Seizure Detection Challenge
The Fitzgerald team solution

Vincent Adam, Joana Soldado-Magraner, Wittawat Jitkritum,
Heiko Strathmann, Balaji Lakshminarayanan, Alessandro Davide Ialongo,
Gergő Bohner, Ben Dongsung Huh, Lea Goetz,
Shaun Dowling, Iulian Vlad Serban, Matthieu Louis

1 Introduction

Accurate early seizure detection is key to building effective devices to treat epilepsy. Intracranial electroencephalography (iEEG) provides a signal rich enough to both detect and predict epileptic seizures online when coupled with efficient computational algorithms. As part of a *hackathon*, graduate students of the Gatsby Computational Neuroscience unit (UCL, London) took part in the seizure Detection Challenge [1] hosted by the machine learning competition organizer Kaggle. We ranked 9/205 and here describe the method we used. Furthermore, we provide a more in depth analysis of our solution based on a set of experiments we carried out after the competition.

In this project, we applied machine learning algorithms to detect seizure occurrence using training datasets collected from both dog and human subjects with naturally occurring epilepsy. Specifically, we aimed to classify small iEEG segments as ictal or interictal events, and further, assess whether the ictal segments occurred early into the seizure.

2 Data and Task

Data

Intracranial EEG data analyzed for this study was provided by the UPenn - Mayo Clinic Seizure Detection Challenge on kaggle.com, sponsored by the American Epilepsy Society, and the data is available via the NIH-sponsored International Epilepsy Electrophysiology Portal (ieeg.org).

The data consisted in 1s long multi-channel intracranial EEG recordings (segments) from 4 dogs and 8 human patients. Each patient had a fixed number of implanted electrodes from which signal was recorded ($N_{channel}$) at a fixed sampling rate f_s. No information was given about the localization of the electrodes.

Labels

Training data was manually classified as ictal or interictal. Ictal segments belonged to episodes lasting multiple seconds. For those ictal segments, time since episode onset was given. A segment was labeled early if its time since onset was below 15s, with onset also manually determined by human experts.

Classification Tasks

The task was to answer the following questions for each test segment:

- is it a seizure segment?
- is it an early seizure segment? (early being defined as time from onset of seizure episode below 15sec.)

3 Method

We proceeded in 3 steps from the raw signal to the classification output: (1) preprocessing, (2) feature extraction, (3) classification. Our approach follows closely that of [4].

3.1 Preprocessing

Signal was downsampled to the divisor of the initial sampling rate f_s the closest to the target sampling rate f_s^{target}. This downsampling also low pass filters the signal. The aim was to reduce dimensionality to speedup feature extraction with minimal loss of information. Electrical noise (United States) was then removed using a band stop filter.

Downsampling

We downsampled each signal by integer factor $k = \lfloor f_s / f_s^{target} \rfloor$.

Name	Value
Downsampling	
Target sampling rate f_s^{target}	$400Hz$
Electrical noise removal	
band cutoff $\left[f_{low}^{BP}, f_{high}^{BP}\right]$	$[59Hz, 61Hz]$
width of filter band w^{BP}	$3Hz$
attenuation in band ρ^{BP}	$60dB$
Low pass filtering	
cutoff f_c^{LP}	$400Hz$
width of filter w^{LP}	$30Hz$
attenuation in band ρ^{LP}	$60dB$

Table 1: Preprocessing Parameters

Bandpass filtering

Electrical noise was removed using a kaiser filter with following parameterization: cutoff frequencies : $\left[f_{low}^{BP}, f_{high}^{BP}\right]$, width: w^{BP}, attenuation in band: ρ^{BP}.

Low pass filtering

Finally a kaiser low pass filter was applied with the following parameterization: cutoff: f_c^{LP}, width: w^{LP}, attenuation: ρ^{LP}

3.2 Feature extraction

The aim of feature extraction is to obtain a set of parameters (feature vectors) which summarize the task-relevant statistics of the iEEG data. These feature vectors can then be used for classification, allowing to identify data segments believed to have originated during different brain regimes. In our case, these regimes correspond to the ictal and interictal periods.

