Human Intracranial EEG Quantitative Analysis and Automatic Feature Learning for Epileptic Seizure Prediction

Ramy Hussein*, Mohamed Osama Ahmed, Rabab Ward, Z. Jane Wang, Levin Kuhlmann, and Yi Guo

Abstract—*Objevtive:* **The aim of this study is to develop an efficient and reliable epileptic seizure prediction system using intracranial EEG (iEEG) data, especially for people with drug-resistant epilepsy. The prediction procedure should yield accurate results in a fast enough fashion to alert patients of impending seizures.** *Methods:* **We quantitatively analyze the human iEEG data to obtain insights into how the human brain behaves before and between epileptic seizures. We then introduce an efficient pre-processing method for reducing the data size and converting the time-series iEEG data into an image-like format that can be used as inputs to convolutional neural networks (CNNs). Further, we propose a seizure prediction algorithm that uses cooperative multi-scale CNNs for automatic feature learning of iEEG data.** *Results:* **1) iEEG channels contain complementary information and excluding individual channels is not advisable to retain the spatial information needed for accurate prediction of epileptic seizures. 2) The traditional PCA is not a reliable method for iEEG data reduction in seizure prediction. 3) Hand-crafted iEEG features may not be suitable for reliable seizure prediction performance as the iEEG data varies between patients and over time for the same patient. 4) Seizure prediction results show that our algorithm outperforms existing methods by achieving an average sensitivity of 87.85% and AUC score of 0.84.** *Conclusion:* **Understanding how the human brain behaves before seizure attacks and far from them facilitates better designs of epileptic seizure predictors.** *Significance:* **Accurate seizure prediction algorithms can warn patients about the next seizure attack so they could avoid dangerous activities. Medications could then be administered to abort the impending seizure and minimize the risk of injury. Closed-loop seizure intervention systems could also help to prevent seizures in patients with drug-resistant epilepsy.**

Index Terms—**Electroencephalogram (EEG), epilepsy, seizure prediction, quantitative analysis, convolutional neural networks.**

I. 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 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, or blurred vision. Unfortunately, 20-40% of people with epilepsy continue to have seizures despite treatment [3]. Further, the patient's quality of life is significantly degraded by the anxiety associated with the unpredictable nature of seizures and the consequences therefrom. This motivated researchers to develop seizure prediction systems [4]. The ability to predict seizures with high accuracies would make individualized epilepsy treatment possible (*e.g.*, tailored therapies with less side effects). With warnings of impending seizures, the patients can take their precautions and avoid any probable injuries. This vision inspired the proposed research.

Epilepsy surgery may improve the quality of life of epileptic patients who are drug-resistant to epilepsy. Brain surgery may reduce the number of seizure attacks and hence limit the risk of permanent brain damage. It may, however, involves serious risks such as stroke, paralysis, speech issues, memory problems, loss of vision, loss of motor skills, and sometimes more seizures [5]. Besides epilepsy surgery, people with drug-resistant epilepsy can benefit from methods that predict seizures far enough in advance. Despite the fact that epileptic seizures seem unpredictable and often occur without warning, recent investigations have demonstrated that seizures do not strike at random [6], [7]. Intracranial electroencephalography (iEEG) is the prime tool used for forecasting epileptic seizure attacks. The iEEG data recorded preceding seizures are analyzed to identify pattern(s) that indicate upcoming seizures.

Over the past decade, researchers have made several attempts to develop robust iEEG-based seizure prediction methods. The prediction performance, however, was affected by the lack of long-term preictal (before seizures) and interictal (between seizuress, *i.e.*, baseline) iEEG recordings [8], [9]. In 2013, Cook *et al.* designed a seizure advisory system that can predict the probability of seizure occurrence in the minutes or hours in advance [10]. Fifteen adults with refractory epilepsy were each implanted with 16 electrodes. This 16-channel seizure prediction device allows uninterrupted recording of intracranial EEG for a long period of time (6-36 months). The seizure prediction algorithm however yielded prediction performances with a large subject variability for 15 patients. For example, the highest seizure prediction sensitivity of 100%

was reached for two patients, while a low sensitivity of 17-45% was reported for three patients. Improving prediction performances for these three patients is important to ensure that seizure prediction is possible for different patients with drug-resistant epilepsy.

Recently, a big dataset of long-term iEEG recordings has been collected from the aforementioned three patients over 374-559 days. A subset of this dataset was made publicly available and used in the Melbourne University AES/MathWorks/NIH Seizure Prediction Competition organized in November, 2016 on Kaggle.com[1]. Kuhlmann *et al.* describe the human iEEG dataset used in this contest and the results achieved by the top eight seizure prediction algorithms [11]. These algorithms adopted a diverse range of iEEG feature extraction, selection, and classification techniques; and achieved superior performance compared to those in [10]. For example, the winning solution was based on an ensemble of 11 different machine learning models; each was designed subject-specific. A detailed description of the 11 models - including the extracted iEEG features as well as the utilized classification models - is provided in [12]. To assess the predictive power of these models, the area under the ROC curve was obtained for each model individually. The best reported prediction score for the contest test set was 0.854 [11], [12].

In this work, we propose a Convolutional Neural Network (CNN)-based algorithm for reliable prediction of epileptic seizures. Major contributions of this study are as follows: (1) We first provide an extensive quantitative analysis of the human iEEG data preceding and between epileptic seizures; (2) We introduce an efficient pre-processing strategy for transferring time-series iEEG data into image-like format that can better leverage the power of CNNs; (3) We propose a multi-scale CNN architecture that can learn different representations of iEEG data efficiently; and (4) We demonstrate that the proposed seizure prediction algorithm works reliably for different patients with refractory focal epilepsy, and outperforms the state-of-the-art algorithms for this problem. To the best of our knowledge, this is the first attempt to automatically learn both local and high-abstracted iEEG features simultaneously. This is in contrast to previous studies that rely on extracting hand-crafted features.

II. HUMAN iEEG QUANTITATIVE ANALYSIS

In this section, we first describe the human intracranial EEG data used in our study and then present some hidden insights of the data, in order to better understand the unpredictable nature of epileptic seizures that we have been attempting to quantify.

A. Subjects and Data

The iEEG data used is from the 2016 Kaggle seizure prediction competition and is described in [11]. The data were recorded constantly from humans suffering from refractory (drug-resistant) focal epilepsy using the NeuroVista Seizure Advisory System (described in [10]). Sixteen electrodes (4×4
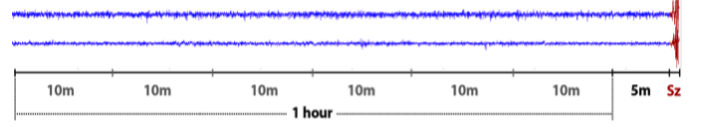
Figure 1. Examples of one-hour preictal iEEG clips collected by a 16-channel device with a 5-minute offset before seizures. For convenience, only two channels are plotted.

contact strips) were implanted in every patient, directed to the presumed seizure focus, and connected to a telemetry unit embedded in the subclavicular area. A rechargeable battery powered the embedded device. 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 saved in a removable flash drive.

Three drug-resistant patients who also had the least seizure prediction performance in [10] were chosen for our study. The selection of these patients was propelled by the intention to demonstrate that accurate seizure prediction is also feasible for patients with drug-resistant epilepsy. The large number of seizures recorded per patient (∼380) gives us the chance to explore the common signature within the iEEG preceding seizures of every patient. The three patients were females and they all had resective surgery before the trial. The first, second, and third patient were 22, 51, and 53 years old at the time of the trial, but were diagnosed with epilepsy at age of 16, 10, and 15, respectively. The first patient suffered from parietotemporal focal epilepsy and was receiving the antiepileptic drugs of carbamazepine, lamotrigine, and phenytoin. The second patient was diagnosed with occipitoparietal focal epilepsy and was taking carbamazepine medication. The third patient had seizures of frontotemporal origin and was taking perampanel, phenytoin, and Lacosamide for treatment.

Neuroscientists have found that the temporal dynamics of the brain activity of epileptic patients can often be categorized into four different states: *preictal* (prior to seizure), *ictal* (seizure), *postictal* (after seizure), and *interictal* (between seizures). The key challenge in seizure prediction is to differentiate between the *preictal* and *interictal* brain states in humans with epilepsy. Only lead seizures, characterized as seizures occurring at least 4 hours after a previous seizure, were used for every patient. The captured iEEG data were labeled and separated into training and testing sets (for training and testing the proposed seizure prediction method). To mimic real-life clinical settings, the testing data was recorded quite much later after recording the training data. To avert the signal non-stationarity in the immediate time period following sensors' implantation, both sets of data (training and testing) were obtained from data recorded between 1 and 7 months after sensors' implantation.

As shown in Fig. 1, 65-min preictal data clips were extracted preceding the lead seizures. They are divided into six 10-minute clips followed by a 5-minute offset segment before each seizure. Every two consecutive segments were also separated by a 10-second gap. In a similar manner, interictal clips were segmented from 61 minutes long recordings that

Table I
DESCRIPTION OF THE SEIZURE PREDICTION INTRACRANIAL EEG DATASET [10].

