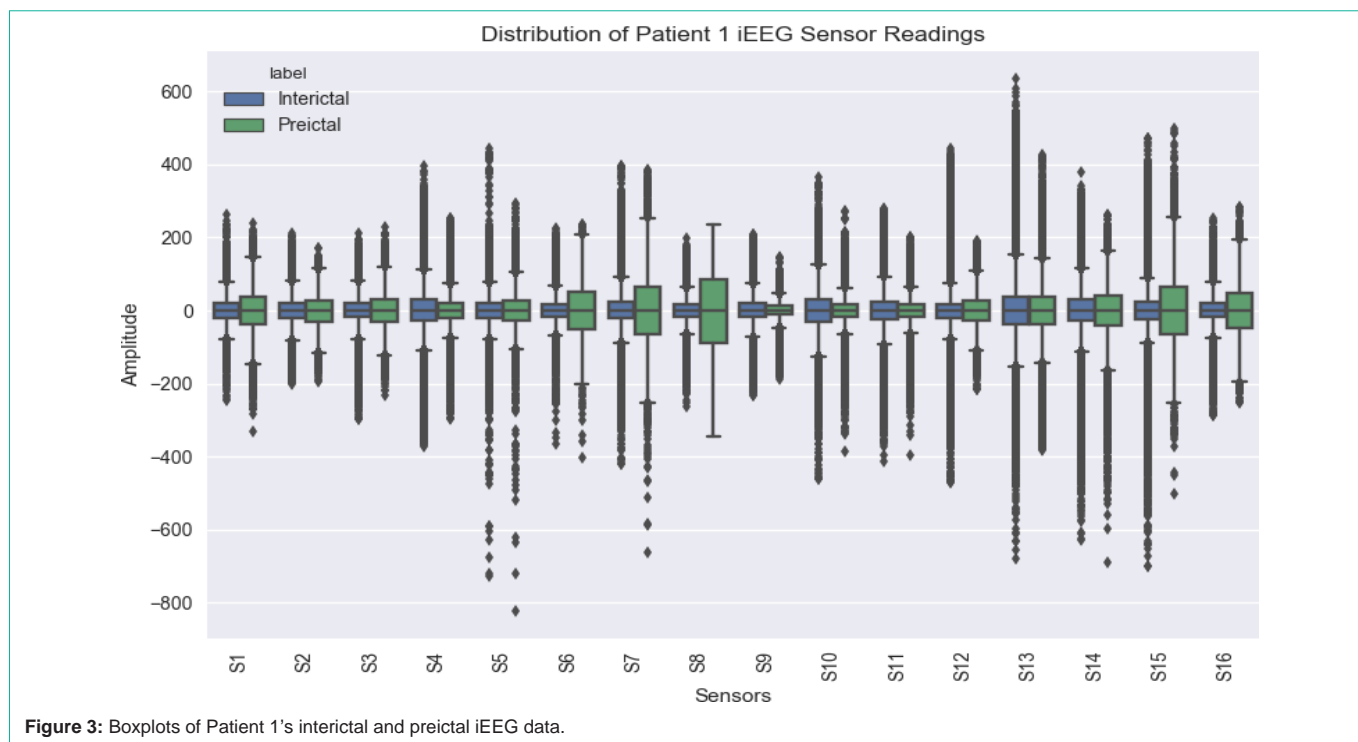


Figure 3: Boxplots of Patient 1's interictal and preictal iEEG data.