We used several features based on single-channel and multichannel measures. Single-channel measures are motivated from the observation that marginally, stationary statistics of the signal coming from a single channel vary within different brain states. Multiple-channel measures focus on capturing interdependence variations between channels as state conditions change.

In our final solution, a combination (stack) of these type of features was used and fed the classifier.

3.2.1 Single-channel feature (univariate)

- **Spectral energy** (**SE**)

Rationale: energy content in different frequency bands changes between interictal and ictal periods.

Name	Value
Spectral Energy	
max frequency f_{max}	$100Hz$
number of bands N_{bands}	40
Vector autoregressive model	
lag τ	2

Table 2: Features Parameters

Method: the spectral energy of the signal - the squared modulus of the signal's Fourier transform-, was computed. Energy below f_{max} was averaged into N_{bands} contiguous bands of spectral width $\lfloor f_{max}/N_{bands} \rfloor$

3.2.2 Multiple-channel features (bivariate and multivariate)

- **Phase Locking Value** (**PLV**)[2]

Rationale: signals between channels become more synchronized during seizures. Phase Locking Value quantifies *locking* between the phases of the signals from two distinct electrodes.

Method: first, for each channel i, we extract the instantaneous phase $\phi_i^a(t)$ of the analytical signal $x_i^a(t)$ of the time series $x_i(t)$.
Then, for each pair (i, j) of channels, we compute the modulus of the time averaged phase difference mapped onto the unit circle

$PLV_{ij} = \left| \frac{1}{T} \sum_t e^{i(\phi_i^a(t) - \phi_j^a(t))} \right|$

- **Vector autoregressive** (**VAR**[τ]) **models** [3]

Rationale: signals from all channels become more synchronized and structured during seizures, being well described by a time-coupled multivariate linear system. This feature is therefore good for seizure detection.

Method: the signal was assumed to be generated from a vector autoregressive model

$x_t = \sum_{k=1}^{\tau} A_k x_{t-k} + \epsilon_t$, where $\epsilon \sim \mathcal{N}(0, Q)$

In this model, the i^{th}channel at time t depends on all the other channels withing $[t-\tau, t-1]$ through matrices $\{A_k\}_{k \in [t-\tau, t-1]}$. Once the model was fitted to the data, the learned parameters $[A_1, A_2, ..., A_\tau, Q]$ were taken as features.

Additional note: Relations between AR, VAR and SE features

AR, VAR and SE features are 2^{nd} order methods. They can all be seen as constructed from auto or cross correlations of the iEEG time series in the following way

- Spectral energy is related to auto-correlation under stationary assumptions through the Wiener-Khinchin Theorem which states that the spectral density of a stationary process is equal the the Fourier transform of its autocorrelation function.
- AR[τ] models of various lag τ parameterize spectral densities which richer expressive power as τ is increased. Parameter estimation through Yule-Walker equations shows that estimated parameters are linear functions of up to τ-lagged autocorrelations. Parameter estimation (residual variance excluded) is scale invariant.
- Parameter estimation in VAR[τ] model amounts to a linear combination of auto and cross correlation up to lag τ

In summary, AR fits are impoverished representations of the single channel spectrum (through its parameterization). VAR fits also impoverish the individual channel spectral density representation but, relative to SE features contain additional cross channel information.

3.3 Classification

Different classifiers were trained individually for each subject.

3.3.1 Random forest classifier

We used a Random forest classifier, parameterized by tree width (with N_{trees} number of trees). We used the scikit-learn implementation `RandomForestClassifier` (version 0.14.1). The rest of parameters were set to their default value.

Name	Value
Random forest	
number of trees N_{trees}	100

Metrics

We evaluate our classifiers using the Area Under the ROC Curve (AUC) as a metric. We report mean and standard deviation of the AUC score for each classifier.

4 Results

In order to get a better understanding of both the dataset and our algorithm, we ran a set of experiments beyond the reproduction of the averaged score of the kaggle competition.

Experiment 1: Reproducing the Kaggle results

We trained and tested our algorithm on the same dataset as the one used in the competition. However, we report the scores per task and per subject for each feature combination evaluated, instead of a unique aggregated score. The results are summarized in table 3 and figure 1.

