

Define the problem - two-decision classification problem.

(Brief overview charted out before the start of the project)

What is Parkinson?

Parkinson's disease - the shaking Palsy

A disorder of the central nervous system that affects movement, often including tremors.

Nerve cell damage in the brain causes dopamine levels to drop, leading to the symptoms of Parkinson's.

Parkinson's Disease is the second most prevalent neurodegenerative disorder after Alzheimer's, affecting more than 10 million people worldwide. Parkinson's is characterized primarily by the deterioration of motor and cognitive ability.

Parkinson's often starts with a tremor in one hand. Other symptoms are slow movement, stiffness and loss of balance.

Medication can help control the symptoms of Parkinson's.

The four main symptoms of Parkinson's are tremor, which means shaking or trembling. Tremor may affect your hands, arms, legs, or head: stiff muscles, slow movement and problems with balance or walking. Other symptoms may include depression and other emotional changes; difficulty in swallowing, chewing, and speaking; urinary problems or constipation; skin problems; and sleep disruptions. It generally affects people over the age of 50.

What do I plan to do with this?

Detecting the early onset of Parkinson's disease is of utmost importance in managing the condition effectively.

The data used for this analysis was obtained from Kaggle and includes a "Status" column, which serves as a binary category indicating whether an individual has the disease or not.

This project revolves around a classification task, where we aim to explore how various attributes influence the disease status. By visualizing and preparing the data, we strive to develop a model that can potentially identify early signs of Parkinson's disease at home. While this may be ambitious, if the model can accurately indicate when it is necessary to consult a doctor, it would be a significant achievement.

In addition, I intend to provide basic guidelines on how individuals can identify potential early signs of Parkinson's disease at home, differentiating them from potential side effects of

medications or other factors. This can be done further in the project when voice extraction is worked on and the features are delved in depth.

Why this data?

There is no single test which can be administered for diagnosis. Instead, doctors must perform a careful clinical analysis of the patient's medical history.

Unfortunately, this method of diagnosis is highly inaccurate. A study from the National Institute of Neurological Disorders finds that early diagnosis (having symptoms for 5 years or less) is only 53% accurate. This is not much better than random guessing, but an early diagnosis is critical to effective treatment.

Why speech features? Speech is very predictive and characteristic of Parkinson's disease; almost every Parkinson's patient experiences severe vocal degradation (inability to produce sustained phonations, tremor, hoarseness), so it makes sense to use voice to diagnose the disease. Voice analysis gives the added benefit of being non-invasive, inexpensive, and very easy to extract clinically.

Metadata:

The dataset provided includes two files: "parkinsons.csv" and "parkinson names," which provides a brief description of the data.

The dataset consists of 195 instances of sustained vowel phonation from 31 individuals, with 23 of them diagnosed with Parkinson's disease [5, 7, 8]. The primary objective of data processing is to distinguish between healthy individuals and those with Parkinson's disease.

The classification problem is based on the "status" attribute, where "non-PD" represents healthy individuals and "PD" denotes people with Parkinson's disease.

Each row in the CSV file corresponds to a voice recording instance, with approximately six recordings per patient. The first column contains the patient's name for identification purposes.

What does success in this project mean?

Success in this project is defined by avoiding overestimation. While classifying a non-PD person as PD is acceptable, misclassifying a PD person as non-PD would be highly dangerous. Therefore, the focus is on minimizing false negatives.

Given the imbalanced nature of the dataset, evaluation metrics such as Cohen Kappa score, AUC ROC, and average precision score were chosen to assess the model's performance.

Parkinson's disease affects the central nervous system (CNS) of the brain and currently has no cure. Early detection is crucial, as late detection often leads to a lack of treatment and loss of life. Therefore, early detection plays a vital role in managing the disease effectively.