Requirements:

* EEGlab with Biosig v3.8.1+ and ICLabel v1.4+

**PREPROCESSING**

Scripts:

* GetData (Support function, only needed if Data202012221319.xlsx needed to be rerun)
  + Creates an excel spreadsheet from a local copy of the TUEG data base v.1.1.0 and v.1.2.0
  + Output: Data202012221319.xlsx
* RecordAnalysis
  + Compares RecordDatabase.mat vs Vivaldi (“Original”) Cohort, and Abnormal   
    (“Lopez”) Cohort.
  + Removes duplicates
  + Uses BERT for database predictions
  + Final Curation
    - Additional keywords
    - Age
  + Outputs: PredictionDatabase.mat
* dreegstruc = DREEG(T,n,all\_data)
  + Runs dimension reduction from table T with location field. This includes cleaning the data and running a subset of features.
  + n denotes the amount of minutes the EEG is segmented into.
  + All\_data (default 0) can be set to 1 to run a larger collection of features.
  + Examples of calcualted feature sets are included as
    - HEA\_3F2.mat
    - STR\_3F2.mat
    - TBI\_3F2.mat
* Out = DREEGplot(dreegstruc,out,N)
  + Calculates 3D tSNE calculations and settings for the output of DREEG.
  + Out can be blank or a previous generated out structure
  + N (default = 1000) is number of iterations of tSNE.
* Names = DREEGd(F,p)
  + Selects and plots the tSNE points in F (dreegstruc) at percentage p (default = 0.5) away from the center of mass. Names output denotes selected points.
* DREEGexport2
  + Prepares and exports selected EEGs (saved as \*\_3F2.mat in the workspace) from local TUEG database dump to usable EEG files. The files are all preprocessed (cleaned, segmented, filtered, reordered, resampled, etc.)
  + Estimated size: 88.7 GB

Files:

* RecordDatabase.mat
  + A: table
    - 6688 EEG sessions and selected demographic fields extracted from the TUEG included .txt file. These have been manually marked as either “Normal”, “TBI”, “Stroke”, or “Unknown”.
    - Fields include:
      * Filename
      * Subject #
      * Session #
      * Category
        + Manually marked (see above)
      * Age
      * Sex
      * Medication #
      * “Plus other” flag
        + If the .txt indicated that there were other medications and they were not specified this flag was 1, else it was 0.
      * Full Note
        + The entire extracted .txt data
      * User Comments
        + Comments inputted by user conducting the Category manual marking.
* Data202012221319.xlsx
  + Spreadsheet of all the TUEG EEG Sessions
* Paper\_Tables.mat
  + T\_HEA: table
    - 121 Subject and Session #s
  + T\_STR: table
    - 344 Subject and Session #s
  + T\_TBI: table
    - 143 Subject and Session #s
* LopezData.mat
  + 2993 selected files and if they were identified as “Abnormal” or “Normal”.
* textnet.mat
  + net: SeriesNetwork
    - Trained net used on the BERT embeddings
* PredictionDatabase.mat
  + T: table
    - 19449 EEG Sessions and selected demographic fields extracted from the TUEG included .txt file. These have been predicted by net as either “Normal”, “TBI”, “Stroke”, or “Unknown”.
    - Fields include:
      * FileID
      * File Location
      * Subject #
      * Session #
      * Flag for TBI keywords
        + Empty if not TBI keywords, else the keywords are included here
      * Age
      * Sex
      * mTBIEvidence
        + Localized txt data for the TBI flag. Empty if flag is empty.
      * Method
        + Extracted Method section of the .txt file.
      * Medication text
        + Extracted Medication section of the .txt file.
      * Full Note
        + The entire extracted .txt data
      * Montage of EEG
      * Date of recording
      * EEG Type
      * EEG Subtype
      * LTM/Routine
      * Filenames
      * Category
        + Predicted categories from net.
      * Score
        + Prediction score from net.
  + net: SeriesNetwork
    - Trained net used on the BERT embeddings. This is imported from textnet.mat unless retrained in the script.
  + PredictedFiles: CohortFiles
    - The table T, converted to a CohortFiles class. This allows some visualization functions such as wordclouds
* HEA\_3F2.mat
  + Normal subject dimension reduction structure. Includes:
    - F: feature matrix
    - Names: list of string names and number of EEG segment
    - Min: the number of minutes used to segment
    - T: table of demographics
    - Out: structure output of DREEGplot
* STR\_3F2.mat
  + Stroke subject dimension reduction structure. Includes:
    - F: feature matrix
    - Names: list of string names and number of EEG segment
    - Min: the number of minutes used to segment
    - T: table of demographics
    - Out: structure output of DREEGplot
* TBI\_3F2.mat
  + TBI subject dimension reduction structure. Includes:
    - F: feature matrix
    - Names: list of string names and number of EEG segment
    - Min: the number of minutes used to segment
    - T: table of demographics
    - Out: structure output of DREEGplot

Classes:

* CohortFiles
  + A selection of files and functions to better use files associated with cohorts and to read the medical txt files associated with TUEG.

**TRAINING**

Scripts:

* MatchSubjects
  + Function to match normal and stroke subjects to the TBI cohort using gender and age.
* TrainFeatures
  + Live script used to train networks with feature data.
* TrainTopo
  + Live script used to train topographic map network
* TrainFusion
  + Live script used to train sensor fusion network
* TrainSTFT
  + Live script used to train the Short time fourier transform network

Files:

* Sep22.mat
  + Generated list of file locations and categories. This can be extracted from the output structure given in DREEGEXPORT
* ReliefF\_SVM.mat
  + Trained SVM on top 100 features identified by ReliefF
* LDA\_SVM.mat
  + Trained SVM using LDA feature selection.
* F\_DL.mat
  + Collection of deep learning networks trained on 100 ReliefF features.
* Chlocs2.mat
  + EEG channel locations
* Topo\_BasicNet.mat
  + Trained TMN
* TopoData2.mat
  + Local training/testing data
* SFnet.mat
  + Trained sensor fusion network
* AllFeatures2.mat
  + Calculated LDA Features
* RelieffScores.mat
  + Calculated ReliefF Features

Classes:

* MdlResults
  + Shared functions for ML/DL models
* TopoDatastore
  + Custom datastore to read in data from file structure
* ResampleDatastore
  + Custom datastore to read in data from file structure