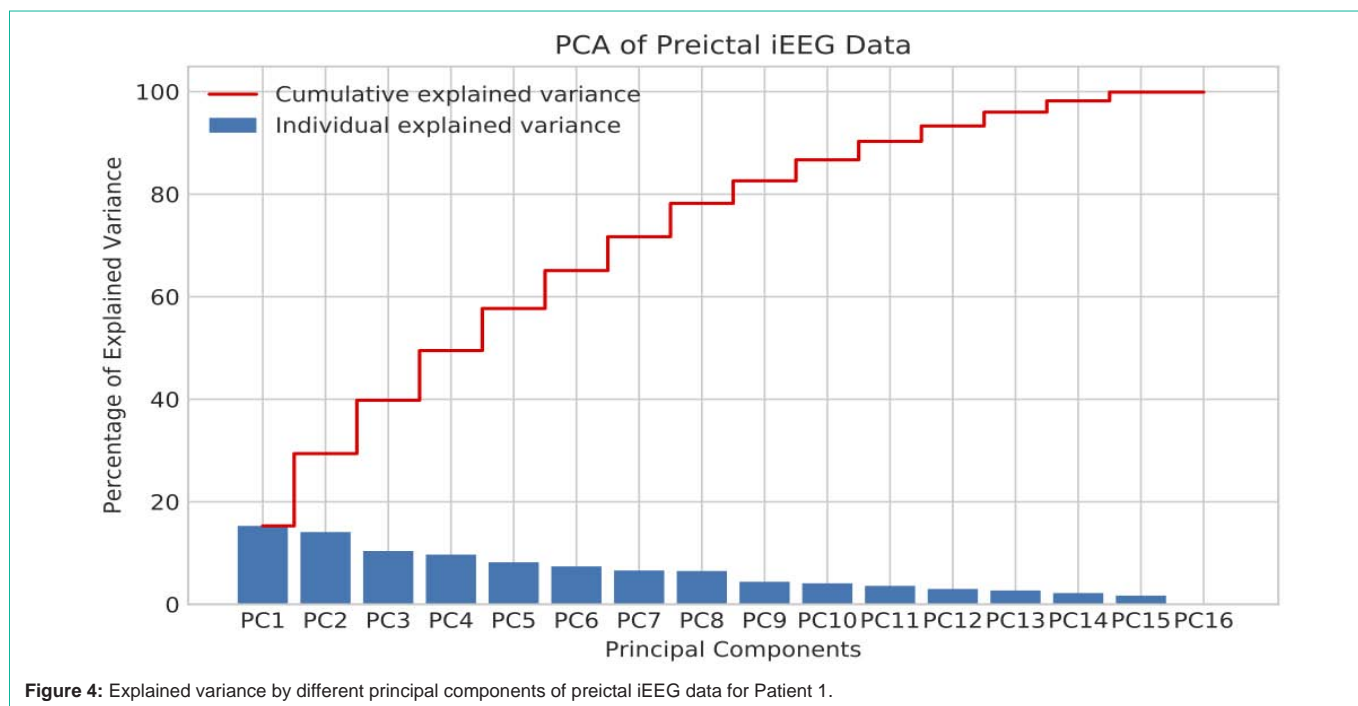


Figure 4: Explained variance by different principal components of preictal iEEG data for Patient 1.

between the first quartile and the third quartile is referred to as the “interquartile range, IQR”.

Figure 3 shows a comparative boxplot for Patient 1’s interictal and preictal iEEG sensor readings. For each of the 16 sensors, the interictal data boxplot (in blue) is on the left of the preictal data boxplot (in green). It can be noticed that, for most of the iEEG channels, the overall range (represented by the vertical distance between the

maximum and the minimum values) and Interquartile Range (IQR: represented by the vertical length of the box) of the preictal data are much greater than those of the interictal data. For the second patient, however, the overall range and IQR of the preictal data are smaller than those of the interictal data. For both patients, we also observe that both interictal and preictal iEEG data series include large outliers (an outlier is a point which falls more than 1.5 times the interquartile range below the first quartile or above the third quartile).