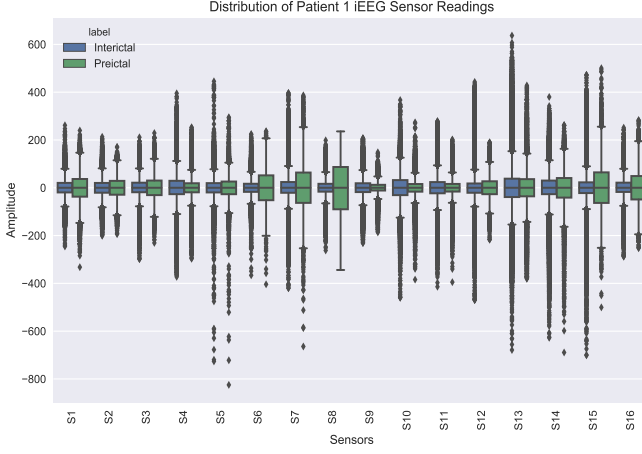| Patient	Age (years)	Gender	Recording duration (days)	Seizures	Lead seizures	No. of training clips		No. of testing clips	
						Preictal	Interictal	Preictal	Interictal
1	22	Female	559	390	231	256	570	16	46
2	51	Female	393	204	186	222	1836	18	279
3	53	Female	374	545	216	255	1908	18	188

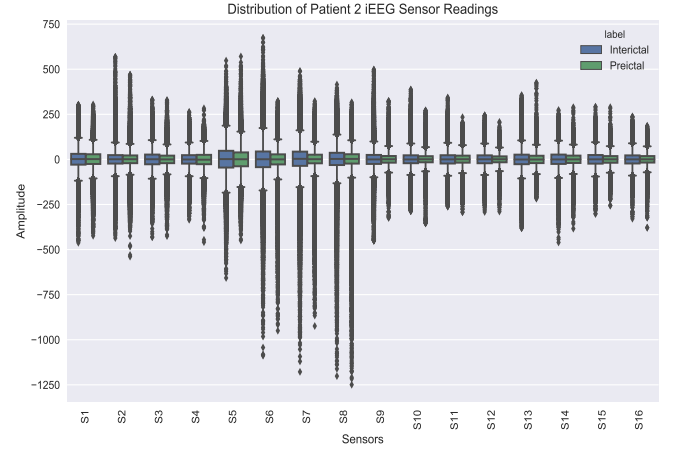Figure 2. Boxplots of Patient 1's *interictal* and *preictal* iEEG data.

Figure 3. Boxplots of Patient 2's *interictal* and *preictal* iEEG data.

started at arbitrarily time with a minimum gap of 3 hours before and 4 hours after any seizure. They also were split into six 10-minute clips with 10-second spacing. A summary of the data for every patient and the number of iEEG clips in the training and testing data sets are shown in Table I.

B. Are Interictal and Preictal iEEGs Statistically Different?

Most seizure prediction problems focus on the discrimination between the interictal (baseline) and preictal brain states. The question that arises here is: Are the interictal and preictal iEEGs statistically different? If yes, can we use their statistical features to differentiate between them? In an effort to address this question, we use descriptive statistics to describe whether these data are similar or different. The boxplot is a descriptive statistics tool that presents the data distribution using the five number summary: minimum, first quartile, median, third quartile, and maximum. To interpret a boxplot, we focus on three measures:

- Median: The median is indicated by the horizontal line in the center of the box.
- Range: The range describes the spread of the data, represented by the vertical distance between the minimum and maximum values.
- Interquartile Range (IQR): IQR shows the likely range of variations and is represented by the length of the box spanning the first quartile to the third quartile.

Fig. 2 shows a comparative boxplot for Patient 1's interictal and preictal iEEG recordings. It can be noticed that, for all 16 sensors, the median values of the interictal and the preictal

iEEG are both around zero. Moreover, for most of the sensor readings (∼11 sensors), the overall range and interquartile range of the preictal data are much greater than those of the interictal data. However, for the rest of the sensor readings, the overall range and interquartile range of the preictal data are smaller than those of the interictal data. Interestingly, both data sets display potential outliers at both ends.

Despite the fact that the interictal and preictal data seem to have different dispersion (which implies that they are statistically different), we believe that such statistical features can only be used for patient-specific seizure prediction systems. Figure 3, for instance, depicts that the distributions of interictal and preictal iEEG sensor readings for Patient 2 are quite different from those of Patient 1. Also, within the same iEEG data class (*e.g.*, preictal), the overall range and IQR of the iEEG sensor readings of Patient 2 are less than those of Patient 1. Based on our observations for different patients, 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 a seizure prediction system for this particular patient.

C. Can PCA be Efficient for iEEG Data Reduction?

Principal Component Analysis (PCA), is a prime tool for data reduction as it projects higher dimensional data to lower dimensional data. PCA has been widely used in reducing the dimensionality of the scalp (surface) EEG data, particularly for seizure onset detection. It also helps reduce the dimension
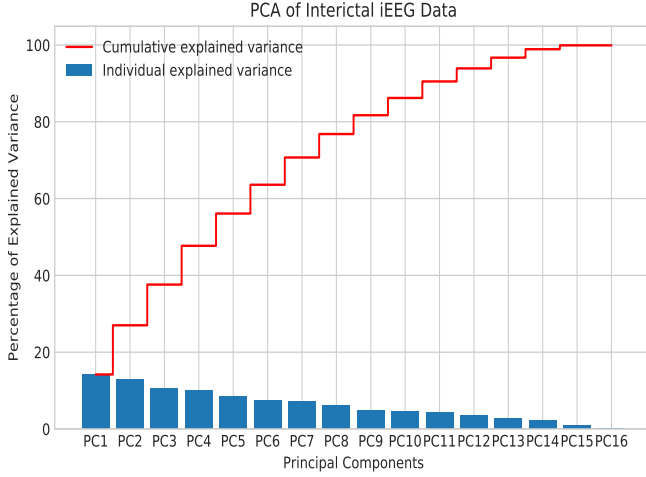
Figure 4. Explained variance by different principle components of interictal iEEG data for Patient 1.
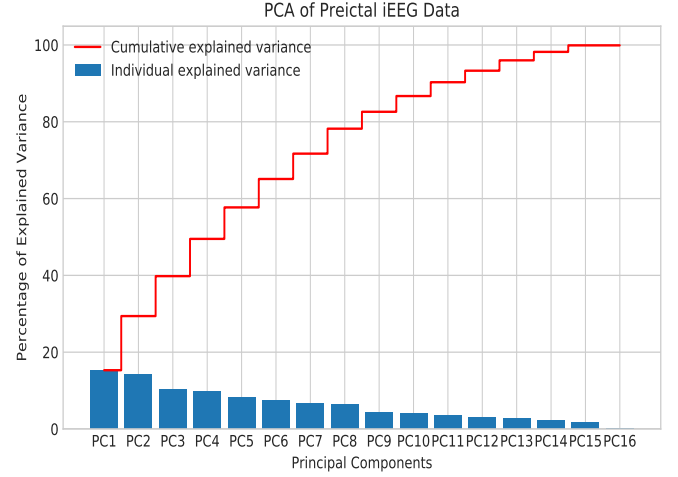

Figure 5. Explained variance by different principle components of preictal iEEG data for Patient 1.

of multi-channel EEG signals, and hence speed up the subsequent machine learning algorithms (*e.g.*, feature extraction and classification), minimizing the computation time of the overall EEG-based diagnosis system of neurological disorders. PCA also helps remove the redundancy (correlation) in multivariate 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 [13] demonstrates how PCA can be effectively used in choosing the optimal feature subset from the original EEG feature set, and therefore improve the epileptic seizure detection performance. PCA has shown to be a valuable data reduction tool that preserves a significant amount of the EEG data variance and achieves stable seizure detection results. The question that arises here is: Can we use PCA to efficiently reduce the iEEG data dimensionality for epileptic seizure prediction? The iEEG dataset under study has 16 feature columns (16 sensors), a training set of 5,047 instances, and a test set of 565 instances (Table I shows the details of the used dataset). To answer the above question, we applied PCA to the interictal and preictal iEEG data under study. We then tested whether the 16-sensor readings can be mapped into a fewer number of principal components as follows:

We chose the minimum number of principal components such that 95% of the variance was retained. Figures 4 and 5 show the individual principal components (colored in blue) and the cumulative principal components (colored in red) for both the interictal and preictal iEEG data clips for Patient 1. Unexpectedly, Fig. 4 shows that, for interictal iEEG, the first principal component accounts for a small amount of the data variance ($\sim$16%), and 95% of the variance is contained in 14 principal components. Only the 15th and 16th principal components can be dropped without losing too much information. Similarly, Fig. 5 represents the explained variance by 16 principal components of the preictal iEEG data. It provides

useful insights on the distribution of variance across different principal components and suggests that, at least, 14 principal components are needed to preserve most of the data variance. This suggests that PCA is not a reliable method for iEEG data reduction for our problem.

D. Can We Exclude some EEG Channels/Sensors?

Since PCA has failed to reduce the dimensionality of the human iEEG data when most of its variance is retained, researchers have considered turning some of the sensors off in order to reduce the amount of data and the computation time. The excluded channels are usually less relevant or redundant. It is a trade-off between selecting fewer EEG channels and retaining as much spatial information as possible. The work presented in [14] investigates the epileptic seizure detection performance for different channel configurations. As expected, using all EEG channels from a 10-20 EEG configuration achieves the best seizure detection performance. Choosing 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 degradation in the seizure detection performance [14].

