Melbourne University AES/MathWorks/NIH Seizure Prediction

Shivam Kalra & Ruifan Yu
{shivam.kalra,ruifan.yu}@uwaterloo.ca

November 29, 2016

Problem Description
Introduction

- Nearly one-third of patients with epilepsy continue to have **seizures** despite optimal medication management [1].
- Seizures are a symptom associated with abnormal electrical activity in the brain.
- What is a seizure? and When to detect it? questions remain elusive.
- Plenty data is available, **machine learning** can help in building seizure forecasting systems.
- **Could save life!**

Problem Description

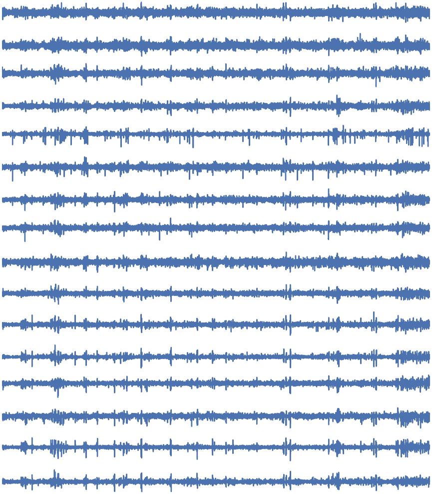
Melbourne University AES/MathWorks/NIH Seizure Prediction

What is given and what is required?

- Human brain activity (intracranial EEG) taken from multiple sensors on brain.
- Each recording is **10 minutes** long, recorded at **400 Hz** resulting **240,000** data points per recording.
- Challenge is to classify unseen recording as **Preictal** (prior to seizure) or **Interictal** (at least an hour before seizure).

Problem Description

Tell me if this is Interictal or Preictal?



Input

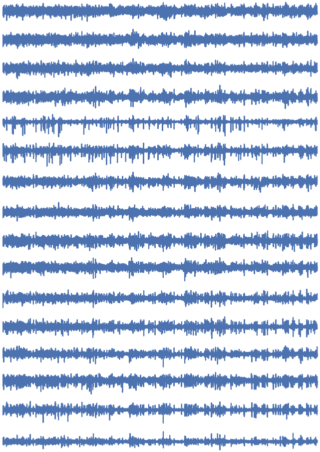
Figure shows **one sample of input data** from Patient 1's data-set. Shape of the data is 16×240000.

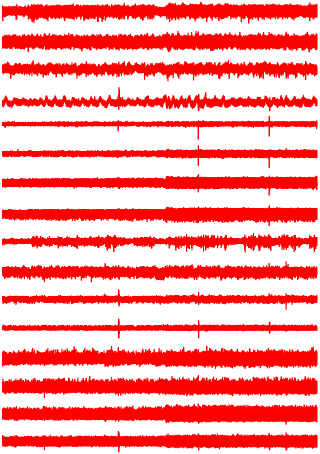
Output

Classifier model needs to figure out if it is **Preictal** or **Interictal**.

Problem Description

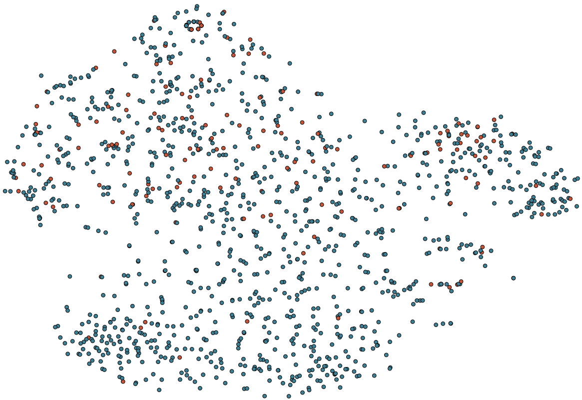
Red is Preictal. But predicting this is not very trivial...





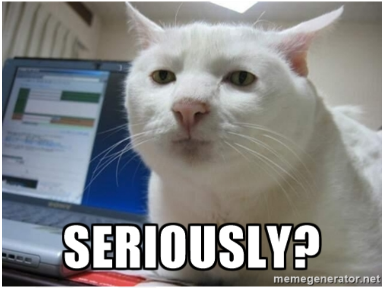
Problem Description

How good is raw data?? TSNE visualization for Patient 1 Channel 3.



