Seizure Prediction Challenge (#3)

Ryan Marshall, Bassel El Mabsout, Aditya Chechani, William Chapman
{ryanmars, bmabsout, adityac, wchapman}@bu.edu
BOSTON UNIVERSITY

Introduction

- Approximately 1.2% of the population is diagnosed with epilepsy, in which over excitability of neural tissue causes widespread activity and seizures.
- For drug-resistant epileptics, seizures must be controlled with a short-acting medication or electrical stimulation. However, such stimulation must be performed before onset of behavioral symptoms. Thus, early detection of oncoming seizures is critical.
- **Goal:** Categorize 10 minute samples as inter-ictal (not preceding a seizure in the next hour), or pre-ictal (seizure within one hour)
- **Performance** is evaluated by the area-under-curve (AUC)

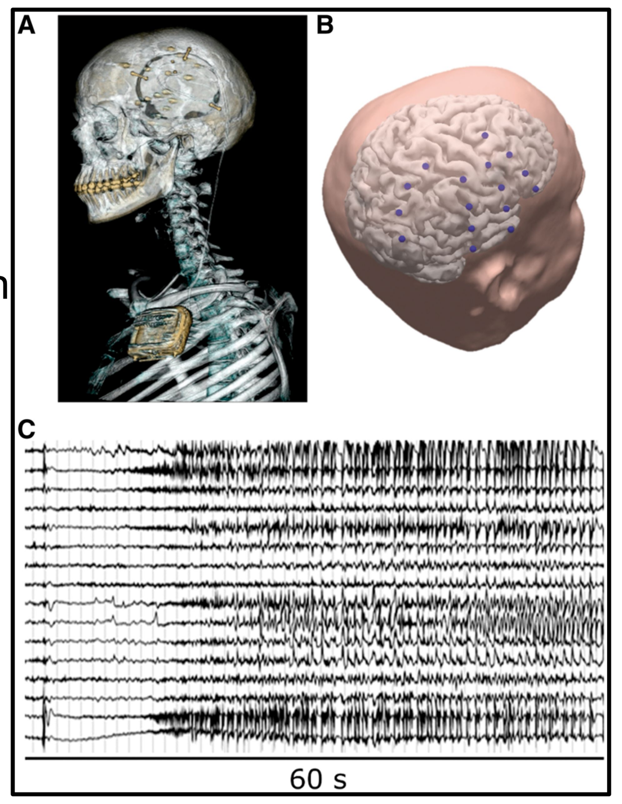


Figure 1. Showing the device used for data collection From: Epilepsyecosystem.org: crowd-sourcing reproducible seizure prediction with long-term human intracranial EEG. Brain.

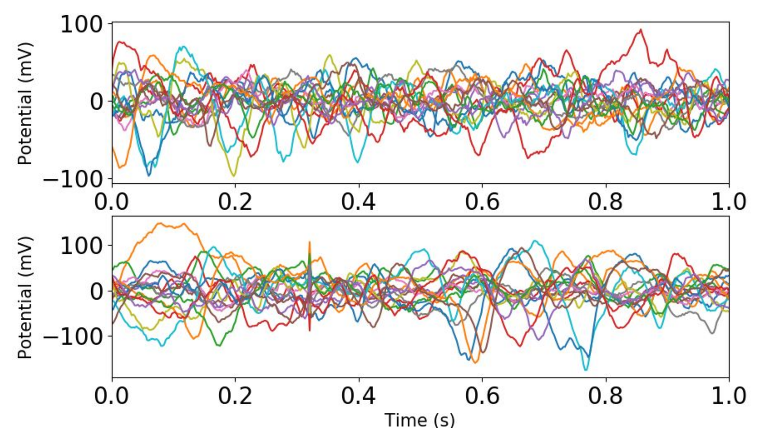
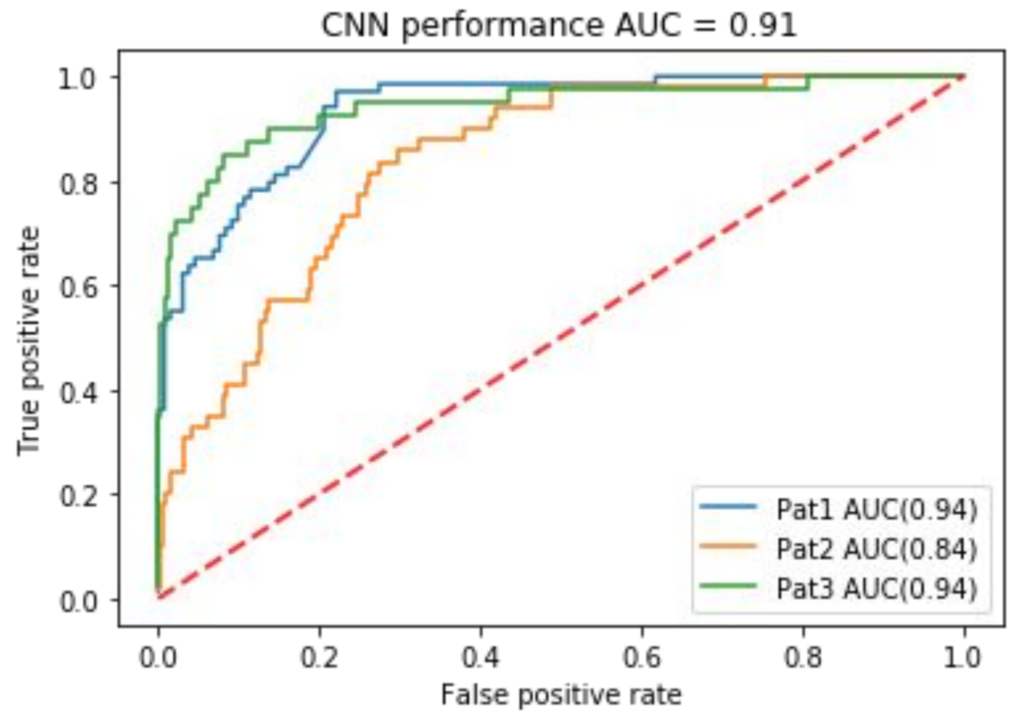


