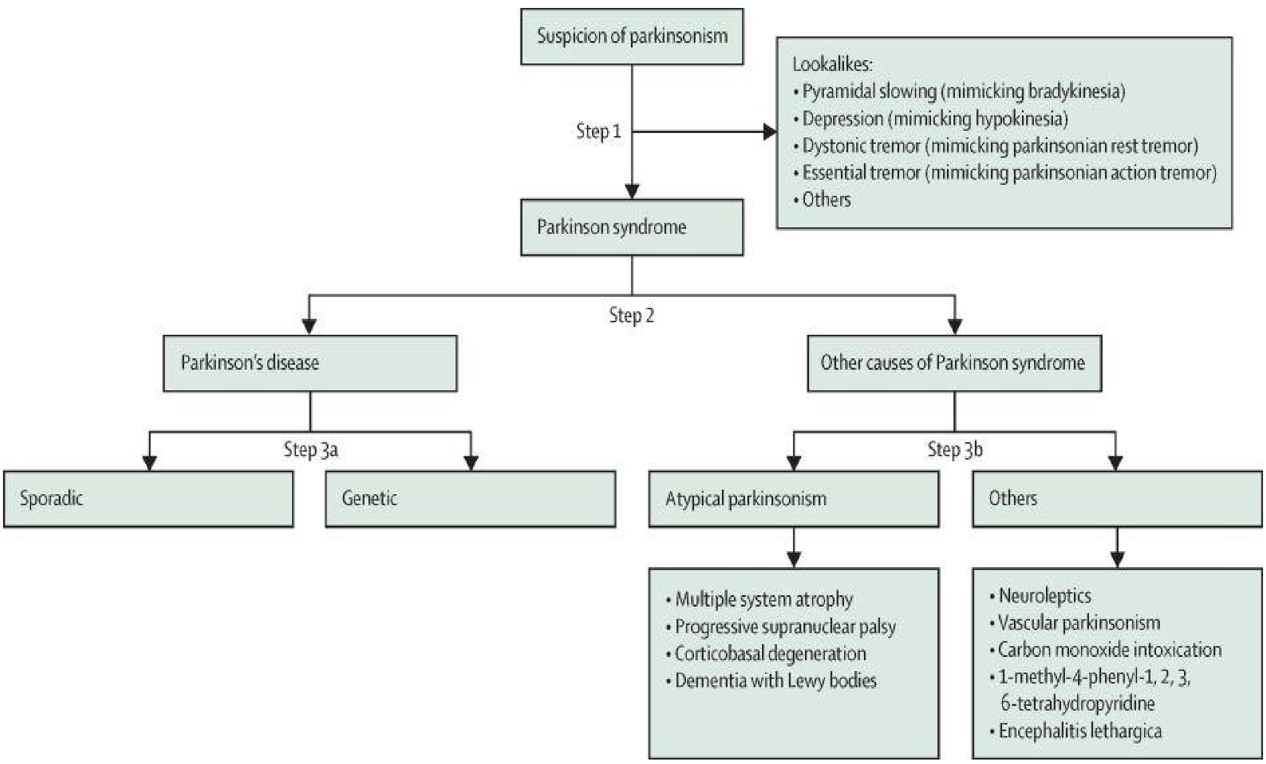
Parkinson's Disease Detection Using ML

Solution Architecture

Abstract: Parkinson's disease is a neurological disorders, causes damage to the underlying cognitive and neuro muscular function. PD is a progressive neuro-degenerative disorder that approximately affects about 5 million people around world, with approximately sixty thousand new clinical diagnoses made each year. Historically, PD has been difficult to quantify and doctors have tended to focus on some symptoms while ignoring others, relying primarily on subjective rating scales. Due to the decrease in motor control that is the hallmark of the disease, voice can be used as a means to detect and diagnose PD. With advancements in technology and the prevalence of audio collecting devices in daily lives, reliable models that can translate this audio data into a diagnostic tool for healthcare professionals would potentially provide diagnoses that are cheaper and more accurate. We provide evidence to validate this concept here using a voice dataset collected from people with and without PD. This paper explores the effectiveness of using supervised classification algorithms, such as deep neural networks, to accurately diagnose individuals with the disease. Our peak accuracy of 85% provided by the machine learning models exceed the average clinical diagnosis accuracy of non-experts (73.8%) and average accuracy of movement disorder specialists (79.6% without follow-up, 83.9% after follow-up) with pathological post-mortem examination as ground truth.