Features	$AUC_{seizure}$	AUC_{early}
VAR(1)	0.939±0.072	0.893±0.093
VAR(2)	0.940±0.081	0.881±0.094
VAR(3)	0.916±0.117	0.864±0.119
VAR(4)	0.917±0.111	0.870±0.113
VAR(7)	0.909±0.101	0.859±0.086
SE	0.938±0.086	0.865±0.156
PLV	0.848±0.115	0.756±0.126
VAR(2)+PLV+SE	0.945±0.084	0.896±0.116
VAR(2)+SE	0.938±0.094	0.896±0.112
VAR(1)+VAR(2)+PLV+SE	0.949±0.073	0.904±0.106

Table 3: Mean performance and standard deviation across patients for both the seizure and early seizure detection tasks. Each row indicates the features/feature combinations tested.

Experiment 2: Classification performance as a function of latency

We reversed the training and test datasets in order to analyze the prediction confidence as a function of latency. The rationale here is to assess the extend to which different features pick up early signatures of seizures. The results of this analysis are exemplified in figure 2, which correspond to a single patient.

5 Discussion

We observe that features extracted with VAR models beyond 2nd order do not improve classification performance, as can be seen in figure 1, top and bottom plots. A combination of SE+VAR+PLV features achieves the best results in both tasks. Using this method, we ranked among the top 10 in the kaggle seizure detection competition [1].

We observe that iEEG data segments are more confidently classified as ictal as seizure progresses (see figure 2, left plot). Low confidence predictions in the first 10 seconds reveal the increasing difficulty of the detection task for segments closest to seizure onset. We also note that in both tasks, PLV features perform poorly regardless of latency. As expected (fig 2, right plot), classifier has lower confidence in telling apart early vs late for earliest segments and close to early/late boundary at 15s.

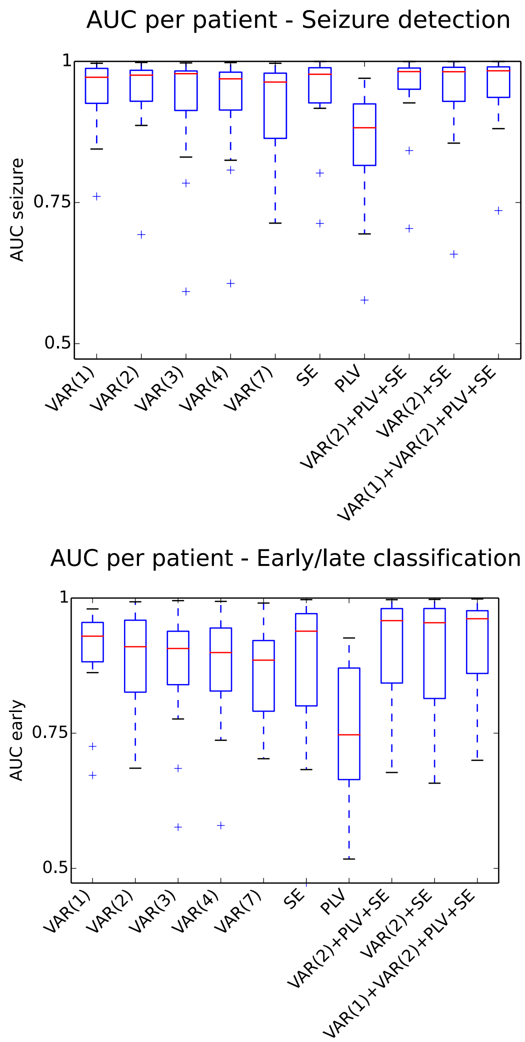

Figure 1: Box plot summarizing classification performance across all subjects. Top figure, seizure detection task. Bottom figure, early seizure detection task.

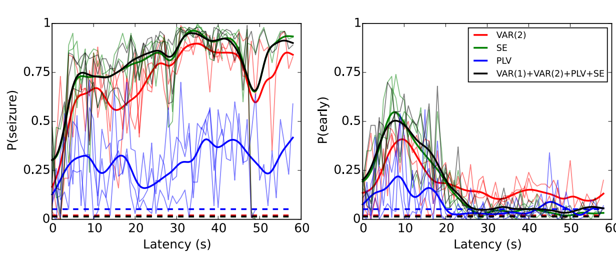

Figure 2: Estimated probabilities for seizure (left) and early seizure (right) as a function of seizure latency, for an example human patient. Thin lines: seizure prediction for individual ictal epochs. Thick lines: mean seizure prediction. Dashed lines: mean seizure prediction for interictal segments.