Figure 2: Showing data for an interictal (top) and preictal (bottom) period.

Subject	Training+CV (pos / neg / %)	Test
1	279 / 619 / 31%	144
2	240 / 2123 / 10%	697
3	273 / 2097 / 11%	483

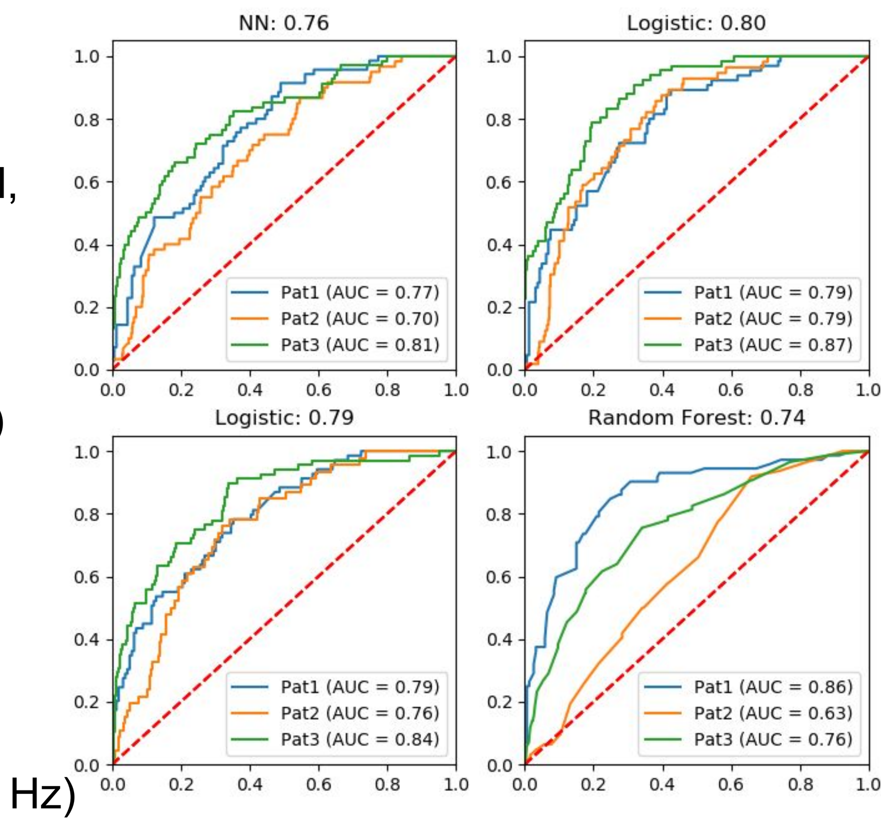
CNN / LSTM

- Due to the temporal nature of the data, we sought LSTMs and 1D CNNs as natural candidates for classification.
- In both cases, several layers of the applicable layer type were stacked, followed by a final layer with a single sigmoidal unit.
- Data was downsampled by 100x allowing for rapid prototyping while preserving enough biological information. A convolutional autoencoder was meant to serve this purpose but bilinear downsampling proved more effective.
- Both network types easily overfit because of the large ratio of features to data points. To mitigate this aggressive regularization techniques were employed. A high dropout rate (0.8-0.9) and L2 regularization (5e-3) along with the regularization effect of downsampling (less features) were most effective.
- RNNs gave good accuracy after regularization (0.9 AUC on patient 1) but proved to be slow to train due to their sequential nature which disallows for parallel processing on GPUs. LSTMs and GRUs performed similarly.
- CNNs' quick training allowed for many architectural iterations and a very high AUC (0.91). The best performing one had 3 layers and 8000 parameters. The deeper the layers the smaller the convolutions and the larger the number of filters.
- Larger neural networks (50000 parameters) had to be regularized more which led to longer training times. In theory, if it is trained more it should lead to even higher performance due to the higher complexity in the transformations being made before classification.



Existing Approach: Feature-Based Classifiers

