Parkinson's Disease Detection

Parth Singh Bargoti , Richa Agrawal, Sayyam Aneja

Acropolis Institute of Technology and Research parthbargotics18@acropolis.in richaagrawalcs18@acropolis.in sayyamanejacs18@acropolis.in

ABSTRACT: -

Classification of recorded acoustic signals dataset to represent if an individual is presenting early onset of Parkinson's Disease, using XGBClassifier from python library XGBoost.

Key-Words: - python, scikit-learn, numpy, pandas, xgboost, xgbclassifier, parkinson's disease, machine learning.

1.Introduction

Parkinson's Disease Detection is a machine learning model built using python. This Model allows us to classify recorded acoustic signals dataset, to represent if an individual is presenting early signs of parkinson's disease.This model can be used by medical professionals to determine if the patient is showing early onset of Parkinson's Disease.

2.ProblemFormulation

Parkinson's disease is a neurological movement disorder. Common symptoms include tremor, slowness of movement, stiff muscles, unsteady walk and balance and coordination problems. There is no cure for the disease. Most patients can maintain a good quality of life with medications. In some patients, surgery can help improve symptoms.

Parkinson's disease occurs when nerve cells (neurons) in an area of the brain called the substantia nigra become impaired or die. These cells normally produce dopamine, a chemical (neurotransmitter) that helps the cells of the brain communicate (transmits signals, "messages," between areas in the brain). When these nerve cells become impaired or die, they produce less dopamine. Dopamine is especially important for the operation of another area of the brain called the basal ganglia. This area of the brain is responsible for organizing the brain's commands for body movement. The loss of dopamine causes the movement symptoms seen in people with Parkinson's disease.

3.Literature Review

In existing system, PD is detected at the secondary stage only (Dopamine deficiency) which leads to medical challenges. Also, doctor has to manually examine and suggest medical diagnosis in which the symptoms might vary from person to person so suggesting medicine is also a challenge. Thus the mental disorders are been poorly characterized and have many health complications. PD is generally diagnosed with the following clinical methods as,

- MRI or CT scan - Conventional MRI cannot detect early signs of Parkinson's disease
- PET scan - is used to assess activity and function of brain regions involved in movement
- SPECT scan - can reveal changes in brain chemistry, such as a decrease in dopamine

This results in a high misdiagnosis rate (up to 25% by non-specialists) and many years before diagnosis, people can have the disease. Thus existing system is not effective in early prediction and accurate medicinal diagnosis to the affected people.

Methodology

Python Libraries such as numpy, pandas, scikit-learn(sklearn) were used alongside xgboost.

1.   *Numpy*

NumPy is a Python library used for working with arrays.It also has functions for working in domain of linear algebra, fourier transform, and matrices.NumPy was created in 2005 by Travis Oliphant. It is an open source project and you can use it freely.NumPy stands for Numerical Python. In Python we have lists that serve the purpose of arrays, but they are slow to process. NumPy aims to provide an array object that is up to 50x faster than traditional Python lists.

2.   *Pandas*

Pandas is an open-source, BSD-licensed Python library providing high-performance, easy-to-use data structures and data analysis tools for the Python programming language. Python with Pandas is used in a wide range of fields including academic and commercial domains including finance, economics, Statistics, analytics, etc.

3.   *XGBoost*

XGBoost is an optimized distributed gradient boosting library designed to be highly efficient, flexible and portable. It implements machine learning algorithms under the Gradient Boosting framework. XGBoost provides a parallel tree boosting (also known as GBDT, GBM) that solve many data science problems in a fast and accurate way. The same code runs on major distributed environment (Hadoop, SGE, MPI) and can solve problems beyond billions of examples.

4.   *Scikit-Learn(Sklearn)*

Scikit-learn (Sklearn) is the most useful and robust library for machine learning in Python. It provides a selection of efficient tools for machine learning and statistical modeling including classification, regression, clustering and dimensionality reduction via a consistence interface in Python. This library, which is largely written in Python, is built upon NumPy, SciPy and Matplotlib.

The model uses XGBClassifier from XGBoost to classify the given dataset.

Table 1. List of measurement methods applied to acoustic signals recorded from each subject.

Name	ASCII subject name and recording number
MDVP:Fo (Hz)	Average vocal fundamental frequency
MDVP:Fhi (Hz)	Maximum vocal fundamental frequency
MDVP:Flo (Hz)	Minimum vocal fundamental frequency
MDVP:Jitter (%)	
MDVP:Jitter (Abs)	
MDVP:RAP	Five measures of variation in fundamental frequency
MDVP:PPQ	
Jitter:DDP	
MDVP:Shimmer	
MDVP:Shimmer (dB)	
Shimmer:APQ3	
Shimmer:APQ5	Six measures of variation in amplitude
MDVP:APQ	
Shimmer:DDA	
NHR	
HNR	Two measures of ratio of noise to tonal components in the voice
RPDE	
D2	Two nonlinear dynamical complexity measures
DFA	Signal fractal scaling exponent
Spread1	
Spread2	Three nonlinear measures of fundamental frequency variation
PPE	
Status	Health status of the subject: one, Parkinson's; zero, healthy

Result Discussions

Through this Python machine learning project, we aimed to detect the presence of Parkinson's Disease in individuals using various factors. We used an XGBClassifier for this and made use of the sklearn library to prepare the dataset. In this Python machine learning project, using the python libraries scikit-learn, numpy, pandas, and xgboost, we will build a model using an XGBClassifier. We'll load the data, get the features and labels, scale the features, then split the dataset, build an XGBClassifier, and then calculate the accuracy of our model.

Conclusion

This model predicts Parkinson's Disease through machine learning technique. This model provides promising results by analysis. This proposed project aims to provide machine learning technique to predict and provide early diagnostics for Parkinson's Disease. In this project the voice input is analyzed using RStudio using python. Thus, this proposed system emphasize the importance of early detection and prediction of Parkinson's Disease, such that treatment and support can be provided to patients as soon as possible.

References

[1] https://my.clevelandclinic.org/health/diseases/8525-parkinsons-disease-an-overview.

[2] https://www.hindawi.com/journals/cmmm/2016/1267919/tab1/

[3] http://ijariie.com/AdminUploadPdf/EARLY_DETECTION_OF_PARKINSON_S_DISEASE_USING_MACHINE_L EARNING_ijariie11591.pdf

[4] https://www.tutorialspoint.com/python_pandas/index.htm

[5] https://www.w3schools.com/python/numpy/numpy_intro.asp

[6] https://xgboost.readthedocs.io/en/latest/

[7] https://www.tutorialspoint.com/scikit_learn/index.htm

Authors

First Author – Parth Singh Bargoti (0827CS181139), Student, Acropolis Institute of Technology and Research.

Second Author – Richa Agrawal (0827CS181166), Student, Acropolis Institute of Technology and Research.

Third Author – Sayyam Aneja (0827CS181188), Student, Acropolis Institute of Technology and Research.