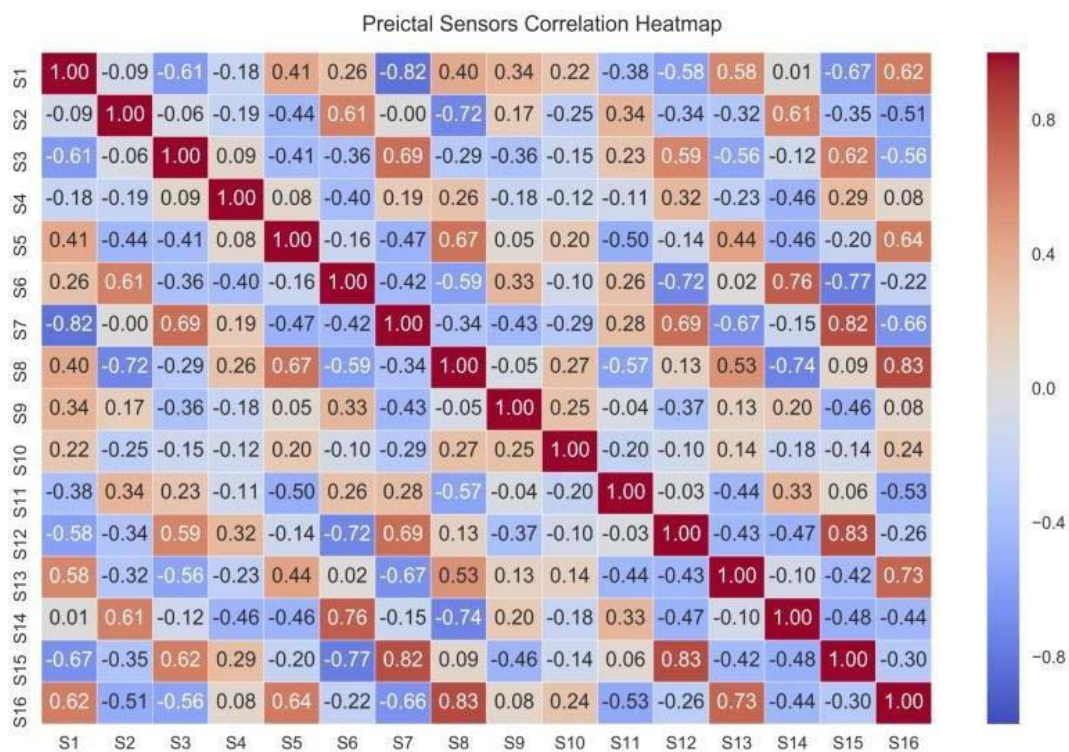


Figure 5: Heatmap of pairwise correlation values of iEEG sensor readings of Patient 1.

The difference between Patient 1 and Patient 2 and also patient 3 shows that the interictal and preictal data of any patient have different dispersion (which implies that they are statistically different), however their statistical features may only be used for building patient-specific seizure prediction systems. Based on our observations for different patients, we came to the conclusion that “There are no typical trends in either interictal or preictal data across different epileptic patients. The iEEG data of each patient has its own characteristics, and the statistical features can solely be meaningful for building patient-specific seizure prediction systems”.

### Dimensionality Reduction for Seizure Prediction Purposes

#### Investigating the use of Principal Component Analysis (PCA) for iEEG dimensionality Reduction in seizure prediction

Health care personnel should be informed as soon as possible when patients are experiencing a seizure attack. Similarly, patients and health personnel should be also warned as soon as possible about an impending seizure a patient may experience. It is therefore crucial to carry seizure detection and seizure prediction in real-time or as fast as possible. For such methods to be of practical use, the reduction of EEG/iEEG data is thus necessary for their fast implementation. Here we investigate two methods for iEEG data reduction. The first is Principal Component Analysis (PCA) and the second is channel reduction. PCA been widely used in reducing the dimensionality of the scalp Electroencephalogram (EEG) data for seizure onset detection. In this study, we investigate the possibility and effectiveness of PCA for iEEG dimensionality reduction for seizure prediction purposes. PCA

has been a popular tool for dimensionality reduction as it projects higher dimensional data to lower dimensional data. Efficient PCA holds a great potential for reducing the dimension of multi-channel EEG data, and hence speed up the subsequent machine learning algorithms (e.g., feature extraction and classification), minimizing the computation time of the overall EEG-based diagnostic tools for neurological disorders. PCA can help remove the redundancy in multi-variate EEG data, while maintaining most of the variance (information) in the observed variables. A useful measure is the “explained variance”, which can be calculated from the eigenvalues to measure how much information can be represented by each of the principal components.

The work presented in [15] demonstrated how PCA could be effectively used in selecting the optimal feature subset from the original EEG feature set, and therefore improve the epileptic seizure detection performance and also detection time. PCA has been shown to be an efficient data reduction tool that preserves the crucial EEG data variance and achieves reliable seizure detection results. The question that arises here is: Can we also use PCA for efficient iEEG dimensionality reduction in epileptic seizure prediction tasks? To answer this question, we applied PCA to the invasive EEG data under study. We then tested whether the 16 iEEG channels can be mapped into a fewer number of principal components. The common criterion is to select the minimum number of principal components such that 90-95% of the data variance is retained.