Whilst many researchers have successfully used channel selection for surface EEG data reduction applied to seizure detection, there is no clear consensus regarding turning off any or some intracranial EEG channels in the seizure prediction task. The correlation heatmap gives us a way to gain some insights into what sensors (channels) are correlated and whether we can exclude certain sensors. Figures 6 and 7 depict the correlation heatmaps of iEEG sensor data during the interictal and preictal brain states for Patient 1. The positive correlations are displayed by increasing the red hues and the negative correlations are displayed by increasing the blue shades. Faded red and blue colors denote low positive and negative correlations, respectively. The correlation coefficients are also included in the heatmaps, displaying the dependency relationships between 16 iEEG sensors.
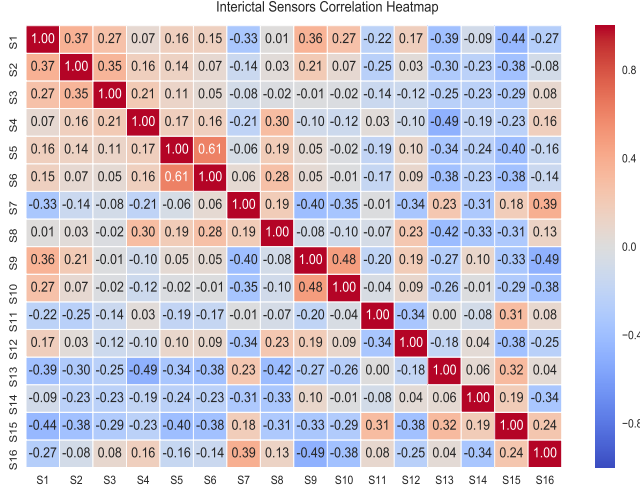
Interictal Sensors Correlation Heatmap

	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16
S1	1.00	0.37	0.27	0.07	0.16	0.15	-0.33	0.01	0.36	0.27	-0.22	0.17	-0.39	-0.09	-0.44	-0.27
S2	0.37	1.00	0.35	0.16	0.14	0.07	-0.14	0.03	0.21	0.07	-0.25	0.03	-0.30	-0.23	-0.38	-0.08
S3	0.27	0.35	1.00	0.21	0.11	0.05	-0.08	-0.02	-0.01	-0.02	-0.14	-0.12	-0.25	-0.23	-0.29	0.08
S4	0.07	0.16	0.21	1.00	0.17	0.16	-0.21	0.30	-0.10	-0.12	0.03	-0.10	-0.49	-0.19	-0.23	0.16
S5	0.16	0.14	0.11	0.17	1.00	0.61	-0.06	0.19	0.05	-0.02	-0.19	0.10	-0.34	-0.24	-0.40	-0.16
S6	0.15	0.07	0.05	0.16	0.61	1.00	0.06	0.28	0.05	-0.01	-0.17	0.09	-0.38	-0.23	-0.38	-0.14
S7	-0.33	-0.14	-0.08	-0.21	-0.06	0.06	1.00	0.19	-0.40	-0.35	-0.01	-0.34	0.23	-0.31	0.18	0.39
S8	0.01	0.03	-0.02	0.30	0.19	0.28	0.19	1.00	-0.08	-0.10	-0.07	0.23	-0.42	-0.33	-0.31	0.13
S9	0.36	0.21	-0.01	-0.10	0.05	0.05	-0.40	-0.08	1.00	0.48	-0.20	0.19	-0.27	0.10	-0.33	-0.49
S10	0.27	0.07	-0.02	-0.12	-0.02	-0.01	-0.35	-0.10	0.48	1.00	-0.04	0.09	-0.26	-0.01	-0.29	-0.38
S11	-0.22	-0.25	-0.14	0.03	-0.19	-0.17	-0.01	-0.07	-0.20	-0.04	1.00	-0.34	0.00	-0.08	0.31	0.08
S12	0.17	0.03	-0.12	-0.10	0.10	0.09	-0.34	0.23	0.19	0.09	-0.34	1.00	-0.18	0.04	-0.38	-0.25
S13	-0.39	-0.30	-0.25	-0.49	-0.34	-0.38	0.23	-0.42	-0.27	-0.26	0.00	-0.18	1.00	0.06	0.32	0.04
S14	-0.09	-0.23	-0.23	-0.19	-0.24	-0.23	-0.31	-0.33	0.10	-0.01	-0.08	0.04	0.06	1.00	0.19	-0.34
S15	-0.44	-0.38	-0.29	-0.23	-0.40	-0.38	0.18	-0.31	-0.33	-0.29	0.31	-0.38	0.32	0.19	1.00	0.24
S16	-0.27	-0.08	0.08	0.16	-0.16	-0.14	0.39	0.13	-0.49	-0.38	0.08	-0.25	0.04	-0.34	0.24	1.00

Figure 6. Heatmap of pairwise correlation values of Patient 1's interictal iEEG's 16 sensors.

Preictal Sensors Correlation Heatmap

	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16
S1	1.00	-0.09	-0.61	-0.18	0.41	0.26	-0.82	0.40	0.34	0.22	-0.38	-0.58	0.58	0.01	-0.67	0.62
S2	-0.09	1.00	-0.06	-0.19	-0.44	0.61	-0.00	-0.72	0.17	-0.25	0.34	-0.34	-0.32	0.61	-0.35	-0.51
S3	-0.61	-0.06	1.00	0.09	-0.41	-0.36	0.69	-0.29	-0.36	-0.15	0.23	0.59	-0.56	-0.12	0.62	-0.56
S4	-0.18	-0.19	0.09	1.00	0.08	-0.40	0.19	0.26	-0.18	-0.12	-0.11	0.32	-0.23	-0.46	0.29	0.08
S5	0.41	-0.44	-0.41	0.08	1.00	-0.16	-0.47	0.67	0.05	0.20	-0.50	-0.14	0.44	-0.46	-0.20	0.64
S6	0.26	0.61	-0.36	-0.40	-0.16	1.00	-0.42	-0.59	0.33	-0.10	0.26	-0.72	0.02	0.76	-0.77	-0.22
S7	-0.82	-0.00	0.69	0.19	-0.47	-0.42	1.00	-0.34	-0.43	-0.29	0.28	0.69	-0.67	-0.15	0.82	-0.66
S8	0.40	-0.72	-0.29	0.26	0.67	-0.59	-0.34	1.00	-0.05	0.27	-0.57	0.13	0.53	-0.74	0.09	0.83
S9	0.34	0.17	-0.36	-0.18	0.05	0.33	-0.43	-0.05	1.00	0.25	-0.04	-0.37	0.13	0.20	-0.46	0.08
S10	0.22	-0.25	-0.15	-0.12	0.20	-0.10	-0.29	0.27	0.25	1.00	-0.20	-0.10	0.14	-0.18	-0.14	0.24
S11	-0.38	0.34	0.23	-0.11	-0.50	0.26	0.28	-0.57	-0.04	-0.20	1.00	-0.03	-0.44	0.33	0.06	-0.53
S12	-0.58	-0.34	0.59	0.32	-0.14	-0.72	0.69	0.13	-0.37	-0.10	-0.03	1.00	-0.43	-0.47	0.83	-0.26
S13	0.58	-0.32	-0.56	-0.23	0.44	0.02	-0.67	0.53	0.13	0.14	-0.44	-0.43	1.00	-0.10	-0.42	0.73
S14	0.01	0.61	-0.12	-0.46	-0.46	0.76	-0.15	-0.74	0.20	-0.18	0.33	-0.47	-0.10	1.00	-0.48	-0.44
S15	-0.67	-0.35	0.62	0.29	-0.20	-0.77	0.82	0.09	-0.46	-0.14	0.06	0.83	-0.42	-0.48	1.00	-0.30
S16	0.62	-0.51	-0.56	0.08	0.64	-0.22	-0.66	0.83	0.08	0.24	-0.53	-0.26	0.73	-0.44	-0.30	1.00

Figure 7. Heatmap of pairwise correlation values of Patient 1's preictal iEEG's 16 sensors.

In [15], Hall *et al.* proposed a correlation-based feature selection method and concluded that: *"a representative feature (sensor) subset is one that comprises features (sensors) highly correlated with the target (iEEG class), yet uncorrelated with each other"*. Following the same rule, we watched out for the iEEG sensors that are highly correlated during the interictal and preictal brain states. If a sensor's predictive ability is covered by another, then it can safely be removed. As shown in Figures 6 and 7, iEEG sensors are not heavily correlated neither in the interictal nor in the preictal brain states. The cross-correlation values range between −0.49 and 0.61 for the interictal iEEG sensor readings, and between −0.82 and 0.83 for the preictal iEEG sensor readings. Therefore, it is advised not to exclude any of the iEEG channels in order to retain the spatial information needed for accurate prediction of epileptic seizures.

E. Are Neighbor iEEG Sensors more Correlated than Distant Ones?

Fig. 8 shows an example of Computerized Tomography (CT) scan of the head of one patient, and reveals the locations of 16 iEEG electrodes on the cortical surface of the brain. The questions here are: Are adjacent iEEG sensors more correlated than distant ones? If yes, are they positively or negatively correlated? In an attempt to answer these questions, we first identified the approximate location of the iEEG electrodes on the cerebral cortex of epileptic patients under study. Unlike most seizure prediction systems that deploy iEEG electrodes implanted across both brain hemispheres, the electrodes of the seizure advisory system used in the studied dataset were placed on one side over the small brain region presumed to include the epileptogenic zone. In patients with bilateral temporal lobe epilepsy, electrodes were positioned over the brain hemisphere that generates most seizures.

Figures 9 and 10 display the clustered heatmaps of preictal iEEG sensor data for Patients 1 and 2, respectively. As shown in Fig. 9, the iEEG sensor data of Patient 1 are grouped into
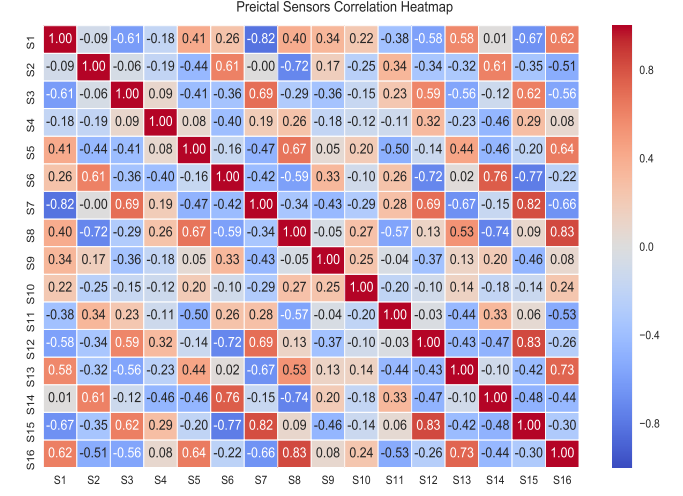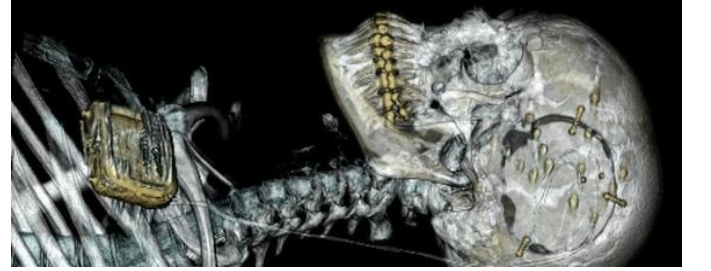
Figure 8. CT scan of the NeuroVista seizure advisory system implanted in a patient [10].

three main clusters; each includes a set of sensors that have similar correlations. The first cluster comprises the sensors S3, S4, S7, S12, and S15 (S3 and S4 are neighbors), while the second cluster includes S2, S6, S11, and S14, and the third cluster includes S1, S5, S8, S9, S10, S13, and S16 (S9 and S10 are neighbors). As can be seen, adjacent sensors do not necessarily have analogous correlation profiles. These clustering methods may allow researchers to determine which particular sensors can act as seizure activity sources and which can act as sinks. However, categorizing iEEG sensors into ictal activity generators and sinks based on the correlation profiles is not well-established yet. In Fig. 10 (for Patient 2), the first cluster includes S6, S7, and S8 (they all are neighbors), the second cluster includes S1, S2, S3, S4, S9, S10, S11, and S12 (where S1, S2, S3, and S4 are neighbors & S9, S10, S11, and S12 are also neighbors), and the third cluster comprises S5, S13, S14, S15, and S16 (S5 is a non-neighbor to others).

