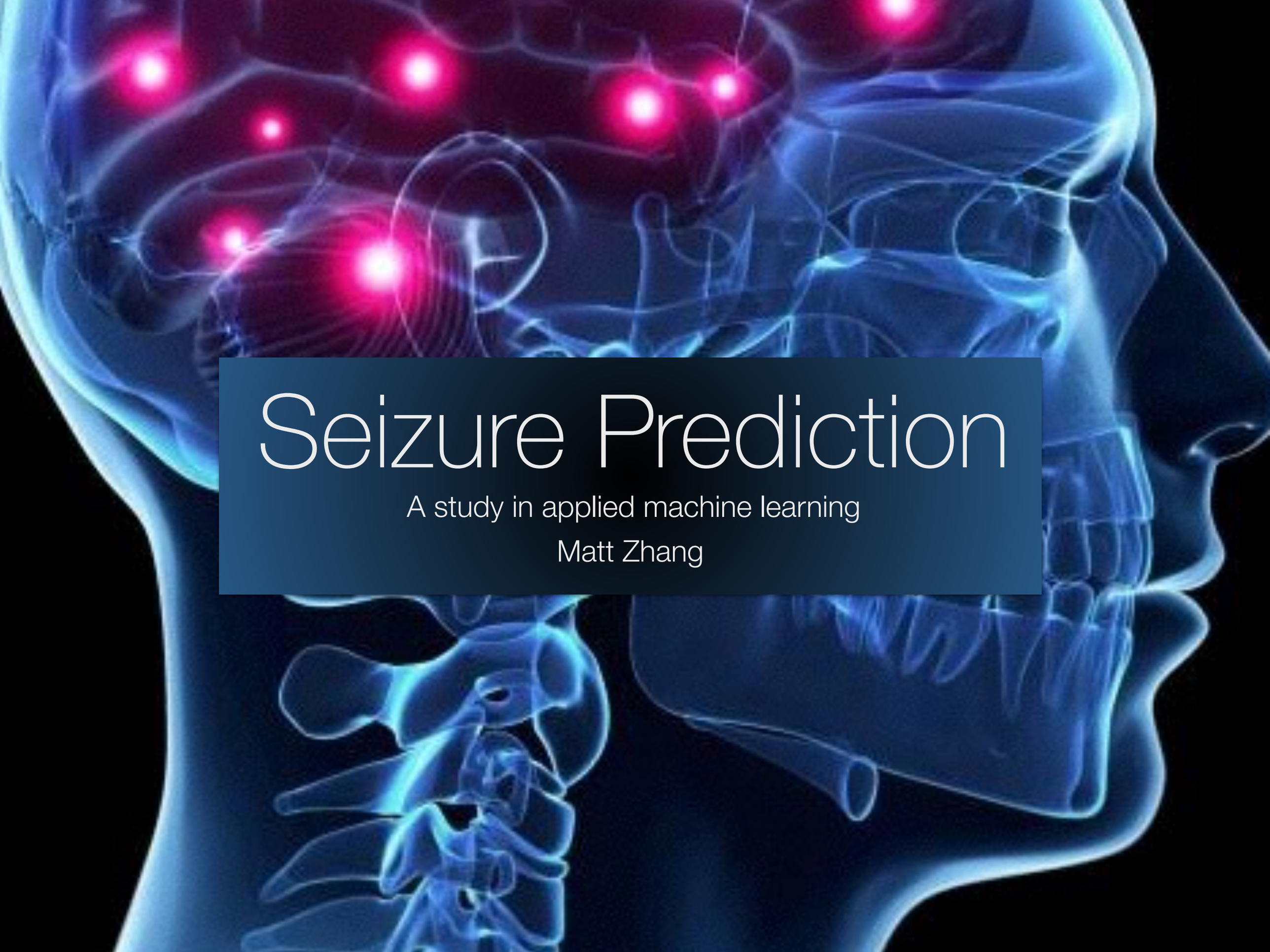


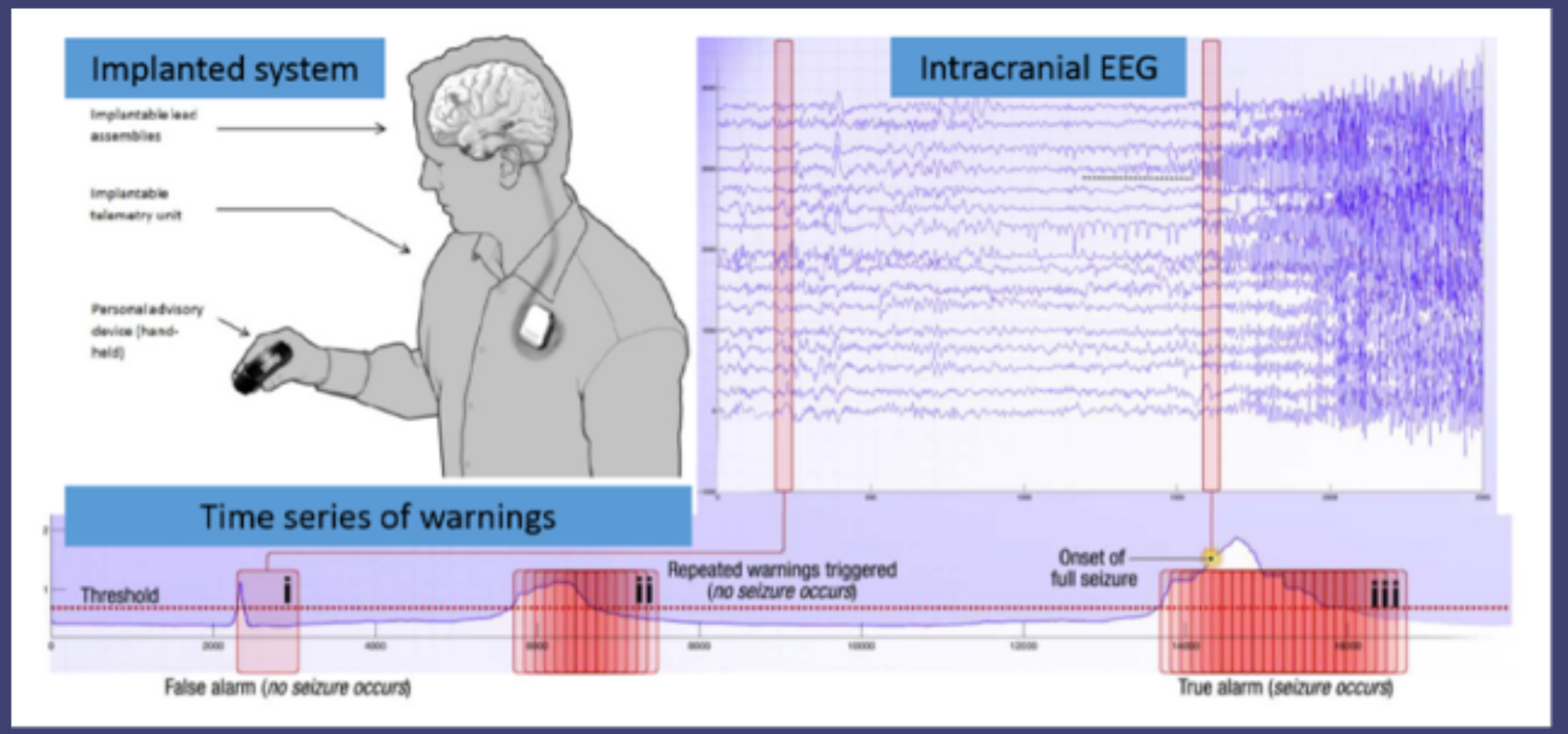
Seizure Prediction

A study in applied machine learning

Matt Zhang

The Problem

- Epilepsy affects 1% of the world population.
- Using implanted devices, researchers hope to be able to predict seizures from EEG signals before they occur. This allows people to stop driving, seek medical attention, etc.
- This is currently the topic of a competition on Kaggle.