General Detection of PD through medical process:



The main deficits of PD speech are loss of intensity, monotony of pitch and loudness, reduced stress, inappropriate silences, short rushes of speech, variable rate, imprecise consonant articulation, and harsh and breathy voice (dysphonia). The range of voice related symptoms is promising for a potential detection tool because recording voice data is non-invasive and can be done easily with mobile devices.

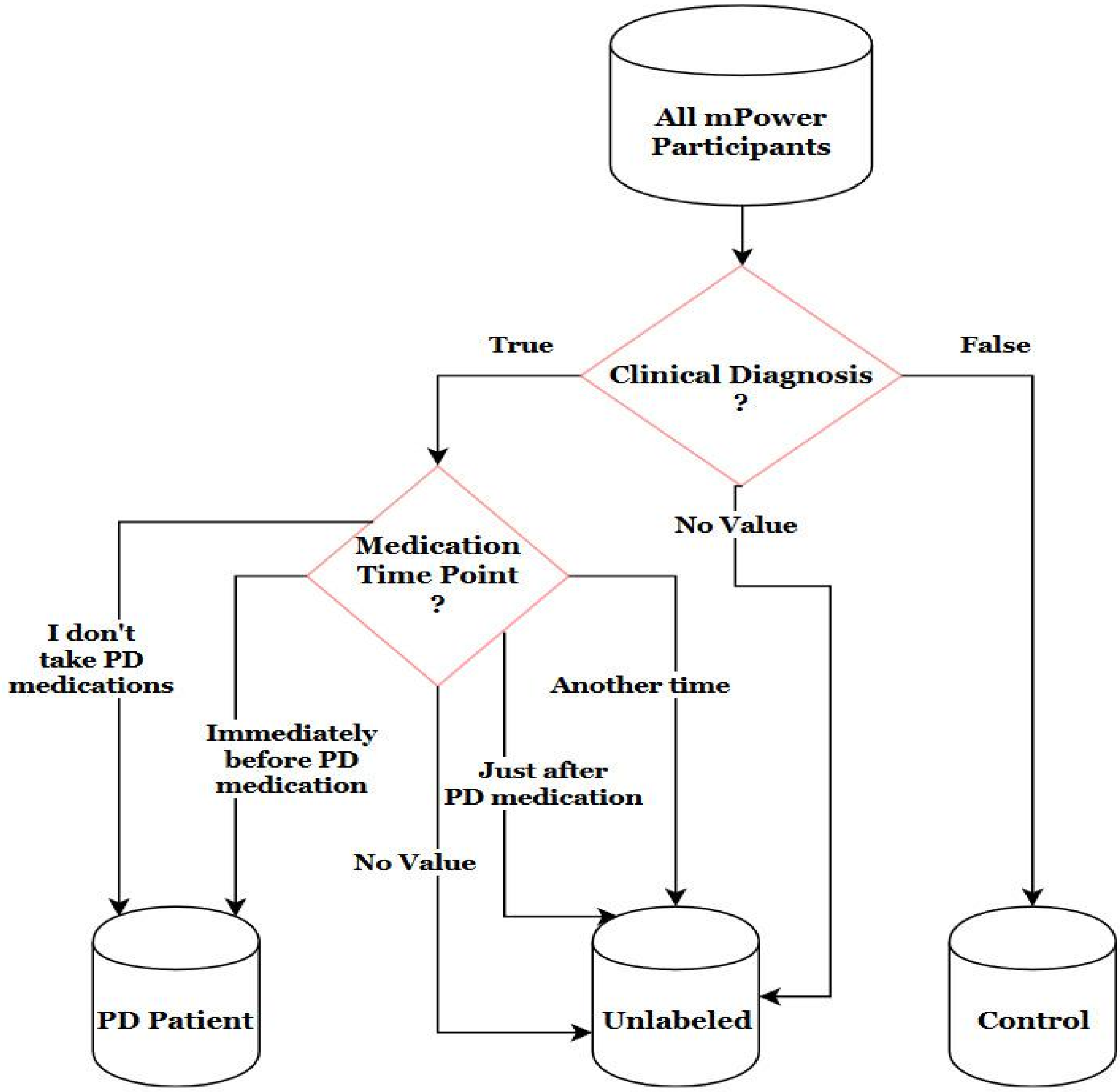
PD is difficult to detect early due to the subtle initial symptoms. There is a signifificant burden to patients and the health care system due to delays in diagnosis [8]. The difficulty in early PD diagnosis has inspired researchers to develop screening tools relying on automated algorithms to differentiate healthy controls from people with PD. This binary diagnosis focuses on the first step of validating digital biomarkers in distinguishing disease from control; it does not offer a form of differential diagnosis where the model may distinguish PD among a variety of disorders that present PD-like symptoms (e.g. Lewy-Body Dementia, Essential Tremor). The current research is a promising first step toward a long-term goal of providing a decision support algorithm for physicians in screening patients for PD [9]. In this paper, we apply several different machine learning models to classify PD from controls using the mPower Voice dataset.