Despite the fact that both Patients 1 and 2 are diagnosed with focal epilepsy, it is important to realize that they have different preictal iEEG data clusters. Similarly, Patient 3 has a totally dissimilar iEEG clustered heatmap (omitted for the lack of space). The main reason why patients have different clustered heatmaps of preictal iEEGs is that they have different epileptogenic zones. Patient 1, for example, was diagnosed by parietal-temporal lobe epilepsy, while Patient 2 was diagnosed by occipito-parietal lobe epilepsy. In brief, the correlations
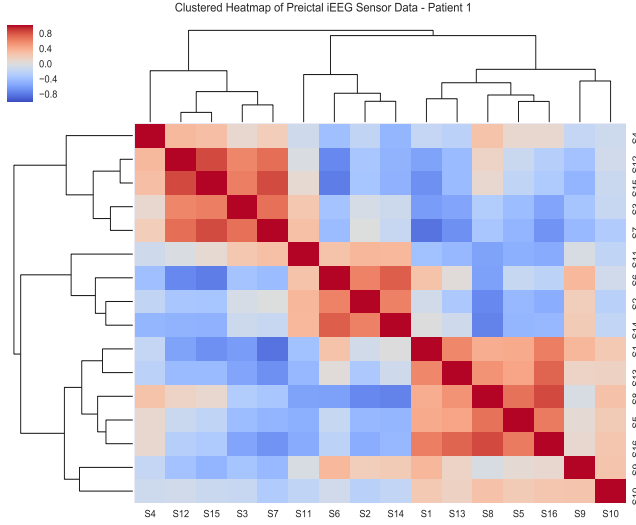
Figure 9. Hierarchically clustered preictal iEEG sensors with dendrograms and clusters in Patient 1.
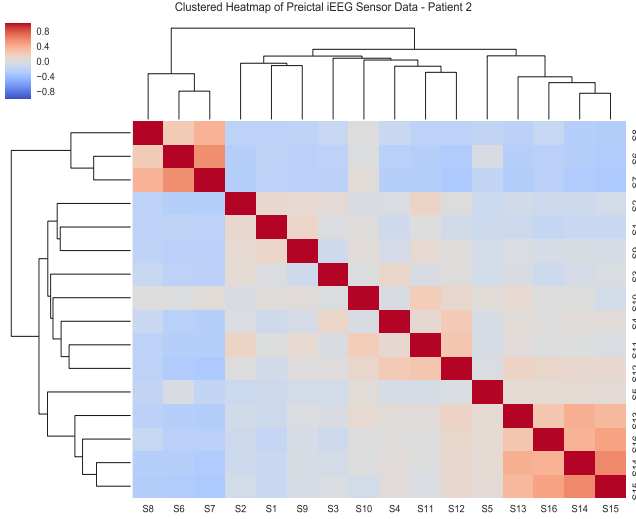

Figure 10. Hierarchically clustered preictal iEEG sensors with dendrograms and clusters in Patient 2.

between iEEG sensors vary depending on which brain region is affected and whether the seizure is focal (partial) or generalized.

To summarize, the conducted quantitative analysis of iEEG data suggests that iEEG data analysis for seizure prediction can be quite different from regular EEG data analysis for other biomedical classification/prediction problems; and it will be desirable to design an iEEG-based seizure prediction algorithm that (i) uses all iEEG channels for proficient feature extraction, (ii) avoids using PCA for iEEG dimensionality reduction, and (iii) avoids using domain-based hand-crafted features for iEEG classification. Such observations help explain why Epileptic Seizure Prediction using iEEG data remains a challenging topic and traditional machine learning algorithms using hand-crafted features did not yield satisfactory performance in the literature. Therefore, the conducted quantitative analysis in

Section II motivates us to explore a deep learning-based approach for seizure prediction using iEEG data. The proposed approach in Section III has the desired advantages as described above.

III. METHODOLOGY

Convolutional Neural Networks (CNNs) have been proven to achieve astonishing results in different research areas such as face recognition [16], object detection [17], image classification [18], and speech recognition [19]. In this study, we propose the use of CNNs for epileptic seizure prediction.

We rely on CNNs to extract different representations of the iEEG data to use for iEEG classification. The proposed neural network architecture is unique as it allows us to recover both the local features (through smaller convolutions) and the high abstracted features (with larger convolutions). Preprocessing and preparing the data for the feature extraction and classification processes are explained below.

A. iEEG Preprocessing

The first step of building our pipeline system is to preprocess the time-series iEEG data by 1) reducing the data dimensionality, 2) dividing the data into smaller segments, and 3) encoding the data into an image-like format.

1) Data Reduction:
A key limitation in the seizure prediction problem is the large data size. The data used had a total of 5,612 iEEG clips (4827 interictal and 785 preictal); each clip is a 10 minutes duration, and given the high sampling rate of 400Hz, each iEEG clip had 240,000 data-points (10min$\times$60sec$\times$400Hz). The total size of the three patients' iEEG data was around 120GB. Reducing the data size can significantly reduce the required memory and computational resources.

As discussed in Section II, PCA is not suitable for dimensionality reduction of human iEEG data. We use a simple alternative for data reduction by resampling the iEEG clips at 100Hz rather than 400Hz. Figure 11 shows an example of the frequency spectra of interictal and preictal iEEGs recorded by the 16-channel seizure advisory system in Patient 1. It can be noticed that, for both interictal and preictal iEEGs, a significant amount of iEEG information is concentrated in the low-frequency band ($<$50Hz), where there is no much relevant information in the high-frequency band ($>$50Hz). Following Nyquist rule, we thus resample the 10-minutes iEEG clips at 100Hz ($2\times$50Hz) instead of 400Hz. This reduces the dimensionality of the data by a factor of 4. The resulting number of data-points in an iEEG clip is thus 60,000 (10min$\times$60sec$\times$100Hz).

Figure 12 demonstrates how the down-sampling process affects iEEG signals in both time and frequency domains. Figure 12(a) shows an example of time-series preictal iEEG signal and Fig. 12(b) depicts its spectrum along the entire frequency range of 0-200Hz. Figure 12(c) shows the down-sampled (by a factor of 4) version of the iEEG signal in Fig. 12(a). The frequency spectrum of the down-sampled signal is shown in Fig. 12(d). It can be noticed in the figure that the power spectrum of the signal is concentrated in the
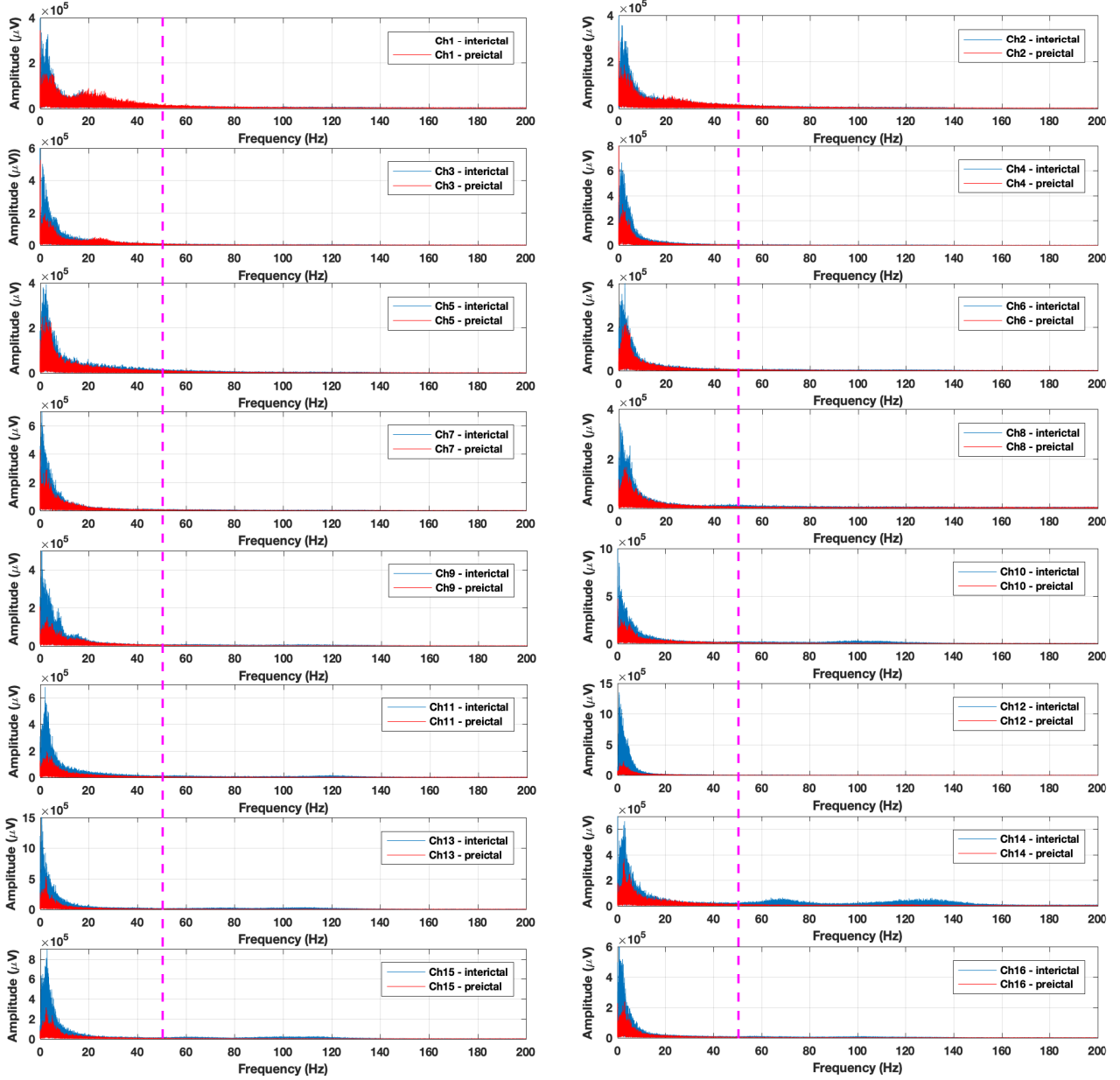
Figure 11. Frequency spectra of interictal (blue) and preictal (red) iEEG signals collected by the 16 implanted electrodes (channels) of the seizure advisory system in Patient 1. Ch1-Ch16 stand for Channels 1 to 16.

