Predicting Seizures and Epilepsy

Although epilepsy is one of the nation's most common neurological disorders, public understanding of it is limited. Many people do not know the causes of epilepsy or what they should do if they see someone having a seizure. Epilepsy is a complex spectrum of disorders that affects an estimated 2.2 million Americans in a variety of ways, and is characterized by unpredictable seizures that differ in type, cause, and severity. Yet, living with epilepsy is about much more than just seizures; the disorder is often defined in practical terms, such as challenges in school, uncertainties about social situations and employment, limitations on driving, and questions about independent living.[1]

Several tests to diagnose epilepsy are performed to determine the cause of seizures. These tests may include:

- Neurological examination. Your doctor may test your behavior, motor abilities, mental function and other areas to diagnose your condition and determine the type of epilepsy you may have.
- Blood tests
- EEG and other brain wave monitoring methods

However, the critical factors and threshold for diagnosis based on these tools is not hard-fast. As such, exploring various factors that might affect epilepsy and/or seizure recurrence could be useful for the population.

Milestones:

1. **Project Selection**: Form teams of 2 or 3 and consider the project a classification problem for predicting future seizures from a rich dataset.
2. **Literature Study**: Go through the following resources for background on the project and write a half to 1 page summary for each one:

   - The Association Between Childhood Seizures and Later Childhood Emotional and Behavioral Problems: Findings From a Nationally Representative Birth Cohort.

   - National Child Development Study (NCDS - 1958 British Birth Cohort Study)

   - Organization for Economic Co-operation and Development Health Statistics (link below)

3. **Data Exploration and Cleaning:** The primary source of data for this project is a publicly available repository:

---

[1] http://nationalacademies.org/HMD/Reports/2012/Epilepsy-Across-the-Spectrum.aspx

- https://www.ukdataservice.ac.uk/use-data/guides/dataset/health-statistics
- http://ije.oxfordjournals.org/content/35/1/34.full
- Choose a suitable database to store the data, and on a computing resource to process the data.

- Visualize the effects of different key factors, such as seizure onset and morbidity

- Check for correlations (if any).

- Remove records with spurious entries (e.g. with invalid entries)

4. Proposal:
Propose methodologies and ideas to be implemented, tested and interpreted for your final project.

5. Baselines:
- Decide on the *performance metric* to evaluate prediction.
- *Feature extraction*: Extract a set of basic features from past data
- Implement the following baseline techniques:
  - *Linear regression*: Train a linear regression model on features extracted from seizure data.
  - *Logistic regression*: Using logistic regression to predict the occurrence of subsequent seizures following the first occurrence.
  - *Linear discriminant analysis:* Using LDA to parse the textual data and categorize factors effecting seizure occurrence into certain groups and with demographics.