prevalence of patients with Parkinson's disease in the dataset and identify abnormalities that will affect model training and evaluation. Fig. 1. Comparison of Data Distribution by Count Correlation Analysis: Analyzing the correlation between the different features and the target variable values (status 1 and 0) and finding out features that have a very strong correlation, because they can play an important role in the diagnostic process. Fig. 2. Correlation Heatmap of features Fig. 3. Proposed architecture The flowchart in Fig. 3. represents a pipeline for a machine learning approach on multi – classification models that can predict Parkinson's disease. 1. Parkinson's Disease Dataset: This process begins with the acquisition of the dataset for Parkinson's disease. The dataset contains features such as voice recordings and other form of clinical data relevant to Parkinson's symptoms. 2. Feature Extraction: Here, features are extracted from the dataset by selecting only the most relevant attributes for 8202/10000 Characters (i) 2 credits left û 