frequency band of 0-50Hz. It is important to note that, in some cases, this down-sampling process may hurt the prediction performance, because some of the iEEG samples have energy beyond 50Hz.

2) iEEG Segmentation:
Human surface and intracranial EEGs are usually non-stationary signals, *i.e.*, their statistical characteristics change over time [20]. The main intuition behind iEEG segmentation is to divide a non-stationary iEEG signal into several pseudo-stationary segments as those are expected to have comparable features [21]. Another advantage of segmentation is the output large number of labeled iEEG samples needed for training deep CNNs to improve their performance. In this work, each down-sampled 10-minute iEEG clip is further split into 10 non-overlapping 1-minute iEEG segments; each has 6,000

data-points. This results in a tenfold increase in the total number of training (interictal and preictal) iEEG samples.

3) Mapping Time-series iEEG Data into Images:
Since CNN has achieved great success in computer vision tasks, we are motivated to convert time-series iEEG data into an image-like format. We propose the use of Short-Time Fourier Transform (STFT) to obtain a two-dimensional representation of each iEEG segment. The (6,000 data-points) 1-minute iEEG segments are first partitioned into shorter chunks of equal length, and then the Fourier transform is computed for each individual chunk. This eventually reveals the changing power spectra as a function of time and frequency. Figure 13 shows the three-dimensional spectrogram of a 1-minute preictal iEEG segment. It can be noticed that the iEEG signal power decays along the frequency axis starting with
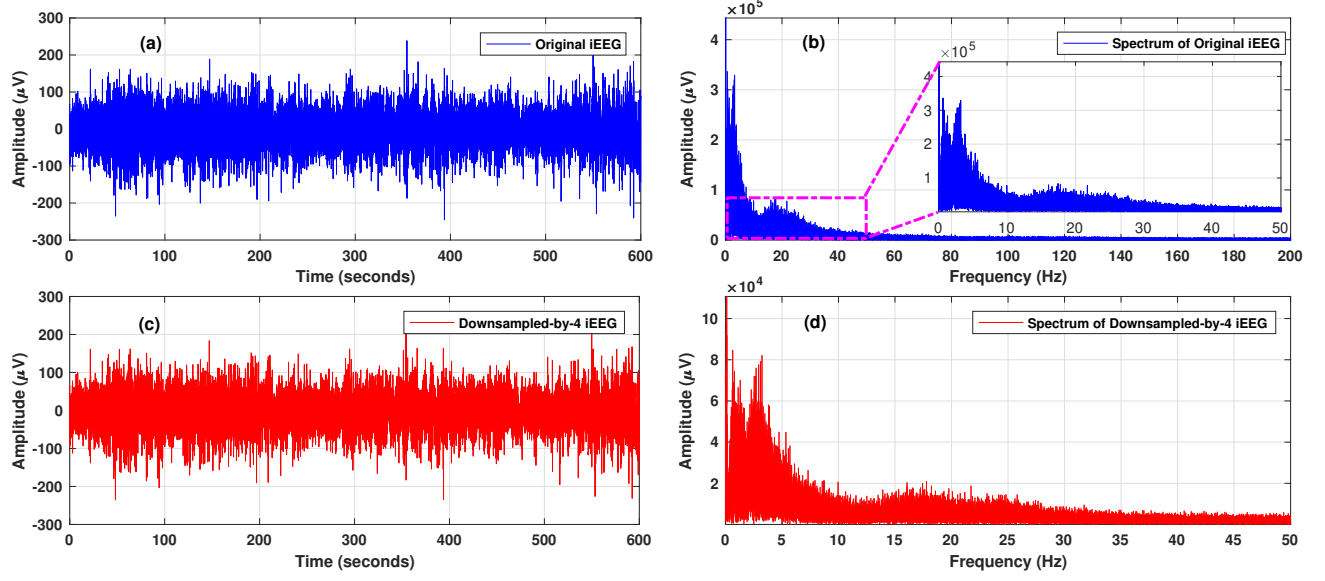
Figure 12. Time-series iEEG signals and their corresponding spectra: (a) and (b) original iEEG clip and its 200Hz frequency spectrum; (c) and (d) downsampled-by-4 iEEG clip and its 50Hz frequency spectrum.
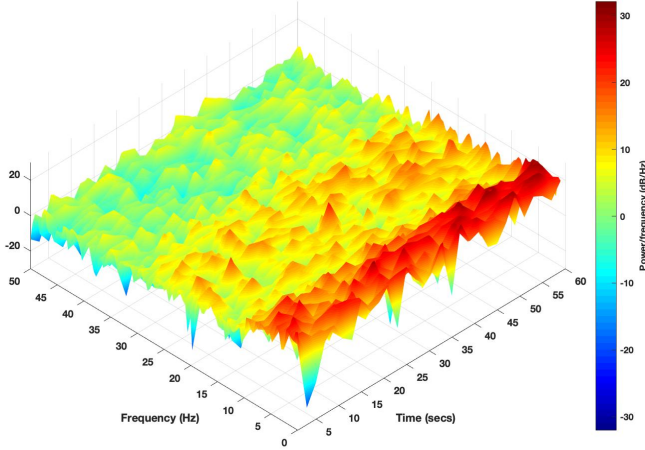

Figure 13. 3D Spectrogram of a 1-minute preictal iEEG segment.

30dB at 0-5Hz and till -30dB at 40-50Hz.

After iEEG downsampling and segmentation, the resulting dimension of the data is $10N \times 16 \times 6,000$; where N is the total number of 10-minute iEEG training clips, 10 is the number of iEEG segments per clip, 16 is the number of iEEG channels, and 6,000 is the length of each iEEG segment. For each iEEG segment, we compute the STFT for the 16 channels separately. The resulting dimension of the data is thus $10N \times 16 \times 129 \times 26$; where 129 and 26 are the numbers of frequencies and time chunks, respectively.

Figure 14 illustrates all the transformations applied to the raw iEEG data before feeding it to the seizure prediction algorithm. The final output is STFTs of 1 minute duration and 50Hz bandwidth. To standardize the training data across all 16 channels, the mean and standard deviation were computed for each STFT image in every channel and then used to standardize the test set prior to their classification.

B. Automatic Feature Learning of iEEG.

Figure 15 depicts the detailed architecture of the proposed CNN-based seizure prediction method. As explained in the previous subsection, the input data takes the shape of $10N \times 16 \times 129 \times 26$; where $10N$ is the total number of iEEG samples. Each sample comprises 16 STFT images; one for each iEEG channel. These images are first fed into a combination of CNNs of the same filter size (1×1) with their output supplied to other CNNs of bigger filter size (3×3 and 5×5). Additionally, since pooling operations have been proven to play a key part in the promising convolutional networks, we add a parallel path in which maximum pooling and convolution operations have been adopted. Ultimately, the outputs of all parallel paths are concatenated into a single feature vector forming the input of the next few layers. The concatenated output is flattened and presented as an input to two Fully Connected (FC) layers. Finally, a Sigmoid function is used to compute the label probabilities and predictions [22].