Problem Description

Okay, so prediction is hard on raw data..but we got more issues!!!



Apart from being already difficult prediction problem, we got bigger issues with data-set. **This is not cool!!**

Overfit

Problems with Data Set

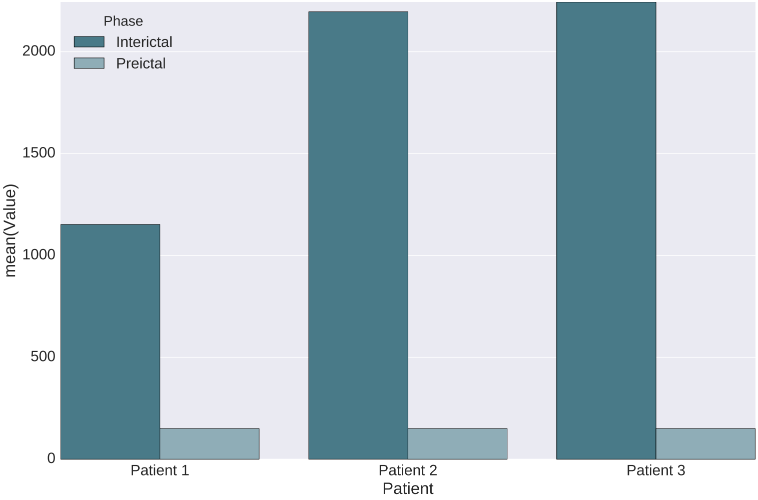
There are only two types of people in the world, those who can extrapolate from incomplete data...

Training Data-set has...

1. Categorical Imbalance
2. Missing Data or Random Dropouts

Problems with Data Set

Categorical Imbalance



Problems with Data Set

Missing/Dropouts in Data Set

- Random dropouts of **10 seconds or more** in the EEG signals across all the 16 channels.
- Exist in **abundance** (even in testing data set).
- Some training data is **entirely empty** (completely missing!!!)

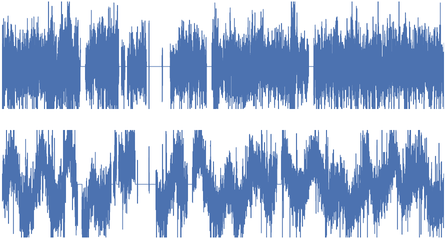


Figure: Missing data in channel 1 and 3 from Patient 1's data set

Model 1: SVM/RF/GB classification using various DSP features

Features Extraction: Frequency Bands

Frequency bands are chosen as in [2].

Table: Frequency bands in Spectrograms

Name	**Frequency Range (Hz)**
Delta	0.4 – 4
Theta	4 – 8
Alpha	8 – 12
Beta	12 – 30
Lowgamma	30 – 80
Highgamma	80 – 180

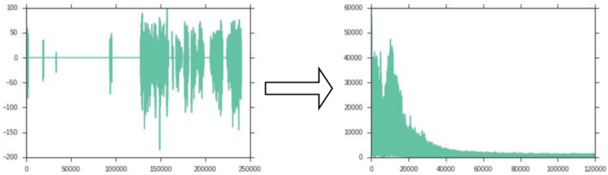
Model 1: SVM/RF/GB classification using various DSP features

Features Extraction: Frequency Domain

Each feature is extracted from window of 20 seconds for a given input EEG wave.

