AML Project Proposal

Team: vv2372, sz3211, ys3748, nk3024, al4363

Problem Statement:

We will develop a machine learning model trained on data collected from a wearable 3D lower back sensor to detect freezing of gait (FOG), a debilitating symptom that afflicts many people with Parkinson's disease.

Background:

Parkinson's disease affects 7 to 10 million people globally, with many experiencing freezing of gait (FOG), a condition where their feet seem stuck during movement attempts. FOG significantly impairs quality of life, leading to depression, increased fall risk, wheelchair dependence, and reduced independence. Despite multiple theories on FOG, its causes remain unclear. Accurate quantification is crucial for understanding and treating FOG. Data science skills can aid in collecting and analyzing FOG events, potentially leading to treatments.Machine learning, particularly with lower back accelerometers, shows promise in accurate FOG detection.

Dataset:

We will be using the dataset from kaggle competition :
https://www.kaggle.com/competitions/tlvmc-parkinsons-freezing-gait-prediction/data

The dataset involves 3D accelerometer data from individuals experiencing freezing of gait (FOG) in Parkinson's disease. The goal is to identify the beginning and end of each freezing episode and categorize them into three types: Start Hesitation, Turn, and Walking.

Brief details of dataset:

- 1044 files (csv and parquet) of 70G
- 65 subjects
- 3 axis motion data AccV, AccML, and AccAP as primary input
- 3 output classes StartHesitation, Turn, Walking Indicator variables
- Age, sex, years since diagnosis, parkinson's score as additional information

Proposed Methods:

Since we are dealing with time-series signal data for classification, we will try to focus more on models like LSTM, 1D CNN, LGBM, XGB along with hyperparameter optimization using tools like Optuna and present a comparison of model performance. However, we also have data which are numerical and categorical values, so we will also explore ensemble approaches along with the time series models to check the effectiveness of such variables. In addition, we will also perform exploratory data analysis and try to identify dataset characteristics and important features and perform data cleaning and signal denoising to improve model performance.