As for the proposed architecture, we were inspired by the successful CNN architectures in computer vision tasks. The goal here is to learn different feature maps and then combine them together so that the following FC layers can learn discriminative features from different scales simultaneously. The proposed CNN model was trained using 50,470 iEEG spectrograms (43,140 interictal and 7,330 preictal), and its seizure prediction performance was tested using 5,830 iEEG spectrograms (5,310 interictal and 520 preictal)[2]. To make

---

[2]Our experiments on the human iEEG-based seizure prediction using multi-scale CNNs were conducted with the open-source software of Keras using TensorFlow backend [23].
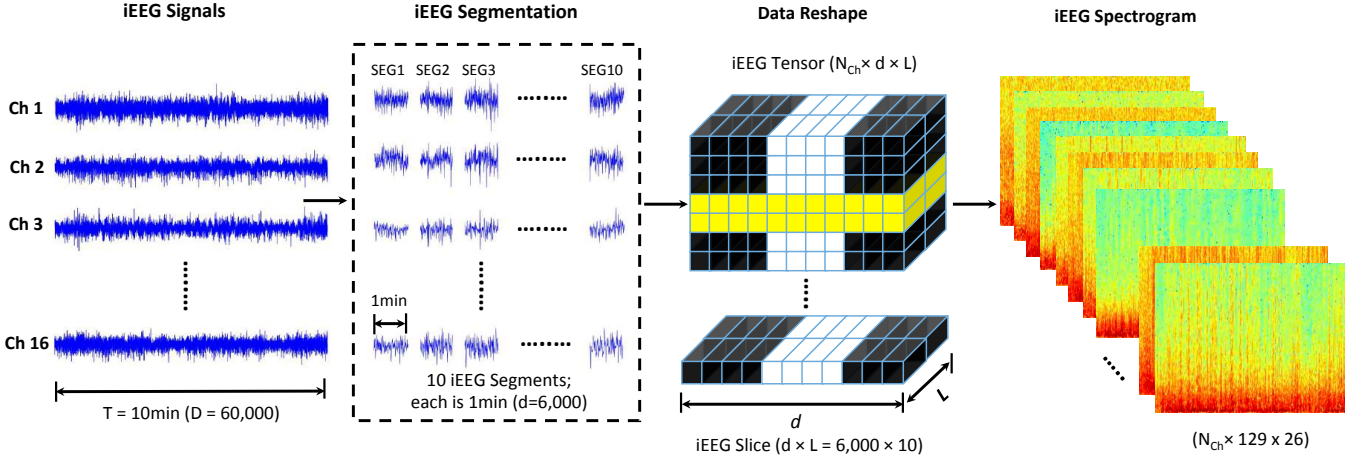
Figure 14. Schematic pipeline of the proposed iEEG data pre-processing approach for epileptic seizure prediction: SEG1, SEG2, $\cdots$, SEG10 are corresponding to the 1st, 2nd, and 10th iEEG segments of each iEEG channel signal; N_{Ch} is the total number of iEEG channels (N_{Ch}=16), L is the number of segments per iEEG clip (L=10), and d is the number of data-points in each iEEG segment (d=6,000).
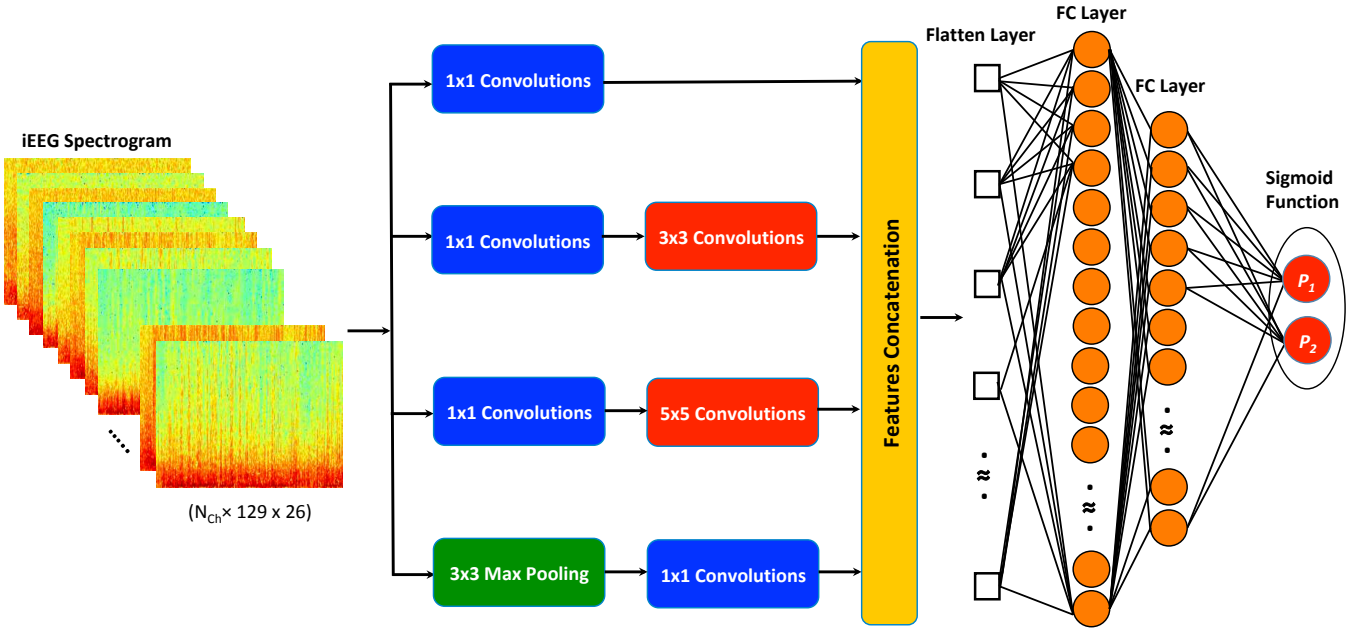

Figure 15. Schematic diagram of the proposed neural network architecture for epileptic seizure prediction: Each iEEG instance is of $N_{Ch}\times129\times26$ (N_{Ch}=16); Max Pooling stands for Maximum Pooling; FC layer stands for Fully Connected layer; P_1 and P_2 are the probabilities produced by the sigmoid function got class 1 and 2, respectively.

a single prediction for each 10-minute iEEG clip, we take the maximum probability over the past consecutive ten 1-minute predictions. The arithmetic mean was also computed for every consecutive ten predictions, however, it achieved inferior prediction results than those of the maximum.

From the network configuration perspective, our CNN model was trained by optimizing the "binary cross-entropy" cost function with "Adam" parameter update and a learning rate of 0.001. The 1×1, 3×3, and 5×5 convolutional networks have 64 units each. The number of units for the two FC layers was set to 124 and 64, respectively. Our implementation was derived in Python using Keras with TensorFlow backend and the training took eight hours on an NVIDIA K40 GPU machine.

C. Performance Evaluation.

Metrics used to assess the prediction performance of the proposed method are sensitivity (SEN) and area under the ROC curve (AUC). To examine the generalizability of our seizure prediction algorithm over different subjects, we first evaluate the performance metrics for the three patients individually and then report the average performance.

IV. RESULTS AND DISCUSSION

In this section, we first assess the seizure prediction performance of the proposed algorithm and compare it to the

state-of-the-art methods when tested on the same benchmark iEEG dataset [10]. We also provide four intuitive reasons why current seizure prediction algorithms achieve limited performance.

A. Seizure Prediction Results

In this section, we test our seizure prediction algorithm and compare the results with those of four recent studies (reported in [10], [24], [25], and [26]) that were tested on the same human iEEG dataset. In [10], Cook *et al.* implanted the first-in-man Seizure Advisory System (SAS) in patients with refractory epilepsy. After SAS implantation, a seizure prediction algorithm was developed to distinguish time intervals of low, medium, and high likelihood of upcoming seizure attacks. Satisfactory seizure prediction results were achieved for most of the patients, proving that seizure prediction was possible. For example, the average seizure prediction sensitivity for all patients was around 61.20%, while the three patients under study had inferior seizure prediction sensitivities with an average of 33.67%. The main reason behind such low seizure prediction performance was the temporal drift (variation) observed in the adopted time-dependent iEEG features (this study did not report what kind of iEEG features were extracted and what classifier was used to evaluate their effectiveness) [10].

In [24], Karoly *et al.* developed a detection method to identify probable seizure activities. They used the preictal spike rate as an indicator that the brain is approaching a seizure. However, the spike rate was found to be an unreliable biomarker for all patients, as it was shown to increase before seizures for 9 out of 15 patients and decrease for the rest of the patients. The classification of iEEG clips was performed manually and the results showed an average seizure prediction sensitivity of 66.54% for all patients and 43.34% for the three patients under study. In [25], the same authors proposed a circadian seizure forecasting method to achieve robust prediction performance across all patients. With the help of the adopted circadian information, a significant improvement in the predictive power of seizure prediction methods was observed for most of the patients. The logistic regression classifier was used to assess the prediction performance of the proposed algorithm. Average seizure prediction sensitivities of 62.10% and 52.67% were achieved for all patients and for the three patients under study, respectively. In [26], Kiral-Kornek *et al.* used deep learning to build an accurate seizure warning system that is patient-specific and can be adjusted to meet patients' needs. The deployed algorithm manifested notable seizure prediction results for all patients; yielding an average sensitivity of 69.00% for the 15 patients and 77.36% for the three patients whose data is studied in this work.

It is worth noting that the seizure prediction algorithms presented in [10], [24], [25], and [26] could not achieve adequate prediction sensitivity for Patients 3, 9, and 11. These are the same as Patients 1, 2, and 3 in our study. The potential reasons for such poor performance are given in the following subsection. In our work, a multi-scale CNN architecture was specifically built to accurately recognize pre-seizure patterns in the preictal iEEG data taken from those three patients.

Table II summarizes the seizure prediction results achieved by the proposed and the state-of-the-art methods. It can be noticed that, for the three patients, our algorithm yields superior seizure prediction sensitivity rates. It is worth mentioning that the seizure prediction performance is affected by the number of training iEEG samples for individual patients. For instance, Patient 3 has the largest number of training samples (see Table I), which in turn helps our CNN algorithm achieve a notably high seizure prediction sensitivity of 91.84%. Nevertheless, a lower prediction sensitivity of 82.92% was achieved for Patient 1 which had the least number of training samples.

As explained in Section III, all the iEEG signals are downsampled by a decimation factor of 4 for the purpose of speeding up the deployed deep learning algorithm. The question that may arise here: Does the downsampled-by-4 data closely represent the original data? Resampling the data at 100Hz means that all iEEG signals' content above 50Hz would be discarded. After careful examination of the frequency spectra of a wide range of interictal and preictal iEEG signals, we have found the following: (i) for most of the iEEG signals, the signal power dwells in the low-frequency band of 0-50Hz (as shown in Fig. 11), and (ii) for a few iEEG signals, a non-trivial amount of the signal power resides in higher frequency bands (>50Hz), implying that downsampling the data at 100Hz entails a considerable loss of information.