Figures 4 shows the individual principal components (colored in blue) and the cumulative principal components (colored in red) for both the preictal iEEG data of Patient 1. Unexpectedly, Figure 4

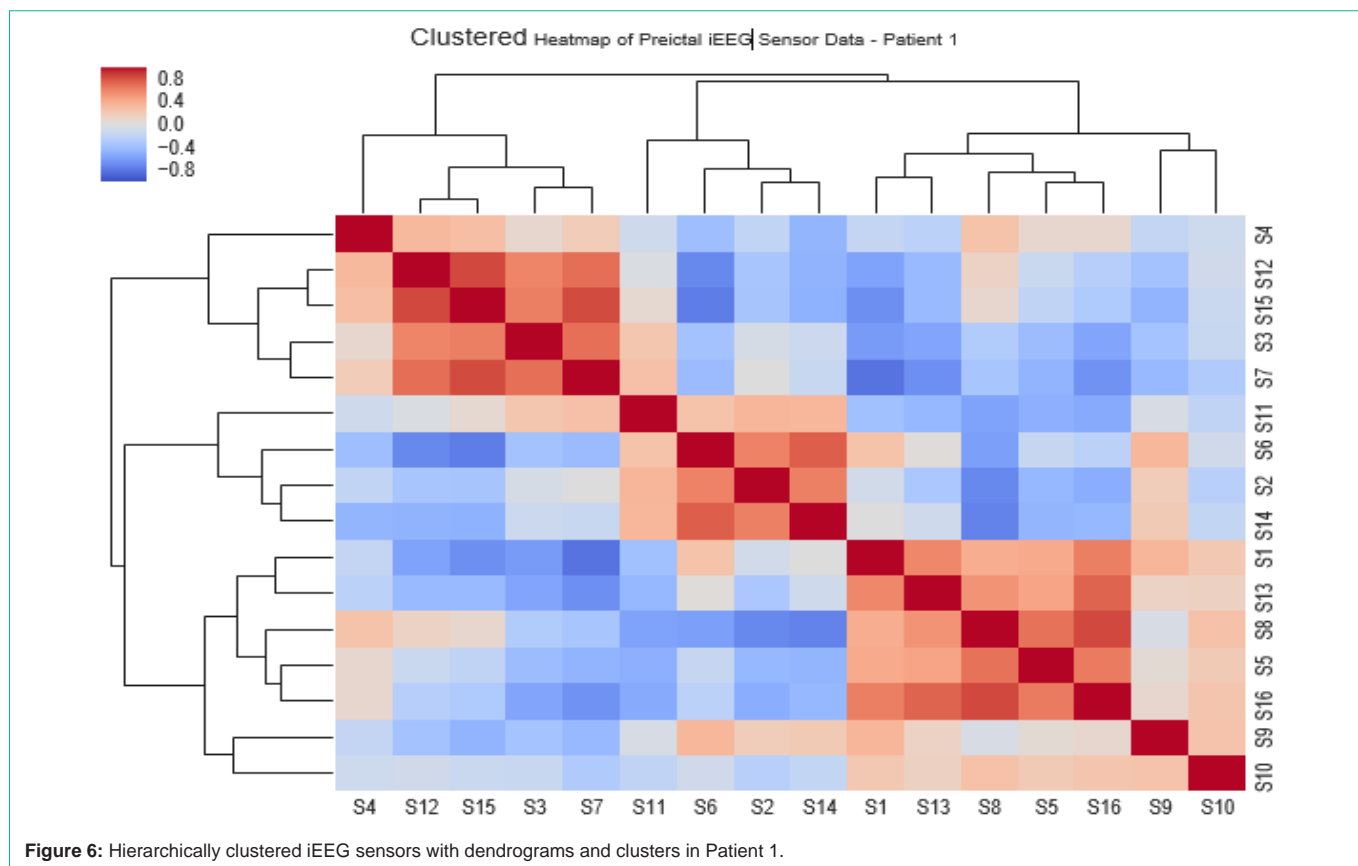


Figure 6: Hierarchically clustered iEEG sensors with dendrograms and clusters in Patient 1.

shows that, the first Principal Component (PC1) accounts for a small amount of the data variance (~16%), and the crucial 95% of the data variance is contained in 14 principal components. Therefore, only the 15<sup>th</sup> and 16<sup>th</sup> principal components can be dropped without losing too much information. This implies that we cannot rely on the PCA method for considerable iEEG dimensionality reduction in seizure prediction problems.