- Previously the best performance approach has been to calculate statistical features for each channel, and use one of three approaches:
  - Logistic Regression
  - Random forest
  - Support Vector Machine (linear)
- Utilized Features:
  - Mean
  - Standard deviation
  - Kurtosis
  - Skew
  - Hurst-Exponents
  - Spectral power (2, 6, 10, 21, 45 Hz)
  - Pairwise channel coherence
- We found similar performance to existing best-ranked submissions.



Parameter Name	Evaluation Points
L1	1e-5, 1e-3, 0
L2	1e-5, 1e-3, 0
Batch_size	4, 32, 64
Dropout	0:0.05:0.3
N_units	[32, 1], [64, 1], [32,64, 1], [64,32,1], [32,64,32,1]
Block Length (seconds)	10, 30, 60, 600

Neural network parameters were optimized over a grid search with these values.

Autoregressive Features

- As a novel approach, we propose the use of *autoregressive coefficients* as features.
- Using the previous N-data samples from all channels, calculate a transformation matrix which gives the current observation on all channels
- Number of time points to use is identified by measuring the Bayes Information Criteria for varying model orders.
- We found optimal order of 5 samples (125ms, about the duration of an alpha cycle).
- Similar time periods were obtained with 400Hz data (46 samples)

$$\begin{bmatrix} C_1(t-1) & C_2(t-1) & \dots & C_{16}(t-1) \\ C_1(t-2) & C_2(t-2) & \dots & C_{16}(t-2) \\ \dots & \dots & \dots & \dots \\ C_1(t-N) & C_2(t-N) & \dots & C_{16}(t-N) \end{bmatrix}$$

$$\begin{bmatrix} C_1(t) & C_2(t) & \dots & C_{16}(t) \end{bmatrix}$$

$$X_t = c + \sum_{i=1}^{p} \rho_i X_{t-i} + \varepsilon_t$$

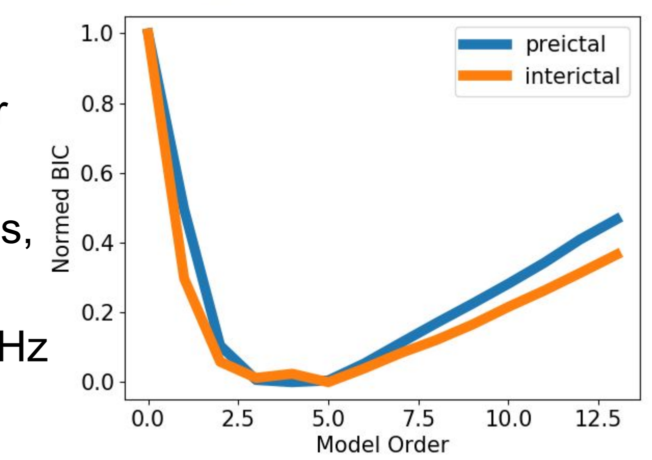


Figure 3: Showing BIC as a function of model order.

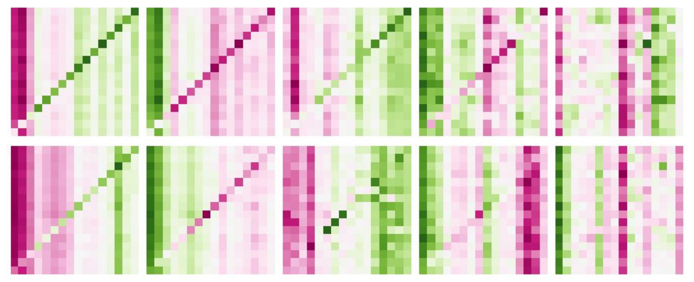


Figure 4. Showing Autoregressive coefficients for an interictal (top) and preictal (bottom) period. Green indicates positive and pink indicates negative.