< previous

next >

“Seizure prediction refers to attempts to forecast epileptic seizures based on the EEG before they occur. As of 2011, no effective mechanism to predict seizures has been developed.” - Wikipedia

< previous

next >

Seizure Mechanisms

- Sources of seizures are not entirely well understood. However, we know the following:
- Many seizures involve defects in neuronal firing suppression. Normally, neurons are not supposed to be able to fire multiple times in rapid succession. However, in epilepsy this resistance is decreased.
- This causes excessive synchronization in neurons, where many surrounding neurons start oscillating in sync and firing together, eventually spreading throughout the brain.

< previous

next >

A History of Seizure Prediction

Stolen from https://www.youtube.com/watch?v=KyUCQ6DYzdc

More than 50 years ago	• Seizures were believed to begin abruptly. • Some patients report auras and prodromes.
1970	• Viglione were the first to propose seizure prediction from the EEG (1970 and 1975)
80's – 90's	• Improvement of technology. • Implementation of nonlinear methods. • First methods suggesting that seizure prediction is possible.
Last 15 years	• Many methods have been proposed, most of them retrospective.
Now	• Need of long-term continuous recordings • Prospective analysis • Improving technical and methodological strategies.

< previous

next >

Learning from Past Research

- Kaggle already held one seizure prediction competition, based on data collected from dogs (and two humans).
- This current competition uses more data, collected for longer periods, in more humans.
- We can look at past solutions.

Held-out Data Results

Team	Algorithm	Public LB	Private LB	Held-out Data	% Change
QMSDP	Combined Bagged SVM, Lasso GLM, Random Forest	0.85951	0.81962	0.75431	-7.968
Jialun He	SVM with Platt scaling	0.83869	0.80079	0.73665	-8.009
Pardo et. al.	Combined Neural Network and kNN Clustering	0.82488	0.79347	0.6331	-20.211
Michael Hills	SVM, Genetic Algorithm Feature Selection	0.86248	0.79251	0.79022	-0.288
Wei Wu	Combined GLM and SVM	0.81803	0.78724	0.7726	-1.86
Ira Korshunova	Convolutional Neural Network	0.82455	0.78513	0.76338	-2.77

https://github.com/MichaelHills/seizure-prediction

< previous

next >

Features

Features used

- Time correlation matrix upper right triangle and sorted eigenvalues
- Frequency correlation matrix upper right triangle and sorted eigenvalues (omits 0Hz bucket)
- FFT Magnitude Log10 buckets for various ranges (see code below), where the power-in-band is calculated between the specified frequencies. The power-in-band is actually the average and not the sum. I saw minor boosts to perform Log10 after calculating power-in-band.
- Power-in-band spectral entropies
- Higuchi fractal dimension with kmax=2
- Petrosian fractal dimension
- Hurst exponent

< previous

next >

Features

- Correlations - Matrices of time and frequency correlations between channels.
- FFT Magnitude and Power in Bands - FFT, and integral of FFT within various frequency ranges.
- Fractal Dimensions - Measures of how much “detail” a complex pattern contains.
- Hurst exponent - “Measure of dependence”. Rate of decrease of a time series’s autocorrelation as the gap between pairs of values increases.

< previous

next >

k-Fold Cross-Validation

- Instead of just a training and test set, we split the data into k components.
- Each component takes a turn being the test set, while the rest form the training set.