Frequency Domain

- Extracted magnitudes of **FFT** in each frequency band.
- **Total Features**: $16 \times 6 = 96$



Model 1: SVM/RF/GB classification using various DSP features

Features Extraction: Time Domain

Time Domain

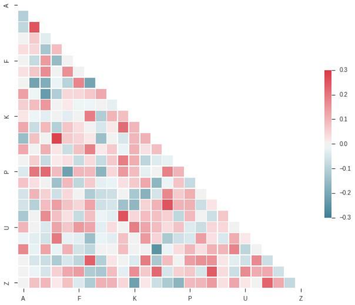
- Extracted energy by **Butterworth Band pass Filter** within each frequency band.
- **Total Features**: $16 \times 6 = 96$

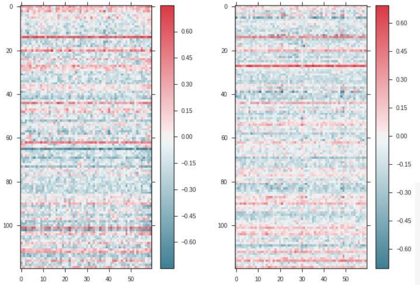
Model 1: SVM/RF/GB classification using various DSP features

Features Extraction: Correlation

Correlation Features

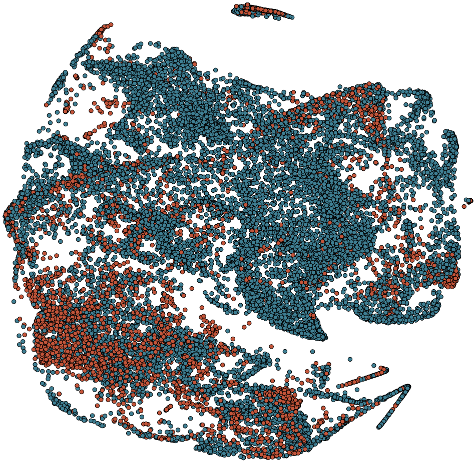
- Calculated correlation between pair of channels.
- **Total Features**: $\frac{(16-1)\times 16}{2} = 120$



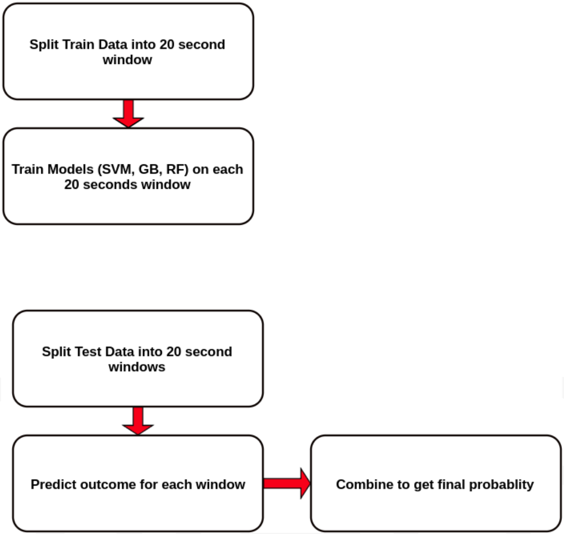


Model 1: SVM/RF/GB classification using various DSP features

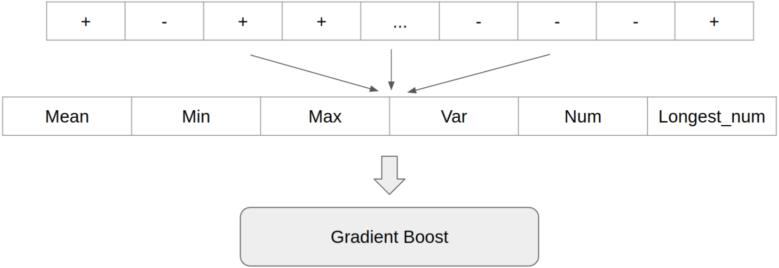
TSNE Visualization of all features combined ($120 + 96 + 96 = 312$)



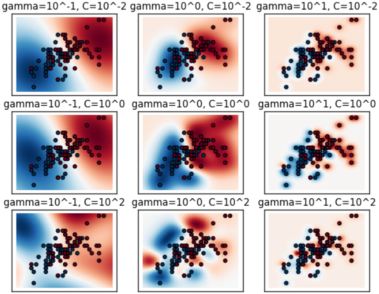
Model 1: Feature Extraction Summarized...



Model 1: Final Prediction



Model 1: Experiences with SVM...



- **Unbalanced labels:** Set `class_weight` to `balanced` in Scikit-Learn library.
- Big `C` brings penalty on mis-classified observations
- Big `gamma` gives us more flexible decision boundary, but very likely to over-fit.

Overfit

Model 1: Experiences with Gradient Boost...

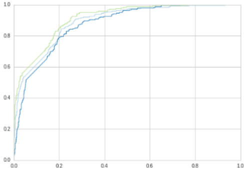
Handle Imbalanced Dataset

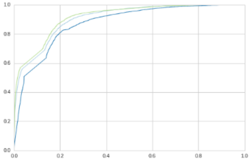
- Balance the positive and negative weights, via `scale_pos_weight`.

Control Over-fitting

- Control model complexity set `max_depth`, `min_child_weight` and `gamma`.
- Add randomness to make training robust to noise, set sub-sample ratio.
- Set early stop round and limit the `max_depth` to avoid over-fitting.

Model 1: Unreal AUC in Training Data





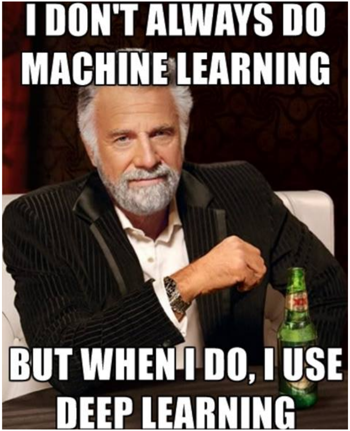
Confusion Matrix

	0	**1**
0	939	21
1	124	116

	0	**1**
0	16822	278
1	2198	2302

- This is caused by shuffling the train test data-set.
- Shuffling causes data leakage in the validation set.
- There should not be overlapping between train and test set.
- This will lead to a higher CV score during training (manually separate the train test set)

Model 2: Deep Learning (CNN) on Spectrograms



Model 2: Deep Learning (CNN) on Spectrograms

Some background research...

Why Convolution Neural Networks?

- Our **eternal love** for deep learning...
- Successful results by CNN in seizure detection as in [2, 3].
- Iryna Korshunova's CNN approach in [2] topped among best 10% in 2014's seizure prediction competition.

Model 2: Deep Learning (CNN) on Spectrograms

Input/Output for training the CNN...

Preprocessing

All the **unsafe** data points are removed as per guideline of competition.

Input

- Raw wave data is converted into **Spectrograms**.
- **Frequencies** are binned into **6 bands**.
- **Time domain** is binned with windows of **10 seconds** each.

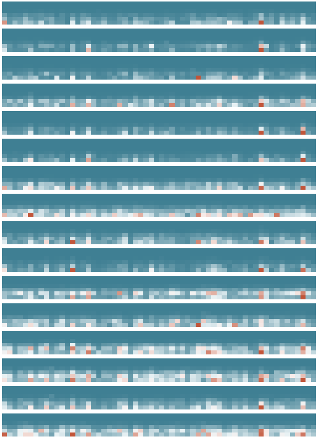
Output

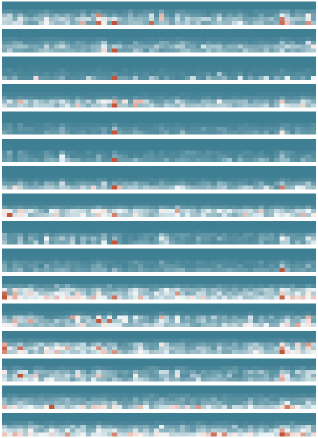
- Interictal $[1, 0]$
- or, Preictal $[0, 1]$

Model 2: Deep Learning (CNN) on Spectrograms

Final Spectrogram as Input to CNN

Final Spectrogram becomes: $16 \times 6 \times 10$. **Right is Preictal, can you guess? Well, lets leave it to CNN to classify!**





Model 2: Deep Learning (CNN) on Spectrograms

What is convolution?