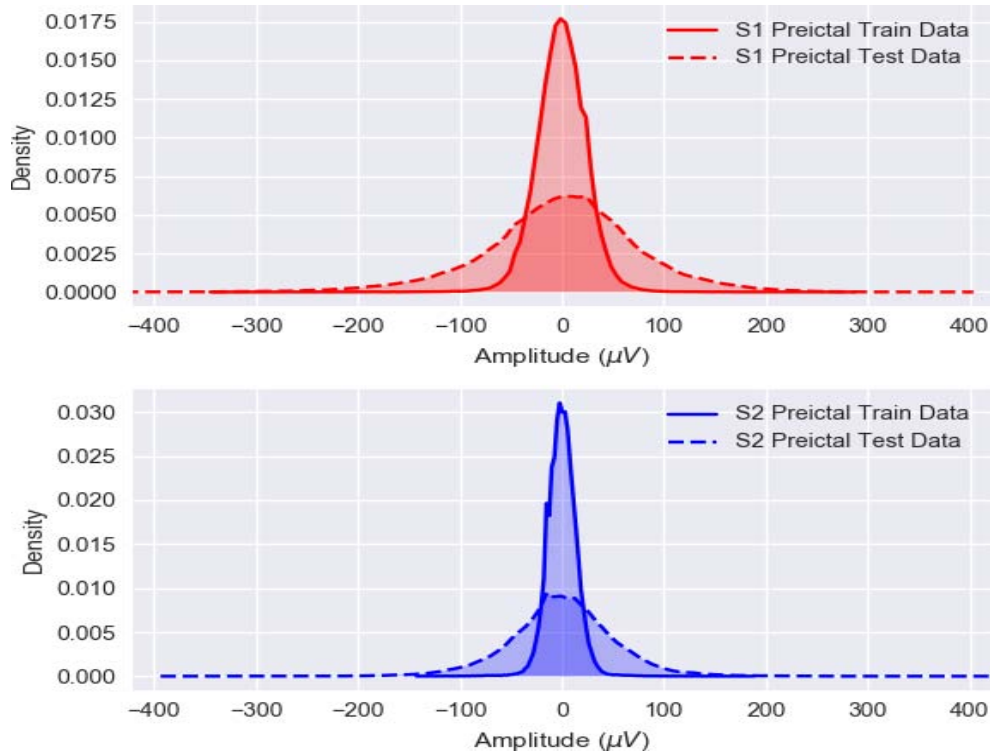
**Investigating the Exclusion of some iEEG Channels/Sensors**

Since PCA has (unexpectedly) been found to be inefficient in reducing the dimensionality of the human iEEG data in seizure prediction problems, we study whether channel selection (using only the data from some particular channels) could be used for reliable iEEG dimensionality reduction. The channels to be excluded are those that will be found to be less relevant or redundant. There is always a trade-off between selecting fewer iEEG channels and retaining as much spatial information as possible. The work presented in [16] investigated the epileptic seizure detection performance for different channel selection configurations. Using all EEG channels (22 channels) achieved the highest seizure detection performance. Selecting a moderately fewer number of EEG channels (e.g., 16 and 8) reduces the amount of data to 36-72 % of its original size, and results in a minor decay in the seizure detection performance [16]. Channel selection of scalp EEG data was then found to be a successful dimensionality reduction tool in seizure detection tasks. However, studies on channel selection of iEEG data for seizure prediction is still lacking and cannot be deduced from those carried for seizure

detection. This is because for seizure prediction, one major factor is the differentiation between preictal and interictal data, while for seizure detection it is the differentiation between ictal and interictal data.