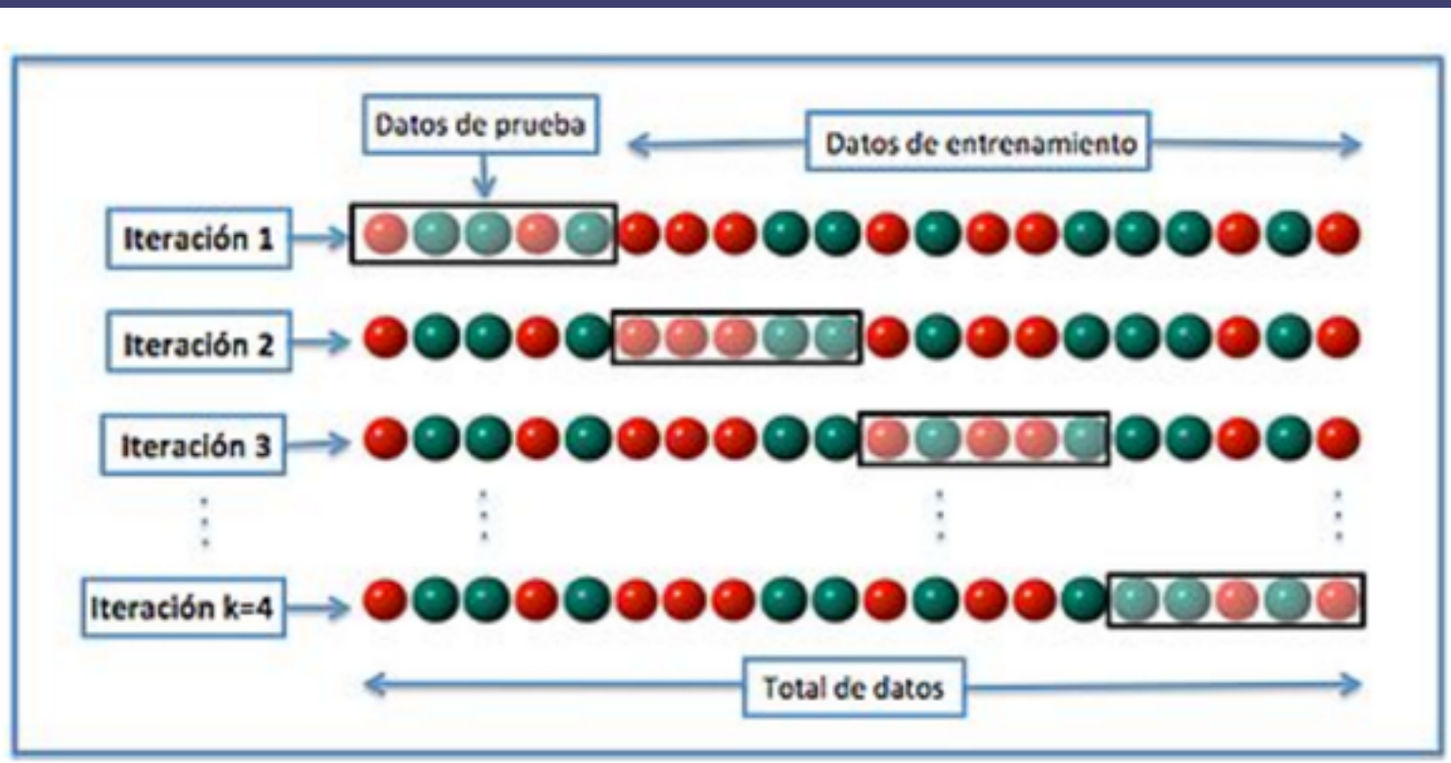


Diagram of k-fold cross-validation with k=4.

< previous

next >

The Algorithm

- After the competition, the algorithm author discovered that his genetic algorithm was beaten by simple linear regression…
- But let's look at the method anyways.
- Discussion of linear regression method: https://www.kaggle.com/c/seizure-prediction/forums/t/10945/congratulations-to-the-winners

The Algorithm

- A genetic algorithm with a population of 30 is run for 10 generations. Each individual has 45% of features masked and the other 55% activated.
- Fitness of an individual is determined by calculating the average cross-validation score for a given algorithm.
- This is repeated for multiple algorithms (several SVM's and linear regression), and the one with the best score is chosen for the final submission.

The Algorithm

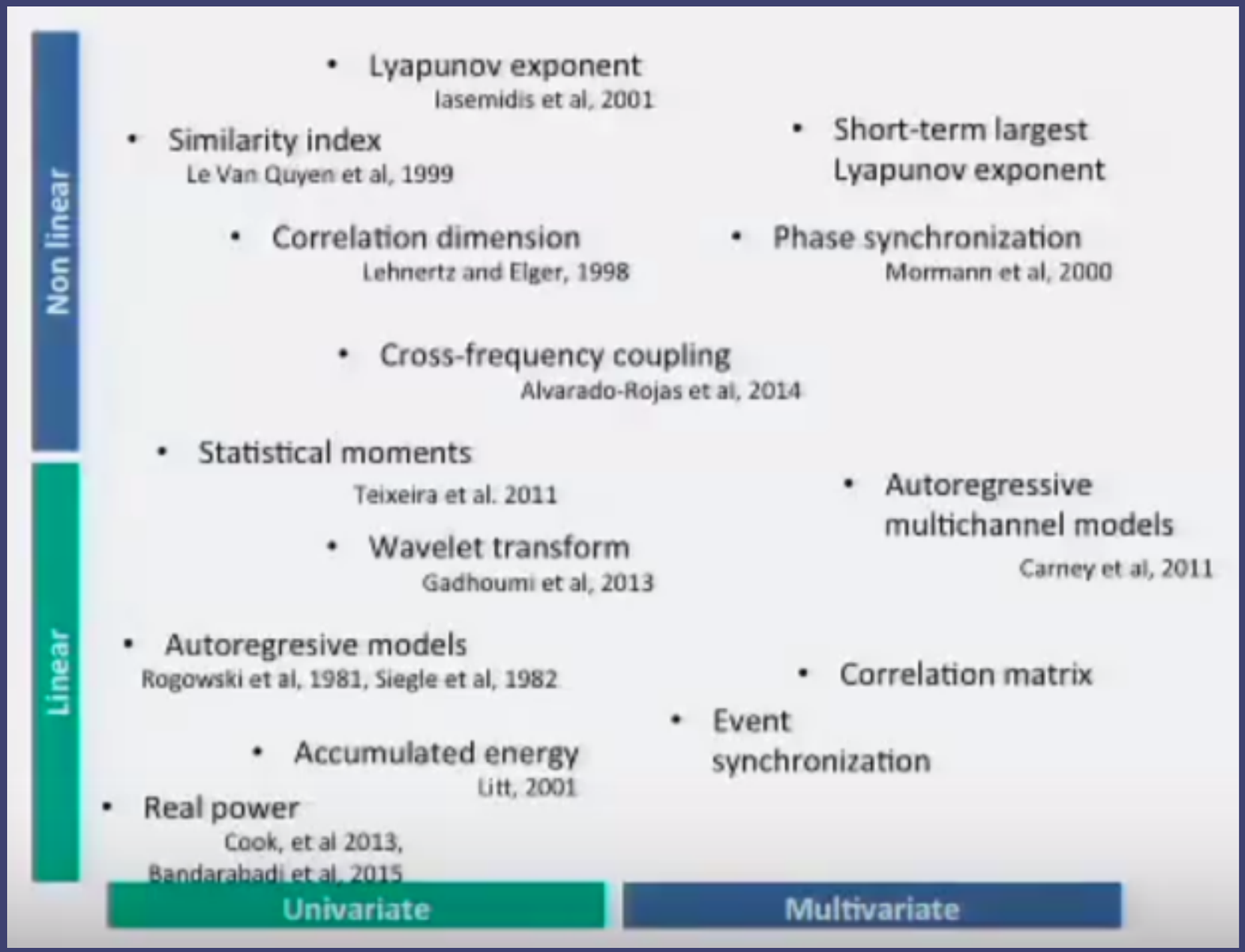
- According to the author, the sub-feature masks used in this genetic algorithm caused predictions to improve for two patients, caused no change in two, and caused two to get worse.
- “…using the linear regression approach more or less makes all of this a waste of time.”

Possible Improvements

- Clean signal before processing. These are standard pre-processing techniques, according to a presentation at the 7th International Workshop on Seizure Prediction (https://www.youtube.com/watch?v=_FcynpGA9yU):
  - Slice timing correction
  - Motion re-alignment
  - GLM-based denoising
  - scrubbing
  - 0.001-0.1 Hz bandpass filtering

Possible Improvements

- Try other standard EEG features:



< previous

next >

Anybody interested in working on this?

- Real world experience
- Fame and glory
- $20,000 prize pool