

Parkinson's disease (PD) is a neurodegenerative disorder primarily affecting movement. At first, we import the necessary libraries, including matplotlib.pyplot for creating visual representations of data and pandas for managing and manipulating data. After that, we load the dataset "PO1_DATA.CSV," which consists of the voice metrics to differentiate between the healthy and Parkinson's disease person. After loading the dataset from a specified path, column headers are assigned to ensure the clarity of analysis. After organizing the data correctly, an initial exploration is conducted to obtain a statistical summary, such as mean, median, and standard deviation, to ensure the accuracy of the analysis. The core of the analysis involves comparing the voice metrics between a healthy person and a person affected by Parkinson's disease. For each voice metric, we conduct a statistical test called a t-test to compare the means of two groups and determine the p-value. The outcomes of these t-tests, along with the p-value and t-statistic, are compiled in a DataFrame and a function is added that flags whether a metric's difference is statistically significant based on a typical p-value threshold of 0.05. Finally, histograms are plotted for all significant voice metrics to visualize these differences.