In an attempt to retain as much iEEG information as possible, the original iEEG signals are downsampled by a factor of 2 (*i.e.*, f_s=200Hz). The proposed seizure prediction algorithm is then tested on the data after downsampling it by a factor of 2. The achieved seizure prediction sensitivities are reported in Table II. The results demonstrate that downsampling the data at 200Hz helps preserve the iEEG signal information below 100Hz, and hence improve the overall seizure prediction sensitivity to 87.85%.

In addition, we compare our results to those of the winning solutions in the 2016 Kaggle seizure prediction challenge. The receiver operating characteristic (ROC) area under the curve (AUC) was the metric used for ranking various solutions. The winning team deployed 11 different classification models and more than 3000 hand-crafted iEEG features, and achieved an average AUC of 0.854 (see Table III) [11], [12]. It is, however, impractical to use such a computationally-intensive process for real-time applications. Our algorithm obtains a comparable but slightly lower seizure prediction AUC score of 0.840 on the iEEG data downsampled by 2. It is, however, much faster in obtaining the results and is thus more suitable for use in ambulatory and clinical applications. For faster iEEG processing (feature extraction and classification), the proposed algorithm is also tested on the downsampled by 4 iEEG data and an average AUC score of 0.787 is achieved.

Table IV shows the AUC scores of our algorithm for individual patients when tested on iEEG data downsampled by 4 and 2. It can be seen that, for the downsampled-by-4 data, the AUC scores of 0.822 and 0.865 are achieved for Patients 2 and 3, respectively, while Patient 1 experiences a lower AUC score of 0.675. Using the downsampled-by-2 data helped improve the AUC scores for all patients. For example, a remarkable AUC score of 0.938 was achieved for Patient 3

Table II
SEIZURE PREDICTION SENSITIVITY SCORES OF THE ALGORITHMS UNDER STUDY.

Subject	Cook *et al.* [10] (2013)	Karoly *et al.* [24] (2016)	Karoly *et al.* [25] (2017)	Kiral-Kornek *et al.* [26] (2018)	Proposed Method (DS-by-4)	Proposed Method (DS-by-2)
Patient 1	45.00	56.00	55.00	71·10	**82.92**	79.65
Patient 2	17.00	18.00	45.00	83·10	88.52	**91.86**
Patient 3	39.00	56.00	58.00	77·90	91.84	**92.05**
Average	33.67	43.34	52.67	77.36	87.76	**87.85**

Patient 1, 2, and 3 in the Kaggle compeition dataset are the same as Patients 3, 9, and 11 in [10], [24], [25], and [26]. DS = downsampled. DS-by-4 and DS-by-2 indicate the seizure prediction results using the downsampled-by-4 and downsampled-by-2 iEEG data, respectively.

Table III
AUC SCORES FOR THE PROPOSED METHOD AND KAGGLE TOP FINISHING CONTESTANTS ON THE PUBLIC TEST SET.

Top 10 Kaggle Scores (2018) [11] min-max	Proposed Method (DS-by-4)	Proposed Method (DS-by-2)
0.820-0.854	0.787	0.840

The first column shows the minimum and maximum AUC scores of the 10 top-scoring solutions of a post-competition study [11]. The second and third columns show the AUC score of our multi-scale CNN model on the iEEG data downsampled by 4 and 2, respectively.

Table IV
PER-SUBJECT AUC SCORES OF THE PROPOSED METHOD.

Subject	Proposed Method (DS-by-4)	Proposed Method (DS-by-2)
Patient 1	0.675	0.692
Patient 2	0.822	0.891
Patient 3	0.865	0.938
Average	0.787	0.840

Figure 16. Examples of corrupted iEEG clips for Patient 1.

(as this patient had the largest number of interictal and preictal iEEG samples). Patient 2 had a little lower AUC score of 0.891 and Patient 1 had the least AUC score of 0.692. The average AUC score produced by our CNN algorithm for all patients is 0.840.

B. Possible Reasons for Limited Performance

Seizure prediction algorithms examined on the 2016 Kaggle/Melbourne University iEEG dataset provide limited performance. The top contestant, for example, achieved the highest AUC of 0.854 [11], [12]. Herein, we list some potential reasons why existing machine learning and deep learning algorithms could not achieve any better performance.

1) iEEG Data Corruption:
After a thorough inspection and data visualization of the iEEG data under study, we realized that many of the 10-minutes iEEG clips contain "data drop-out". This is when the implanted device failed to communicate with the storage device for several possible reasons. This data drop-out corresponds to iEEG signal values of zeros across all channels at a given time sample. A handful of 10-minute clips comprise 100% data drop-out and cannot be classified. Other clips, however, are partially corrupted and contain different percentages of data drop-out. Fig. 16 displays various examples of iEEG signals comprising data drop-outs at different time samples.
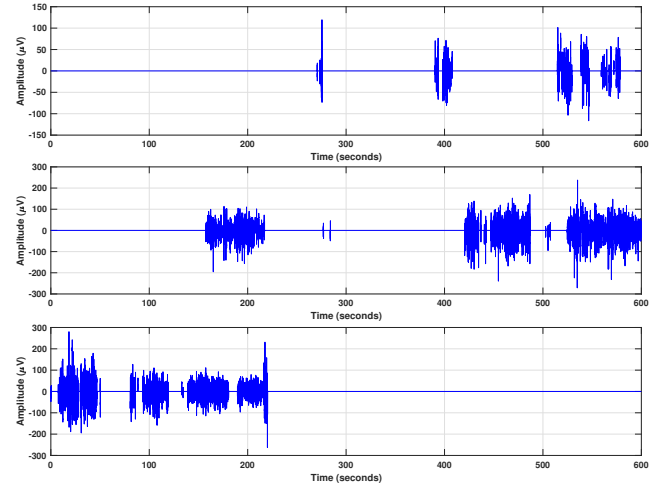
It is worth noting that, if for any reason, a dataset includes corrupted instances, standard machine learning algorithms can break down.

2) iEEG Data Mismatch:
The terms "data mismatch", "concept drift", and "covariate shift" have been used to refer to the situation where data characteristics (distribution) change over time [27]. 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 just the same as the distribution of the new data. This holds true for many machine learning problems, but not 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 attentive screening to the characteristics of the iEEG data under study, we realized that, for each individual patient, the distribution of either interictal or preictal iEEG data is different between the training set and testing set. A potential reason for such a data shift is that the testing data was recorded a long time after recording the training data. During this time, the patient may be positively or negatively influenced by the medications. Fig. 17 depicts how the preictal iEEG data distributions for Patient 1 have changed over time. The top plot explains how the density of preictal iEEG signals recorded by Channel 1 differs between training and testing sets. Similarly, the bottom plot demonstrates the variations in preictal iEEG density for
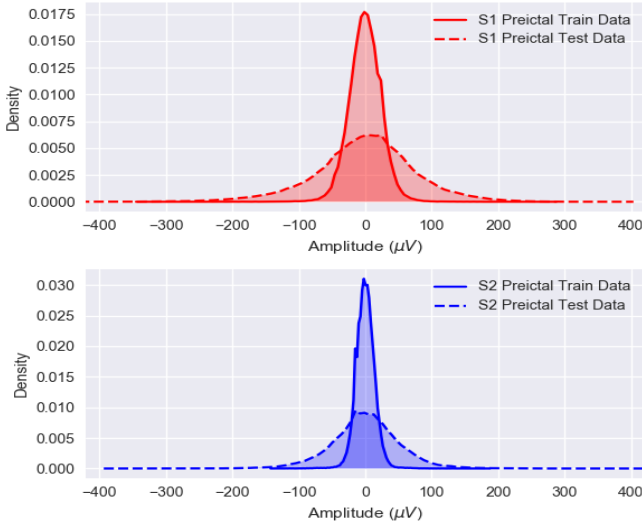
Figure 17. Data mismatch in Patient 1's preictal iEEG sensor data.

Figure 18. Scatter plot to identify outliers in preictal iEEG data of Patient 1: Preictal S1 readings vs. Preictal S2 readings.

Channel 2.

3) Outliers in iEEG Data:

Since iEEGs are recorded with electrodes attached directly to the surface of the brain, they are deemed to be free of body movement artifacts (*e.g.*, muscle activities and eye blinking) and their signal-to-noise ratios are expected to be much higher than those of surface EEGs [28]. However, there exist some inevitable sources of artifacts (*e.g.*, static magnetic field and environmental noise) that interfere with the iEEG signals producing undesired outliers. The scatter plot in Fig. 18 shows an example of "data outliers" which were detected in the S1 and S2 readings for some preictal iEEG clips taken from Patient 1. Other scatter plots (omitted for the lack of space) demonstrate that, for the same iEEG clips, the remaining 14 sensor readings also include similar outliers. The presence of outliers often has a negative influence on the extracted features and certainly on the predictive power of the fitted models. If the outliers are removed from the fitting process, then the resulting fit will maintain subtle performance every time it is tested on new unseen data.