6 Software

All mathematical analyses were executed using the following Python libraries

- statsmodels (Vector Autoregressive model fitting)
- scikit.learn [5] (svm, random forest, cross-validation)
- scipy.signal (signal filtering, downsampling)
- numpy (fourier transform)

References

[1] Seizure detection challenge, June 2014.

[2] Jean-Philippe Lachaux, Eugenio Rodriguez, Jacques Martinerie, Francisco J Varela, et al. Measuring phase synchrony in brain signals. *Human brain mapping*, 8(4):194–208, 1999.

[3] H Lutkepohl. New introduction to multiple time series analysis. *Econometric theory*, 22(5):961–967, 2005.

[4] Piotr W Mirowski, Yann LeCun, Deepak Madhavan, and Ruben Kuzniecky. Comparing svm and convolutional networks for epileptic seizure prediction from intracranial eeg. In *Machine Learning for Signal Processing, 2008. MLSP 2008. IEEE Workshop on*, pages 244–249. IEEE, 2008.

[5] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

Features	$AUC_{seizure}$	AUC_{early}	AUC_{tot}
VAR(1)	0.952	0.893	0.922
VAR(2)	0.951	0.895	0.923
VAR(3)	0.947	0.895	0.921
VAR(4)	0.943	0.894	0.919
VAR(7)	0.939	0.892	0.916
SE	0.942	0.898	0.920
PLV	0.931	0.884	0.908
VAR(2)+PLV+SE	0.935	0.891	0.913
VAR(2)+SE	0.937	0.896	0.917
VAR(1)+VAR(2)+SE	0.939	0.900	0.919
VAR(1)+VAR(2)+PLV+SE	0.941	0.903	0.922
VAR(2)+VAR(4)+PLV+SE	0.942	0.905	0.924

Table 4: Kaggle competition results

Appendix

A. Reproduction of Kaggle competition results

In order to reproduce the kaggle competition results, we computed the AUC for the predicted scores ($P_{seizure}$, P_{early}) of all patients together. We chose the measure $AUC_{tot} = \frac{1}{2}(AUC_{seizure} + AUC_{early})$ to report overall performance, as used in the competition. The results are summarized in Table 4. We provide the overall scores for all the different feature combinations tested in this report, including the combination of features we used for the competition, $VAR(2) + PLV + SE$. Note that, due to lack of time, we were unable to reproduce exactly the score obtained in the competition, as we lost the details of the exact parameters we used. We expect to be able to obtain an even better performance once we carry out a more rigorous parameter search.

B. Summary of other methods

In this section we provide a brief summary of the methods used by other kaggle participants, only for comparison purposes.

Feature extraction

1st position (Michael Hills, winner)

Features were kept or discarded based on their cross-validation performance. Combinations of multiple features eventually proved to provide a better classification score once the right features were combined.

Three sources of features are used to form the whole feature-set:

1. log10(FFT magnitudes) in the low frequency range 1-47 Hz.
   This frequency range offered the best result compared to other frequency ranges. These features alone, combined with the selected classifier (Random Forest), offered already excellent performance. The resulting FFT magnitudes matrix has dimensions $\#_{channels} * 47$.
2. Correlation coefficients, CC, between iEEG channels.
   The FFT magnitudes matrix is first normalized across frequencies, column by column, subtracting the mean and dividing by the std. The correlation coefficient CC matrix ($\#_{channels} * \#_{channels}$) is then calculated for this normalized matrix, taking the upper triangular bit for the features.
   The same is performed using the time-series data, by computing the CC matrix from the original iEEG data matrix, which has dimensions $\#_{channels} * time$.
3. Eigenvalues .
   Eigenvalues are computed from the CC matrices, from both time and frequency domains. All real eigenvalues and the magnitude of the complex eigenvalues are taken as features. These values are then sorted by magnitude.

Classifier

1st position (Michael Hills, winner)

The chosen method for classification was selected with the help of the scikit-learn python machine learning library. Random Forest offered the best performance.