Figure 5 depicts the correlation heatmaps of preictal iEEG sensor data of Patient 1, displaying the dependency relationships between the 16 iEEG sensor readings. As proven in [17], if the predictive ability of a certain channel X is covered by another channel, then channel X can be safely removed. (Figure 6) displays the clustered heatmaps of the preictal iEEG sensor data for Patient 1. The data displayed in (Figure 5) can be re-organized to uncover that it can be categorized into three main clusters; each cluster includes a set of sensors (channels) that have similar correlations. These are shown in (Figure 6). The first cluster comprises the sensors S3, S4, S7, S12, and S15, while the second cluster includes S2, S6, S11 and S14, and the third cluster includes S1, S5, S8, S9, S10, S13 and S16. If two sensors have analogous correlation profiles, one of them could be excluded. Following this rule, the first cluster could be represented only by S3, S4 and S12, and the second cluster could be represented only by S6 and S11, and the third cluster could be represented by S1, S5, S8 and S9.

For Patient 2, the same procedure was carried out, but a different combination of channel clustering was obtained. This is despite the fact that both Patients 1 and 2 have been diagnosed with focal epilepsy, yet (it should be noted that) we have found that they have different preictal iEEG channel correlation clusters. The main reason why the iEEG of different patients have different clustered heatmaps is that



**Figure 7:** Data mismatch in Patient 1's preictal iEEG sensor data.

patients have seizures with different epileptogenic zones. Patient 1, for example, was diagnosed as parietal-temporal lobe epilepsy, while Patient 2 was diagnosed as occipito-parietal lobe epilepsy. *In brief, the correlations between iEEG sensors vary depending on which brain region is affected and whether the seizure is focal (partial) or generalized.* Therefore, if seizure prediction researchers decided to use channel selection for the important reason of dimensionality reduction, it is important that the selection be customized for individual patients.

### iEEG Data Mismatch?

The terms “data mismatch”, “concept drift” and “covariate shift” have been used to refer to situations where the data distribution changes over time [18]. Often, the distribution of a particular data class (e.g., interictal or preictal) is assumed to not change over time, implying that the distribution of the historical data is the same as the distribution of the new data. While this assumption holds for many machine learning problems, it is not necessarily true for all problems. In some cases, the characteristics of the data vary over time, and hence the predictive models trained on historical data are no longer valid for making predictions on new unseen data. After screening the characteristics of the iEEG data under study, we found that, for each individual patient, the distribution of the interictal or the preictal iEEG data in the training set is different from that in the testing set. A potential reason for such a data mismatch is that the testing data is recorded a few months after recording the training data. During this time, the patient may have been positively or negatively influenced by factors such as anti-epileptic medications.

Figure 7 depicts how the preictal iEEG data distributions of Patient 1 have changed over time. The top plot explains how the density of

the training set of the preictal iEEG signals recorded by Channel S1 differs from that of the testing set. The testing data was recorded a few months after recording the training data. Similarly, the bottom plot demonstrates the variation in the preictal iEEG density for Channel 2. The above implies that the training of a seizure prediction system should be repeated once the system is observed to not work accurately, or the system should be designed so that its training is done in an adaptive fashion, i.e., the system continuously learns from its new experiences and updates its coefficients accordingly.

### Conclusion

This work provided, for the first time, a detailed quantitative and qualitative examination of the characteristics and behavior of the human brain iEEG data during the preictal and interictal brain states. From the conducted analyses, we can deduce the following recommendations for researchers building future EEG-based seizure prediction algorithms: i) There is no typical trend in either interictal or preictal data across different epileptic patients. The iEEG data of each patient has its own characteristics, and its statistical features can solely be meaningful for building patient-specific seizure prediction systems. The statistical features extracted from the data of a group of patients may not generalize to other patients or even to the same patients at later time; ii) for iEEG dimensionality reduction, the use of PCA is not recommended, neither is the use of channel deletion. Each individual principal component or iEEG channel carries a non-trivial amount of information and most of the principal components or iEEG channels should be employed for efficient feature learning and classification. Therefore, more efficient techniques for iEEG data reduction should be sought. Such observations demonstrate why

epileptic seizure prediction using iEEG data remains a challenging topic and why traditional machine learning algorithms that rely on domain-based features have a limited seizure prediction performance.