During data collection, patients were asked to give information regarding when, relative to taking medication, they provided their data.These medication time points were interpreted to mean: time of best symptom control, on medication but not immediately before or after, time of worst symptoms, not on medications, and not applicable, respectively.

crossed with the clinical diagnosis responses from the demographics survey led to three groups of patients and data, as shown in Figure 2. Patients that had medication prior to the voice test were not used as participants in the analysis. The rationale for this parameter selection is that the voice of the patient will depict the most extreme effects of the PD without the effect of any medication. The assumption is that the voice features will be noticeably different from those of the controls. The control in this experiment is a participant who has not been professionally diagnosed with PD.

Each patient could contribute to multiple voice submissions, so the number of unique audio files exceeds the total number of patients surveyed. Based on the data extracted from these studies, a csv file was created that contained the demographics data linked with the health codes unique to each patient. The voice data was also pre-processed using the PyAudioAnalysis library in Python. This preliminary audio analysis resulted in eleven unique features as shown in Table IV in Supplementary Material.

Fig.2 Flowchart of medical process for detection of PD:



Working architecture:
Prior to being fed into the feature extraction algorithms, the raw audio was cleaned with VoiceBox's Voice Activation Detection (VAD) algorithm, activlev, [12] to extract and remove background noise of the audio.

This prepossessing step was required in order to pass only raw voice into the audio feature extraction algorithms. This cleaned audio was then passed through two separate algorithms for feature extraction before being input into the machine learning models were used for preliminary audio analysis and the method of Minimum Redundancy Maximum Relevance (mRMR) was applied to these audio features. mRMR extracts the most relevant features of a given dataset with respect to an output class, while minimizing the redundancy.