4) Imbalanced Class Distribution:

The seizure prediction dataset under study has 5,047 iEEG samples. A total of 4,314 belongs to Class-1 (interictal iEEG) and the remaining 733 samples belong to Class-2 (preictal iEEG). This represents a binary classification problem with imbalanced class distribution, where the ratio of Class-1 to Class-2 samples is 4,314:733, *i.e.*, 5.885:1. With such an imbalanced dataset, the classification rules that predict the minority class tend to be more flimsy than those that predict the majority class; therefore, test samples belonging to the minority class are misclassified more often than those belonging to the majority class [29]. Thus, one of the possible reasons for limited seizure prediction performance is that the deployed machine learning algorithms do not take into consideration the interictal and preictal class distributions. They pay less attention to the minority class of preictal iEEGs, which is of crucial importance for forecasting impending seizures.
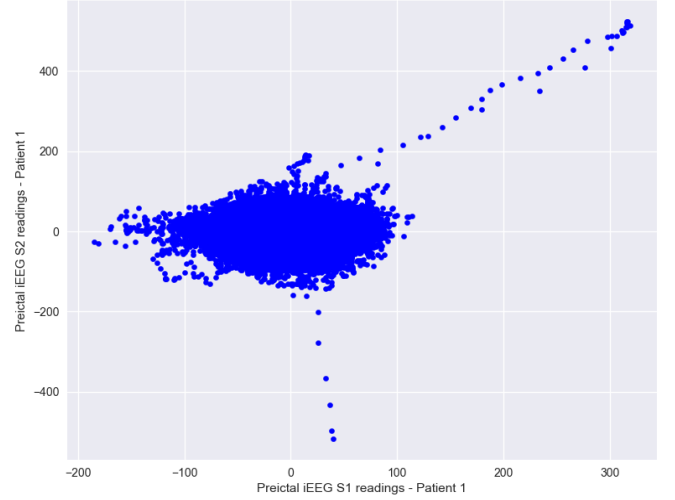
There exist many approaches that can tackle classification problems having imbalanced data. One of the effective procedures is to generate synthetic samples from the under-represented class. In this study, we have first used the Synthetic Minority Over-sampling Technique (SMOTE) [30] to generate more samples from the minority class (preictal iEEG). However, a serious degradation in the seizure prediction performance was noticed (as the trained CNN model was overfitting). As an alternative solution, we employed the class weights in training the proposed CNN model in order to force the model to treat every sample of preictal iEEG as ~6 samples of interictal iEEG. This helped achieve the remarkable seizure prediction performance reported in Tables II, III, and IV.

V. CONCLUSION

Big data is crucial for reliable seizure prediction. The Kaggle/Melbourne University Seizure Prediction Competition provides a big dataset. It includes a long-term human intracranial EEG data recorded continuously from three (drug-resistant) epileptic patients for over 373-559 days. Leveraging this dataset properly can lead to substantial improvements in the field of epileptic seizure prediction.

We have therefore done an extensive analysis of the iEEG data, and from which we have provided, for the first time, a detailed quantitative analysis of the human iEEG data during the preictal and interictal brain states. This analysis reveals why hand-crafted iEEG features that have been successfully used in seizure detection are not the best candidates for robust prediction of epileptic seizures. We then proposed a convolution neural network (CNN) model for accurate prediction of epileptic seizures. The data was first divided into smaller segments on which we applied STFT. This resulted in a much larger number of interictal and preictal iEEG samples, which is important in deep learning in achieving better seizure prediction performance. The resulting time-series segments were converted into image-like formats to enable its use as inputs to

our CNN model. The CNN model we proposed has an efficient multi-scale CNN architecture that automatically learns distinctive iEEG features from different spatial scales contemporaneously. Unlike previous seizure prediction algorithms that used hand-crafted features, the proposed algorithm can yield reliable performances across different patients, achieving a superior average prediction sensitivity performance as 87.85%.

Though training the proposed CNN model takes 6-8 hours, the trained model takes less than a second to test new unseen data. Compared to the Kaggle leaderboard winning solution that uses over 3,000 custom features, the proposed algorithm achieves comparable performances in a much more computationally-efficient fashion.

REFERENCES

[1] G. Rogers, "Epilepsy: the facts," *Primary Health Care Research & Development*, vol. 11, no. 4, p. 413, 2010.

[2] U. R. Acharya, S. V. Sree, G. Swapna, R. J. Martis, and J. S. Suri, "Automated EEG analysis of epilepsy: a review," *Knowledge-Based Systems*, vol. 45, pp. 147–165, 2013.

[3] J. A. French, "Refractory epilepsy: clinical overview," *Epilepsia*, vol. 48, pp. 3–7, 2007.

[4] M. D'Alessandro, R. Esteller, G. Vachtsevanos, A. Hinson, J. Echauz, and B. Litt, "Epileptic seizure prediction using hybrid feature selection over multiple intracranial eeg electrode contacts: a report of four patients," *IEEE transactions on biomedical engineering*, vol. 50, no. 5, pp. 603–615, 2003.

[5] J. F. Téllez-Zenteno, L. H. Ronquillo, F. Moien-Afshari, and S. Wiebe, "Surgical outcomes in lesional and non-lesional epilepsy: a systematic review and meta-analysis," *Epilepsy research*, vol. 89, no. 2-3, pp. 310–318, 2010.

[6] K. Gadhoumi, J.-M. Lina, F. Mormann, and J. Gotman, "Seizure prediction for therapeutic devices: A review," *Journal of neuroscience methods*, vol. 260, pp. 270–282, 2016.

[7] E. B. Assi, D. K. Nguyen, S. Rihana, and M. Sawan, "Towards accurate prediction of epileptic seizures: A review," *Biomedical Signal Processing and Control*, vol. 34, pp. 144–157, 2017.

[8] F. Mormann, R. G. Andrzejak, C. E. Elger, and K. Lehnertz, "Seizure prediction: the long and winding road," *Brain*, vol. 130, no. 2, pp. 314–333, 2006.

[9] L. Kuhlmann, D. Freestone, A. Lai, A. N. Burkitt, K. Fuller, D. B. Grayden, L. Seiderer, S. Vogrin, I. M. Mareels, and M. J. Cook, "Patient-specific bivariate-synchrony-based seizure prediction for short prediction horizons," *Epilepsy research*, vol. 91, no. 2-3, pp. 214–231, 2010.

[10] M. J. Cook, T. J. O'Brien, S. F. Berkovic, M. Murphy, A. Morokoff, G. Fabinyi, W. D'Souza, R. Yerra, J. Archer, L. Litewka *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*, vol. 12, no. 6, pp. 563–571, 2013.

[11] L. Kuhlmann, P. Karoly, D. R. Freestone, B. H. Brinkmann, A. Temko, A. Barachant, F. Li, G. Titericz Jr, B. W. Lang, D. Lavery *et al.*, "Epilepsyecosystem. org: crowd-sourcing reproducible seizure prediction with long-term human intracranial eeg," *Brain*, vol. 141, no. 9, pp. 2619–2630, 2018.

[12] A. Barachant, A. Temko, F. Li, and G. Titericz Junior, "Code and documentation for the winning solution to the melbourne university aes/mathworks/nih seizure prediction challenge on kaggle," December 2016. [Online]. Available: https://github.com/alexandrebarachant/kaggle-seizure-prediction-challenge-2016

[13] L. Wang, W. Xue, Y. Li, M. Luo, J. Huang, W. Cui, and C. Huang, "Automatic epileptic seizure detection in eeg signals using multi-domain feature extraction and nonlinear analysis," *Entropy*, vol. 19, no. 6, p. 222, 2017.

[14] V. Shah, M. Golmohammadi, S. Ziyabari, E. Von Weltin, I. Obeid, and J. Picone, "Optimizing channel selection for seizure detection," in *Signal Processing in Medicine and Biology Symposium (SPMB), 2017 IEEE*. IEEE, 2017, pp. 1–5.

[15] M. A. Hall, "Correlation-based feature selection for machine learning," 1999.

[16] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "Deepface: Closing the gap to human-level performance in face verification," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 1701–1708.

[17] D. Erhan, C. Szegedy, A. Toshev, and D. Anguelov, "Scalable object detection using deep neural networks," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2014.

[18] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[19] A. Graves, A.-r. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," in *Acoustics, speech and signal processing (icassp), 2013 ieee international conference on*. IEEE, 2013, pp. 6645–6649.

[20] H. Azami, K. Mohammadi, and H. Hassanpour, "An improved signal segmentation method using genetic algorithm," *International Journal of Computer Applications*, vol. 29, no. 8, pp. 5–9, 2011.

[21] H. Hassanpour and M. Shahiri, "Adaptive segmentation using wavelet transform," in *Electrical Engineering, 2007. ICEE'07. International Conference on*. IEEE, 2007, pp. 1–5.

[22] J. Friedman, T. Hastie, and R. Tibshirani, "Regularization paths for generalized linear models via coordinate descent," *Journal of statistical software*, vol. 33, no. 1, p. 1, 2010.

[23] R. Hussein, *Multi-scale CNNs for Epileptic Seizure Prediction*, August 2018. [Online]. Available: https://github.com/ramyh/Multiscale-CNNs-for-Epileptic-Seizure-Prediction.git

[24] P. J. Karoly, D. R. Freestone, R. Boston, D. B. Grayden, D. Himes, K. Leyde, U. Seneviratne, S. Berkovic, T. O'brien, and M. J. Cook, "Interictal spikes and epileptic seizures: their relationship and underlying rhythmicity," *Brain*, vol. 139, no. 4, pp. 1066–1078, 2016.

[25] P. J. Karoly, H. Ung, D. B. Grayden, L. Kuhlmann, K. Leyde, M. J. Cook, and D. R. Freestone, "The circadian profile of epilepsy improves seizure forecasting," *Brain*, vol. 140, no. 8, pp. 2169–2182, 2017.

[26] I. Kiral-Kornek, S. Roy, E. Nurse, B. Mashford, P. Karoly, T. Carroll, D. Payne, S. Saha, S. Baldassano, T. O'Brien *et al.*, "Epileptic seizure prediction using big data and deep learning: toward a mobile system," *EBioMedicine*, vol. 27, pp. 103–111, 2018.

[27] I. Žliobaitė, M. Pechenizkiy, and J. Gama, "An overview of concept drift applications," in *Big data analysis: new algorithms for a new society*. Springer, 2016, pp. 91–114.

[28] P. L. Nunez and R. Srinivasan, "Scale and frequency chauvinism in brain dynamics: too much emphasis on gamma band oscillations," *Brain Structure and Function*, vol. 215, no. 2, pp. 67–71, Dec 2010. [Online]. Available: https://doi.org/10.1007/s00429-010-0277-6

[29] Y. Sun, M. S. Kamel, and Y. Wang, "Boosting for learning multiple classes with imbalanced class distribution," in *Sixth International Conference on Data Mining (ICDM'06)*. IEEE, 2006, pp. 592–602.

[30] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "Smote: synthetic minority over-sampling technique," *Journal of artificial intelligence research*, vol. 16, pp. 321–357, 2002.