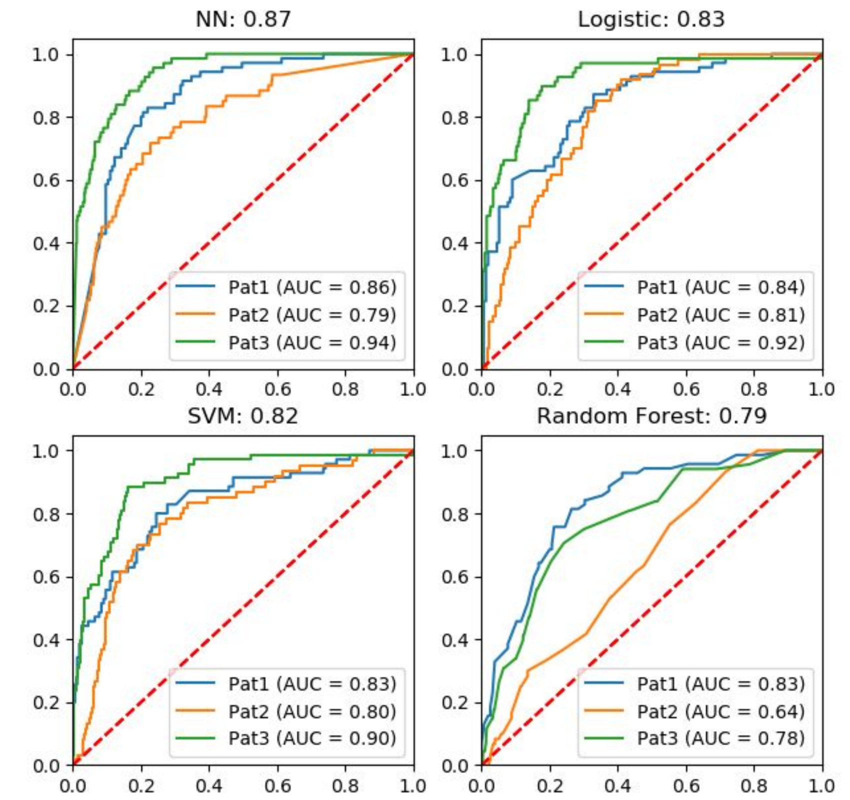
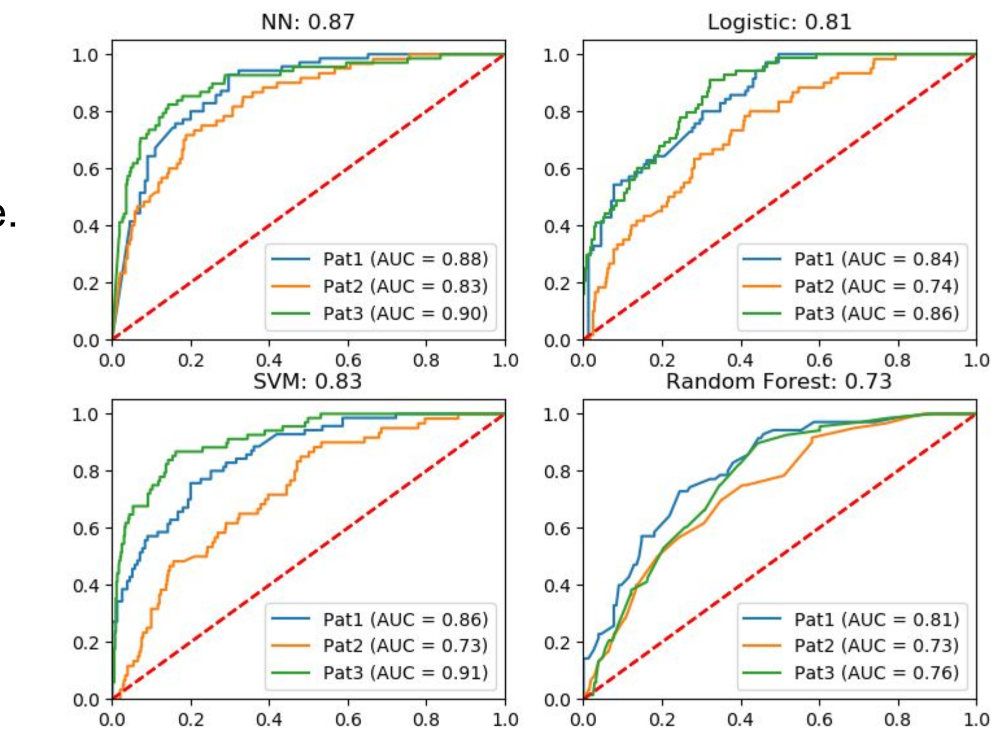