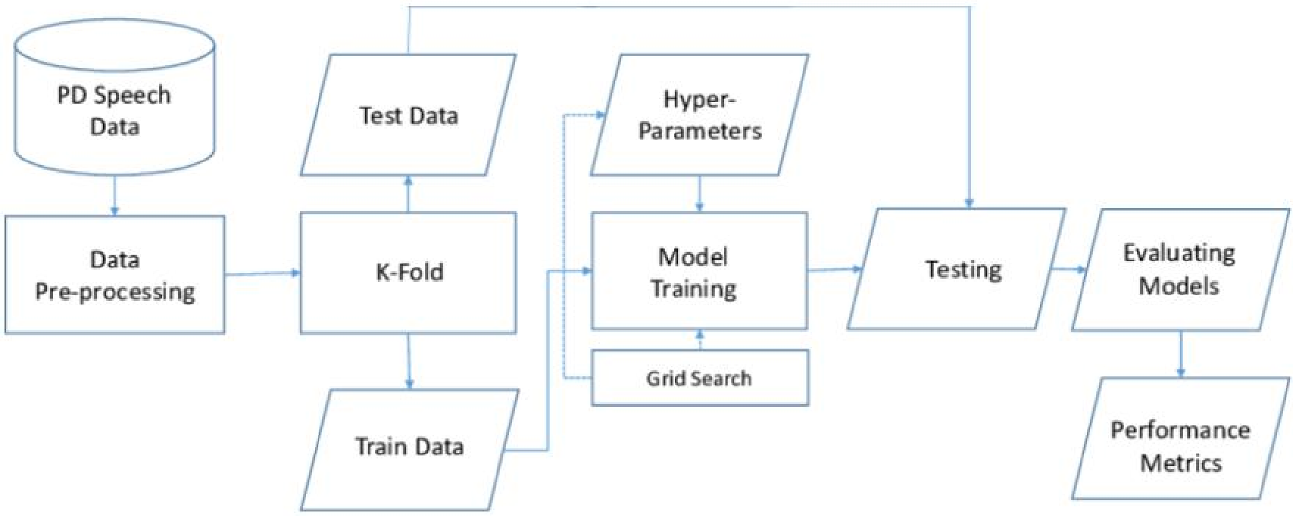
The mRMR technique yielded an array of ranked features indexed from highest to lowest predictive correlation on the labeled data. The ranked feature indexes were then used to further pre-process the data before being fed into the machine learning models (e.g. random forest, support vector machine etc.). The accuracy of the models on the testing set were assessed using varying lengths of the extracted features. Features of size 3, 4, 5, 10, 15, 20, 40, 80, 100, 200, 400, 800, 1000, 1200, 1500, 2000, 2200 were used. 1200 features offered the best categorical accuracy with all classifiers outperforming other baselines of less features on each model. The raw audio was also passed into the algorithm that extracted GeMaps using the openSMILE toolkit [16] for feature ex
traction before being sent to the machine learning models.

The GeMaps feature algorithm extracts a number of lower level features such as pitch, jitter, shimmer, loudness, and harmonics-to-noise ratio, in addition to temporal features, such as rate of loudness and number of continuous voiced regions per second. In total, this analysis yielded 62 features per audio sample.

1.A diverse range of machine learning classifiers were examined to find the highest categorical accuracy for PD diagnosis. The decision tree and support vector machine classifiers were developed
2. Algorithm for PD Diagnosis machine learning library as well as the TensorFlow and Keras Deep Learning Libraries. Models were optimized through stratified cross validation with accuracy, F-1, recall and precision as metrics.
3.A series of decision tree classifiers were used to classify the dataset including standard decision trees, random forest, gradient boosted decision trees and extra tree classifiers. A decision tree operates by creating binary decision boundaries about features to separate the data homogeneously between the two classes by using a metric that minimizes information entropy. In aggregate, these separations create a classification accuracy over the training set that is the applied to the testing set to assess generalization.
4.Random forests are an extension of decision trees that use arbitrary mixing of the data to create different subsets of the training data which are then run through decision tree models.
5.These models are then tested for accuracy for samples not used in the sub trees and parameters are tuned to maximize the expected accuracy of the model over the training set.
6.Extra tree classifiers are another variation of decision tree classifiers that rely on stochastic methods that create shallower but wider decision trees. Gradient boosted decision trees work by creating simple poor classifiers that divide the sample space.
7.The poor classifiers are combined to minimize a differentiable loss function [27]. The algorithm iteratively modifies the previous classification state by creating another classifier for the training set.

8.This process is repeated to produce an ensemble of classifiers that are able to classify the training set accurately.
9.Another popular and powerful classifier is the Support Vector Machine (SVM). SVMs, much like logistic regression, aim to construct an optimal separating hyperplane in the feature space between the two classes.

Block Diagram for Parkinson's disease detection:



Algorithm for PD diagnosis: