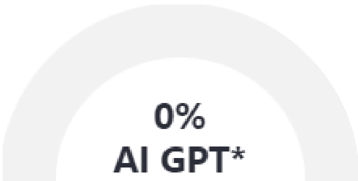
Your Text is Human written



Abstract— Parkinson's disease (PD) is one such progressive neurodegenerative disease that affects millions of people around the world, and it is still critical to be diagnosed early for effective treatment. In the last few years, numerous novel types of machine learning approaches have been developed for PD diagnosis using data represented in the form of speech patterns, handwriting samples, sensor data, and many more. However, with this, the performance of such algorithms has significant variation based on the kind of data and features chosen.

In this paper, we have done the performance comparison of a number of machine learning algorithms in order to be used for Parkinson's disease. The dataset for this study was acquired from Oxford UCI Machine repository. We obtained the dataset for the study from the Oxford UCI Machine repository. After preprocessing the data and extracting relevant features, we applied various algorithms, including logistic regression, support vector machines, random forests to classify individuals as either PD patients or healthy controls.

The performance evaluation parameters, including accuracy, precision, recall, F1 score, and Precision-Recall curve (PR curve), were used to compare the algorithms. Our findings revealed that random forests

Your Text is Human written

0%
AI GPT*

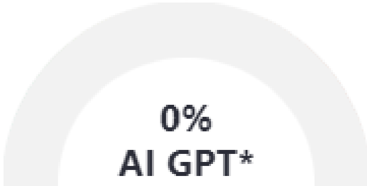
Parkinson's disease has destroyed the lives of 10 million people around the world and is the second most deadly neurodegenerative disease after Alzheimer's disease. The symptoms include: "frozen" facial expression, bradykinesia or slowness of movement, akinesia or impairment of voluntary movement, tremor, and impairment of the voice. By the time a diagnosis is made, typically 60% of nigrostriatal neurons have degenerated, as does 80% of striatal dopamine.

There is no single test in which this condition can be diagnosed. Doctors must perform careful clinical analysis of the patient's medical history. However, this result was inaccurate. According to the National Institute of Neurological Disorders, early diagnosis (with symptoms for ≤ 5 years) is only 53% accurate. This is hardly much better than expected, but early diagnosis makes all the differences in the world for effective treatment.

These problems motivated us to investigate a machine learning approach to accurately diagnose Parkinson's, using a dataset of various speech features (a non-invasive yet characteristic tool) from the University of Oxford.

The purpose of this project is to discover a machine learning technique that can be effectively used for the prediction of Parkinson's disease using relevant data. This study will conduct a comparative analysis of feature selection and representation techniques to identify the most relevant and enlightening features from the available data useful in the treatment of patients.

Your Text is Human written



Dataset

We have collected the required Parkinson's disease dataset, created by the University of Oxford, called the Oxford Parkinson's Disease Detection dataset, containing 197 recordings and 23 features, created by Max Little from the University of Oxford, along with the National Centre for Voice and Speech, Denver, Colorado, who recorded the speech signals.

The data contains biomedical voice recordings from 31 individuals, 23 of whom had Parkinson's disease (PD). Each column in the table is a voice measurement, and each row corresponds to one of the 195 voices recorded by that individual ("name row"). For patients with Parkinson's disease, the "Status" column is set to 0 for healthy patients and 1 for Parkinson's disease. The file is in ASCII CSV format. The rows of the CSV file contain examples for a recording. There are approximately six entries for each patient, and the patient's name is identified in the first row.

Overview of Vocal Analysis Attributes and Their Purposes

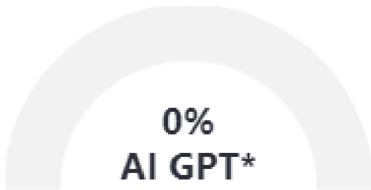
Attribute

Purpose

Name

Data is stored in ASCII CSV format where patient name and recording number is stored

Your Text is Human written



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.

Correlation Heatmap of features

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.

Model Training:

Several machine learning models are implemented within the pipeline. They are:

LR (Logistic Regression)

DT (Decision Tree)

RF (Random Forest)

SVM (Support Vector Machine)

KNN (K-Nearest Neighbors)

Your Text is Human written

0%
AI GPT*

The classification accuracy obtained on the basis of vowel phonation data of Parkinson's disease is calculated to be 98.3051%, for the Random Forest classifier. The results obtained from the Random Forest model are pretty robust and are based on its exceptional aptitude in the representation of complex data structures and associations. Yet another reason for good performance is that Random Forest treats all 22 attributes in the MDVP (Multidimensional Voice Program) dataset equally important. In other words, it gives a fair consideration to each independent vocal attribute without showing any bias toward any feature. This is very crucial since Parkinson's disease is usually heralded by slight changes in speech; it is an expression of a set of characteristics that might possibly allow for a correct diagnosis as against any single characteristic.

Its strength and accuracy make the Random Forest classifier one of the models to predict the existence of Parkinson's disease from the vowel phonation data. The high accuracy of the model and the noninvasive aspect of voice analysis make this approach highly practical for real-world applications. This model may be integrated into long-term health monitoring systems, ensuring that PWP are under continuous and reliable diagnosis. The integration will offer an easy, inexpensive, and accessible tool for the control of diseases, bearing a lot of benefits to patients worldwide regarding earlier detection, monitoring of disease progression, and subsequent tailoring of treatment plans.

We thus advocate that the Random Forest model be given particular capabilities and used as a transforming agent to deliver relief, on a long-term basis, thereby enhancing the quality of life for Parkinson's patients worldwide.

Highlighted text is suspected to be most likely generated by AI*

Export to PDF

1,775 Characters
267 Words