DDT: Dementia Detection Tool

Team members: Himanshu Khanchandani, Clark Butler, Cisil Karaguzel, Selman Ipek, Shreya Shukla
GitHub: https://github.com/HimanshuKhanchandani/Dementia-Detection-Tool

Overview:
Alzheimer's and Frontotemporal Dementia are neurogenerative disorders and are among the most common types of Dementia. Electroencephalography (EEG), which is an non invasive technique to measure the electrical activity in brain using external electrodes, may provide improved diagnosis of the disease.

Goals: To develop a robust machine learning model which predicts Alzheimer's disease accurately using EEG data. And to discover specific EEG markers that distinguish Alzheimer's patients from healthy individuals.

Stakeholders: Hospitals and labs for non-invasive screening of Alzheimer's.

KPI: Prediction accuracy in distinguishing Alzheimer's from healthy patients.

Dataset and features:
The dataset for the project is publicly available at OpenNeuroDatasets/ds004504 and may be downloaded using datalad. The data consists of resting state eyes closed EEG of 88 subjects out of which 36 have Alzheimer's, 23 have frontotemporal dementia and the other 29 are healthy subjects. The EEG was recorded from 19 scalp electrodes for 10- 13 min for each subject with a sampling frequency of 500 Hz, and is available as a .set file. This is a standard size of a dataset in the EEG literature.

To extract **features**, we divide the EEG of each subject into smaller epochs of 2 minutes each with fifty percent overlap amongst epochs. The features consist of power spectrum of each epoch in each channel integrated over six frequency bands namely : (0.5- 4 Hz), (4.0 - 7.0 Hz), (7.0 - 9.0 Hz), (11.0- 13.0 Hz), (13.0 - 25.0 Hz). To be more precise, the features are relative band powers in these frequency bands. Based on cross-validation we chose to use 15 out of the original 19 channels, so there are in total 90 features (15 channels for six frequency bands).

Results:
After conducting cross-validation with different classification algorithms, we found that a support vector machine classifier with a linear kernel achieves the highest validation accuracy. It gives a validation accuracy of 89% and test set accuracy of 81% for Alzheimer's vs healthy subject and outperforms all the existing models in the literature. Looking at the coefficients, the model also tells us the feature importances of various bands and channels, which will be helpful in future experiment designs. In particular, O2 channel (located at the rear of the scalp) in the frequency band of (7.0 - 9.0 Hz) is the most important contributor in this classification problem.

We have also hosted the trained model on a web-based app on Hugging Face . It takes in a EEG file and uses our model to predict the probability that the EEG belongs to a patient with Alzheimer's.

Future directions While relative band power proved to be an useful feature, in future we would like to combine it with other features such as spectral connectivity which measures coherence between different regions of the brain. This could help in further improving the accuracy.

Another related problem is to build a classifier to distinguish Alzheimer's from other types of dementia such as Frontotemporal dementia. This is useful because treatment and diagnostic methods of different types of dementia are different. Our approach will be useful in building classifiers for this important problem.