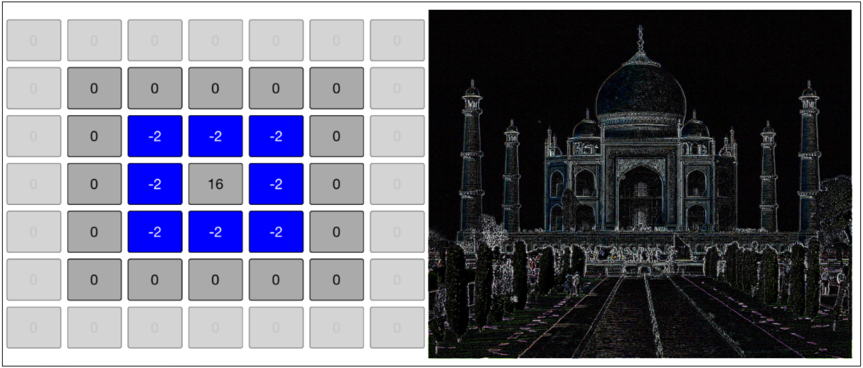
Convolution Operator

$$(f * g)(t) = \int_{-\infty}^{\infty} f(\tau)g(t-\tau)d\tau$$

- Can be generalized to *n* dimensions.
- Fast to calculate.
- Very important concept in discrete numerical analysis.

Model 2: Deep Learning (CNN) on Spectrograms

Example of Convolution in Image Processing

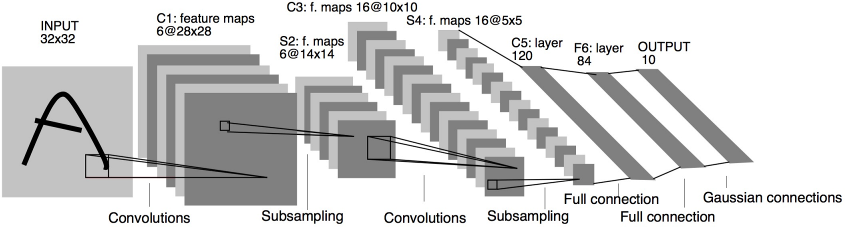


Ref: https://flexmonkey.blogspot.ca/2015/05/convolution-filters-in-swift-with.html

Model 2: Deep Learning (CNN) on Spectrograms

Overview of Convolution Neural Network (CNN)

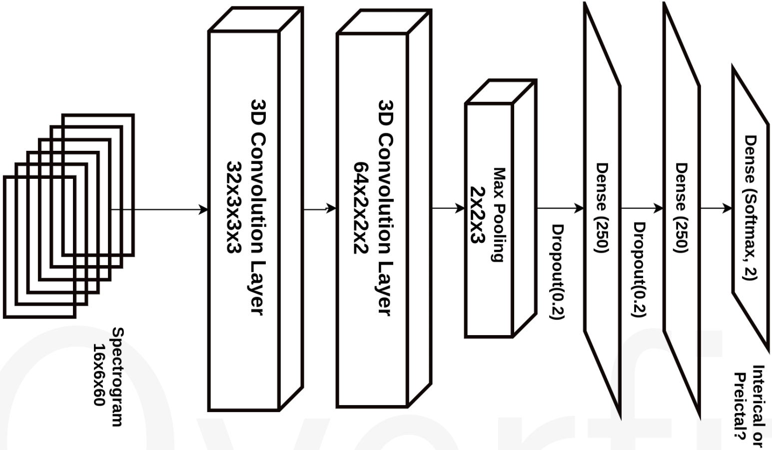
- Learns **Convolution Kernels** (more native to image operations).
- Apply **Pooling** to reduce dimension.
- Perform classification using **Dense** layers.



Ref: `https://culurciello.github.io/assets/nets/lenet5.jpg`

Model 2: Deep Learning (CNN) on Spectrograms

Overview of our CNN Architecture (AUC $\sim$ 0.72)



Overfit

Model 2: Deep Learning (CNN) on Spectrograms

Network fine tuning...

Data Imbalance

- **Data Augmentation:** Added more augmented data/even repetition.
- **Bagging:** Train models with **5 bags** and average out.

Missing Data

Data Augmentation: Augmented white patches on more data. Let network handle bad data.

Are we over-fitting yet?

- **Early Stopping:** Stop training if validation accuracy isn't changing.
- **Folding:** Trained models using **5-Fold Cross Validation** and averaged them out.
- **Dropout:** Regularization to prevent over-fitting.