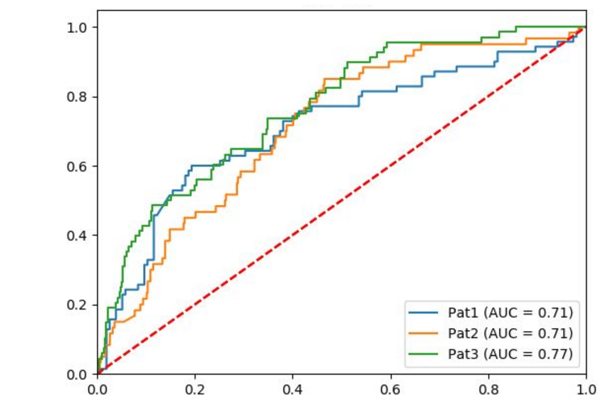
Figure 5. Showing AUC performance for the autoregressive coefficients alone. Perfromance is higher in each method than with traditional features, and places above previously-best algorith,s

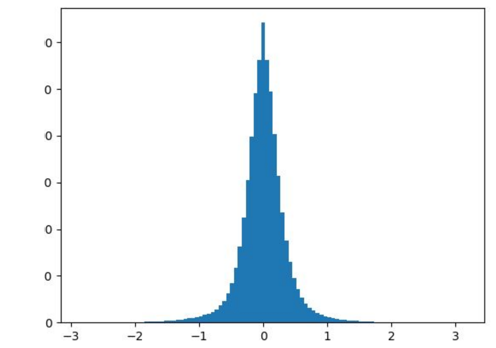
Unified Models

- Using a combination of both feature sets provided no improvement over using autoregressive features alone.
  - Additionally, accuracy dropped slightly for random-forest based approach



- Time-blocked (30 second intervals) autoregressive features were used as an input into a shallow NN, followed by LSTM showed significant overfitting. Training AUC was 0.95, but validation was only 0.73. Simple RNN and 1D CNN provided similar results.





- To assure that overfitting wasn't due to non-normal distribution of features, feature distribution was assessed. However, features were highly normal, with a mean of 0.001, and standard deviation of 1.43

Discussion

- We attempted two broad classes of machine learning approaches in an attempt to classify ECoG recordings as interictal or preictal in a clinically relevant dataset.
- We found that the RNN and CNN approaches, despite their propensity to overfit, are the highest performing methods.
- Our second approach was based on hand engineered features that represent a time-delayed coherence between channels. We used these features in several types of classifiers, and found that a shallow neural network performed highest, with an AUC of 0.87, though logistic regression and SVM performed closely.
- Patient 2 has the lowest AUC with almost every method. This suggests that some patients may have inherently more predictable seizures.
- While our results are only on the public leaderboard, we expect a 0.01 - 0.05 drop on true test data, as other algorithms experienced. This result would still place us in the top tier.

#	change	Team	Score	Entries
0	-	Group3 **(ours)**	0.91	1
1	+1	Not-so-random-any more	0.807	260
2	+35	Arete Associates	0.798	56
3	+12	ZhengYi	0.796	74
...	-			
42 (10%)	+19	CDS_Grp5	0.745	22

References

1) https://www.kaggle.com/c/melbourne-university-seizure-prediction
2) Kuhlmann, L. Crowd-Sourcing Reproducible Seizure Prediction with Long-Term Human Intracranial EEG, Brain, awy210, https://doi.org/10.1093/brain/awy210
3) Wang, Y., Yao, H., & Zhao, S. (2016). Auto-encoder based dimensionality reduction. Neurocomputing, 184, 232–242. https://doi.org/10.1016/J.NEUCOM.2015.08.104
4) M. Mursalin, Y. Zhang, Y. Chen, and N. V Chawla, "Automated epileptic seizure detection using improved correlation-based feature selection with random forest classifier," *Neurocomputing*, vol. 241, pp. 204–214, Jun. 2017.
5) Scikit-learn: Machine Learning in Python, Pedregosa *et al.*, JMLR 12, pp. 2825-2830, 2011.