

15-781
Project Proposal
Revanth Bhattaram
Vipul Singh
Archit Karandikar
1st October 2014

- **Title:** American Epilepsy Society Seizure Prediction Challenge

- **Group Members:**

- Revanth Bhattaram (rbhattar@andrew.cmu.edu)
- Vipul Singh (vipuls@andrew.cmu.edu)
- Archit Karandikar (akarandi@andrew.cmu.edu)

- **Task:** Epilepsy is a neurological disorder. It is characterized by epileptic seizures. Patients of epilepsy are weary of the possibility of occurrence of seizures. This constant anxiety is despite the fact the the seizures themselves may be infrequent. This is why seizure forecasting systems can potentially help epileptic patients to lead more normal lives. In this project, we will address the challenge of forecasting seizures in epilepsy patients.

Electroencephalography (EEG) is the recording of the voltage fluctuations resulting from ionic current flows within the neurons of the brain. Our forecast will be based on the results of the EEG signals.

This task is a [competition](#) on *kaggle*, an online platform for data analytics competitions. It is sponsored by the National Institutes of Health (NINDS), the Epilepsy Foundation, and the American Epilepsy Society.

- **Type of Learning:** This is a two class classification problem where given an iEEG dataclip, we'll have to classify it as preictal(prior to a seizure) activity or interictal(the baseline state) activity. Since we'll be working with datasets that contained labeled data (iEEG dataclips), this problem will be solved using a **supervised** learning approach.
- **Dataset:** The dataset we'll be using for this project is provided by kaggle on the challenge page. Additional annotated intracranial EEG data is freely available at the International Epilepsy Electrophysiology Portal.

The dataset consists of Intracranial EEG (iEEG) data clips recorded from canine and human test subjects.

The dataset is provided in a *coder-friendly* format. This would ensure that preprocessing (if any) isn't too much work.

- **Training Data Set :** The training data is organized into one hour sequences of ten minute iEEG clips labeled "Preictal" for pre-seizure data segments, or "Interictal" for non-seizure data segments.

The interictal activity data, in the case of canines, is recorded at least a week before and after a seizure. For human training data, the interictal training data is recorded at least 4 hours before and after a seizure.

The preictal activity data is recorded one hour prior to a seizure with a 5 minute offset.

- **Testing Data Set :** The testing dataset consists of 10 minute iEEG clips which are provided in random order. This data is provided each of the canine and human test

subjects. Our task would be to predict whether this data counts as interictal activity or preictal activity (which will be let us predict if the patient will have a seizure or not).

- **Validation:** The contest has a provision for online submission with instant feedback. The feedback is in terms of the collective accuracy of the predictions. The results of the training data (not available to contestants) are used for evaluating submissions.

In addition, it's always possible for us to be able to split the training dataset and use a part of it to test the model we build.

- **Mid-way target:** We plan to apply multiple supervised Machine Learning algorithms customized to this setting and compare their results. By the mid-way deadline we hope to identify the set of algorithms we will use and implement basic solutions for each of them.

- **Minimum Goal:**

- Successful implementation of at least two different algorithms.
- Accuracy at least 65% or rank within 100.

- **Ambitious Goal:**

- Successful implementation of at least three different algorithms, each with at least 60% accuracy.
- Accuracy at least 75% and rank within 20.