Model 2: Deep Learning (CNN) on Spectrograms

Network fine tuning...

Data Imbalance

- **Data Augmentation:** Added more augmented data/even repetition.
- **Bagging:** Train models with **5 bags** and average out.

Missing Data

Data Augmentation: Augmented white patches on more data. Let network handle bad data.

Are we over-fitting yet?

- **Early Stopping:** Stop training if validation accuracy isn't changing.
- **Folding:** Trained models using **5-Fold Cross Validation** and averaged them out.
- **Dropout:** Regularization to prevent over-fitting.

Model 2: Deep Learning (CNN) on Spectrograms

Network fine tuning...

Data Imbalance

- **Data Augmentation:** Added more augmented data/even repetition.
- **Bagging:** Train models with **5 bags** and average out.

Missing Data

Data Augmentation: Augmented white patches on more data. Let network handle bad data.

Are we over-fitting yet?

- **Early Stopping:** Stop training if validation accuracy isn't changing.
- **Folding:** Trained models using **5-Fold Cross Validation** and averaged them out.
- **Dropout:** Regularization to prevent over-fitting.

Model 2: Deep Learning (CNN) on Spectrograms

Network fine tuning...

Data Imbalance

- **Data Augmentation:** Added more augmented data/even repetition.
- **Bagging:** Train models with **5 bags** and average out.

Missing Data

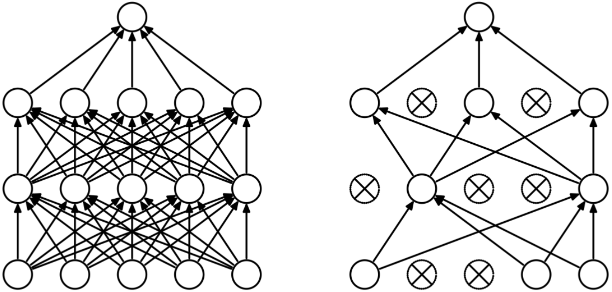
Data Augmentation: Augmented white patches on more data. Let network handle bad data.

Are we over-fitting yet?

- **Early Stopping:** Stop training if validation accuracy isn't changing.
- **Folding:** Trained models using **5-Fold Cross Validation** and averaged them out.
- **Dropout:** Regularization to prevent over-fitting.

Model 2: Deep Learning (CNN) on Spectrograms

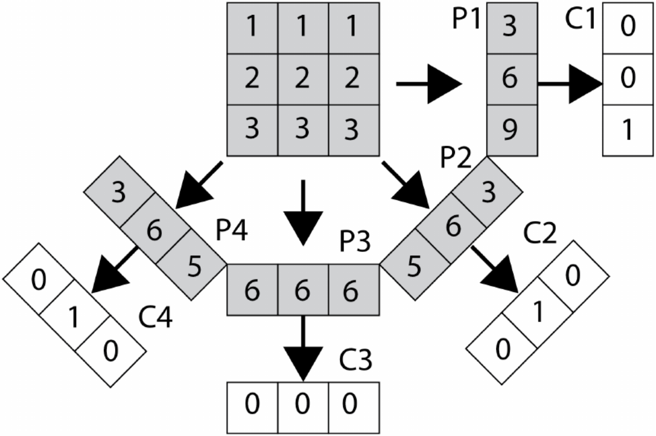
Dropout: An effective technique to prevent over-fitting...



(a) Standard Neural Net (b) After applying dropout.

Overfit

Model 3: SVM Classification on Radon Transforms



Ref: https://arxiv.org/pdf/1604.04675.pdf (Page 3)

Model 3: SVM Classification on Radon Transforms

Motivation behind using Radon Transform...

Why Radon Transform?

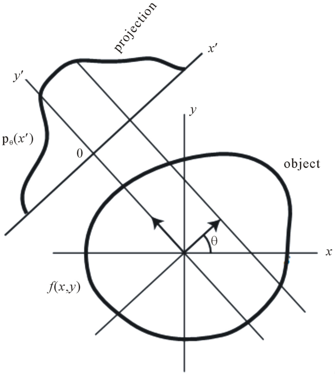
- It is my research topic...so, I could not resist!
- It has been successful in recognition of speaker from sound signals [4].
- Well known/established in medical fields (basis of CT Scans, Tomography)
- Can provide with interesting robust features which are interpret-able as well.

