PROJECT PROPOSAL – 844 Seminar in Computational Biomedicine

Mauricio Ferrato

Mayara Fernanda Gimenes de Souza

Parkinson's disease (PD) is a progressive disorder of the central nervous system affecting movement and inducing tremors and stiffness. PD is a chronic condition without cure that affects more than 1 million people in North America only. It is a neurodegenerative disorder affecting dopamine-producing neurons in the brain. For several people with Parkinson's (PWP) is required physical visits to clinics for treatment monitoring and most of these patients will eventually need clinical intervention and that is not practical for all these individuals. Research has show that approximately 90% of PWP have voice impairment, consequently we can detect in early stages the disease in a noninvasive a simpler way.

Our main objective in this study is to build a model to detect the presence of Parkinson's disease in an individual. We intend to use four different Machine Learning methods to identify the presence of PD and then compare the results between these methods. Support Vector Machine is the technique used in our referenced paper; hence we will check the results obtained on the paper. Random Forest and Neural network are decision-tree based algorithm and we will be able to compare results between them. Finally, we will introduce a fourth technique (XGBoost) to check its results toward the dataset presented.

The Dataset

The paper "Max A. Little, Patrick E. McSharry, Eric J. Hunter, Lorraine O. Ramig (2008), 'Suitability of dysphonia measurements for telemonitoring of Parkinson's disease', IEEE Transactions on Biomedical Engineering." uses speech signals to train an SVM model to decide if a patient suffers from PD. In order to train and validate our systems we will use a dataset that consist of 195 vowel phonations from 31 people, which 23 were diagnosed with PD.

The Machine Learning techniques we will apply to this data set are:

- Support Vector Machine (SVM)
    o SVM can directly measure the extent to which PD can be discriminated from healthy controls based on measures of dysphonia, addressing the problem of classifying subjects as healthy or PD. With such classification methods, it is also possible to combine measures to create more effective discrimination in practice.
- XGBoost (Gradient Boosted decision trees)
    o Boost is a Machine Learning algorithm designed with speed and performance in mind. XGBoost stands for eXtreme Gradient Boosting and is based on decision trees. In this project, we will import the XGBClassifier from the xgboost library.
- Random Forest
    o Random forest consists of a large number of individual decision trees that operate as an ensemble. Each individual tree in the random forest outputs a class prediction and the class with the most votes becomes the model's prediction. We will implement it through scikit-learn library.
- Neural Network
    o Neural Network is a set of algorithms that are designed to recognize patterns, interpreting sensory data through a machine perceptron, labeling and clustering the inputs. Keras library will be used since is a simple tool for constructing a neural network.

Reference:

1. Max A. Little, Patrick E. McSharry, Eric J. Hunter, Lorraine O. Ramig (2008), 'Suitability of dysphonia measurements for telemonitoring of Parkinson's disease', IEEE Transactions on Biomedical Engineering.