## Acknowledgment

Rabab Ward and Ramy Hussein were supported by NPRP grant # NPRP12S-0305-190231 from the Qatar National Research Fund (a member of Qatar Foundation). The findings achieved herein are solely the responsibility of the authors.

## References

1. G Rogers. "Epilepsy: the facts". Primary Health Care Research & Development. 2010; 11: 413.
2. Acharya UR, Sree SV, Swapna G, Martis RJ, Suri JS. "Automated EEG analysis of epilepsy: a review". Knowledge-Based Systems. 2013; 45: 147-165.
3. Alessandro M, Esteller R, Vachtsevanos G, Hinson A, Echaz J, Litt B, et al. "Epileptic seizure prediction using hybrid feature selection over multiple intracranial EEG electrode contacts: a report of four patients". IEEE transactions on biomedical engineering. 2003; 50: 603-615.
4. Gadhoumi K, Lina JM, Mormann F, Gotman J. "Seizure prediction for therapeutic devices: A review". Journal of neuroscience methods. 2016; 260: 270-282.
5. Assi EB, Nguyen DK, Rihana S, Sawan M. "Towards accurate prediction of epileptic seizures: A review". Biomedical Signal Processing and Control. 2017; 34: 144-157.
6. Mormann F, Andrzejak RG, CE Elger, et al. "Seizure prediction: the long and winding road". Brain. 2006; 130: 314-333.
7. Kuhlmann L, Freestone D, Lai A, Burkitt AN, Fuller K, Grayden DB, et al. "Patient-specific bivariate synchrony-based seizure prediction for short prediction horizons". Epilepsy research. 2010; 91: 214-231.
8. Cook MJ, O'Brien TJ, Berkovic SF, Murphy M, Morokoff A, Fabinyi G, et al. "Prediction of seizure likelihood with a long-term, implanted seizure advisory system in patients with drug-resistant epilepsy: a first-in-man study". The Lancet Neurology. 2013; 12: 563-571.
9. Karoly PJ, Ung H, Grayden DB, Kuhlmann L, Leyde K, Cook MJ, et al. "The circadian profile of epilepsy improves seizure forecasting". Brain. 2017; 140: 2169-2182.
10. Eberlein M, Hildebrand R, Tetzla R, Homann N, Kuhlmann L, Brinkmann B, et al. "Convolutional Neural Networks for Epileptic Seizure Prediction". 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). 2018; IEEE: 2577-2582.
11. Truong ND, Nguyen AD, Kuhlmann L, Bonyadi MR, Yang J, Ippolito S, et al. "Convolutional neural networks for seizure prediction using intracranial and scalp electroencephalogram". Neural Networks. 2018; 105: 104-111.
12. Kuhlmann L, Karoly P, Freestone DR, Brinkmann BH, Temko A, Barachant A, et al. "Epilepsyecosystem.org: crowd-sourcing reproducible seizure prediction with long-term human intracranial EEG". Brain. 2018; 141: 2619-2630.
13. Daoud H, Bayoumi MA. "Efficient Epileptic Seizure Prediction Based on Deep Learning". IEEE Transactions on Biomedical Circuits and Systems. 2019; 13: 804-813.
14. Cook MJ, O'Brien TJ, Berkovic SF, Murphy M, Morokoff A, Fabinyi G, et al. "Prediction of seizure likelihood with a long-term, implanted seizure advisory system in patients with drug-resistant epilepsy: a first-in-man study". The Lancet Neurology. 2013; 12: 563-571.
15. Wang L, Xue W, Li Y, Luo M, Huang J, Cui W, et al. "Automatic epileptic seizure detection in EEG signals using multi-domain feature extraction and nonlinear analysis". Entropy. 2017; 19: 222.
16. Shah V, Golmohammadi M, Ziyabari S, Von Weltin E, Obeid I, Picone J. "Optimizing channel selection for seizure detection". In Signal Processing in Medicine and Biology Symposium (SPMB). 2017; IEEE: 1-5.
17. Hall MA. "Correlation-based feature selection for machine learning". 2000.
18. Zliobaite I, Pechenizkiy M, Gama J. "An overview of concept drift applications". In Big data analysis: new algorithms for a new society. Springer. 2016: 91-114.