Model 3: SVM Classification on Radon Transforms

SVM with Radon Features AUC $\sim$ 0.65

Input

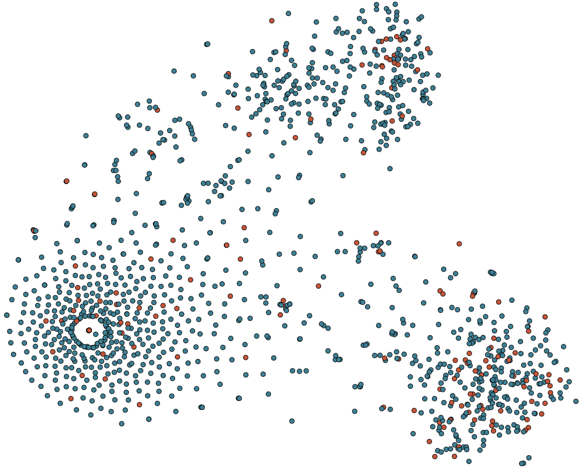
Radon projections from stitched Spectrogram using **8** equidistant angles.



- Projection from each 8 of the angles are concatenated into one vector.
- Vector length: $\sqrt{(6 \times 16)^2 + 10^2} \times 8 = 776$.
- Can run PCA to further reduce dimensions to around 512.
- Use **SVM** to classify using 5-Folds Cross Validation and 5 Bags averaging model (AUC $\sim$ 0.65).

Model 3: SVM Classification on Radon Transforms

TSNE Visualization of Radon Projections



Classification

Current best score on Kaggle...

Ensemble of 3 different classifier models:

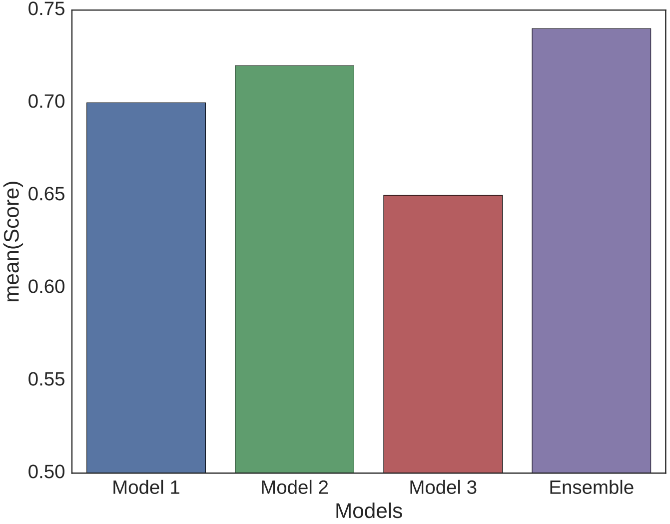
1. **SVM/RF/GB** classification using various **DSP features**
2. **Deep Learning:** CNN classification of Spectrograms
3. SVM classification on **Radon Transforms** of Spectrograms

Current Kaggle Score

~ 0.74

Overfit

Finals Scores



Thanks

- S. Ramgopal, S. Thome-Souza, M. Jackson, N. E. Kadish, I. Sánchez Fernández, J. Klehm, W. Bosl, C. Reinsberger, S. Schachter, and T. Loddenkemper, "Seizure detection, seizure prediction, and closed-loop warning systems in epilepsy," vol. 37, pp. 291–307.
- I. Korshunova, "Faculty of sciences,"
- P. W. Mirowski, Y. LeCun, D. Madhavan, and R. Kuzniecky, "Comparing svm and convolutional networks for epileptic seizure prediction from intracranial eeg," in *2008 IEEE Workshop on Machine Learning for Signal Processing*, pp. 244–249, IEEE, 2008.
- P. K. Ajmera, D. V. Jadhav, and R. S. Holambe, "Text-independent speaker identification using Radon and discrete cosine transforms based features from speech spectrogram," vol. 44, pp. 2749–2759.