LITERATURE SURVEY

Detecting Parkinson's Disease Using Machine Learning

Date	19 October 2022
Team ID	PNT2022TMID51528
Project Name	Detecting Parkinson's Disease Using Machine Learning
Maximum Marks	2 Marks

Abstract — Identification of the correct biomarkers with respect to particular health issues and detection of the same is of paramount importance for the development ofclinical decision support systems. For the patients suffering from Parkinson's Disease(PD), it has been duly observed that impairment in the handwriting is directly proportional to the severity of the disease. Also, the speed and pressure applied to the pen while sketching or writing something are also much lower in patients suffering from Parkinson's disease. Therefore, correctly identifying such biomarkers accurately and precisely at the onset of the disease will lead to a better clinical diagnosis. Therefore,we planned to design a system which analyzes Spiral drawing patterns and wave drawing patterns in patients suffering from Parkinson's disease and healthy subjects.The system developed in the study leverages Histogram of Oriented Gradients (HOG)image descriptor for analyzing the drawing patters of both spiral and wave sketchesrespectively. Further, the prediction probabilities are trained on a random forest classifier based on ensemble voting to provide a weighted prediction from both the spiral and wave sketch.

INTRODUCTION

Parkinson's disease (PD) is one of the most common neurodegenerative diseases with a prevalence rate of 1% in the population above 60 years old, affecting 1 - 2 per 1000 people. The estimated global population affected by PD has more than doubled from 1990 to 2016 (from 2.5 million to 6.1 million), which is a result of increased number of elderly people and age-standardized prevalence rates. PD is a progressive neurological disorder associated with motor and non-motor features which comprises multiple aspects of movements,including planning, initiation and execution. The diagnosis of PD is traditionally based on motor symptoms. Despite the establishment of cardinal signs of PD in clinical assessments, most of the rating scales used in the evaluation of disease severity have not been fully evaluatedand validated. This section of the describes the theoretical background of this project, starting with anexplanation of Parkinson's disease, followed by overviews of machine learning, deep learning, related work and finally PD diagnosis problems. The detection of PD is extremely important at the first stage. This entails the literature survey that was conducted for the purpose of better understanding the problem at hand and to explore possible solutions.

Importance of Voice data:

Speech or voice data is assumed to be 90% helpful to diagnose a person for identifying presence of disease. In general, Person with PD suffer from speech problems, which can be categorized into two: hypophonia and dysarthria. Hypophonia indicates very soft and weak voice from a person and dysarthria indicate slow speech or voice, that can hardly be understood at one time and this causes because of damage to central nervous system.

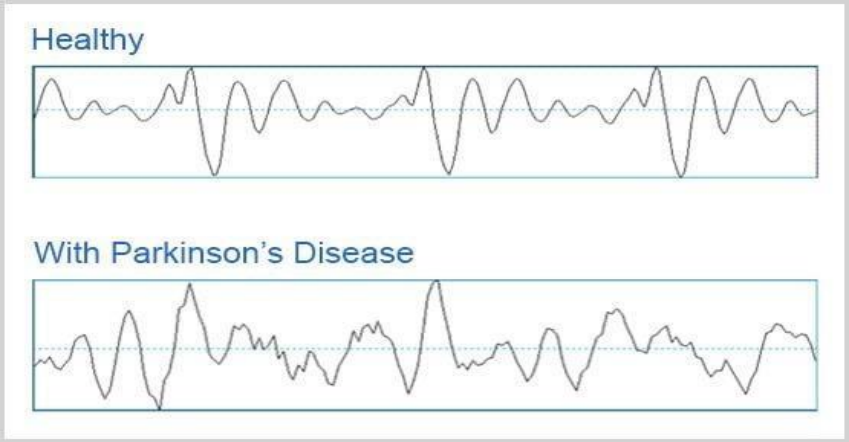

Fig 1.1 Voice data analysis of affected vs healthy person

Literature survey:

Parkinson's disease (PD) is known to be one of the most common neurodegenerative diseases among older people aged more than 65. Since this diseaseis progressive in nature, negligence in the diagnosis of this disease in the early stage and monitoring at different stages would create a severe negative impact on the
patients in terms of healthcare costs as well as the severe health-related disorders.Some of the movement disorders symptoms such as rigidity, instability in posture,
tremor, and bradykinesia are usually observed on the PD patients at different stages . Toprevent the major negative impact on PD patient's it is necessary to detect the PD at theearly stage. One of the most common effects that are easily noticeable among the PD
patients and used most commonly in the early stage of diagnosis is finding the
difference in handwriting and sketching abilities. The non-invasive measures such assketching of a shape such as spiral, waves, and other handwritten texts could be easilydistinguished from one person to another person as well as a person with PD and a person without PD.Previous researchers and clinicians already found some kind of association between the sketching of the spirals and handwriting in the early stages ofthe PD. However, the major drawback of these kinds of diagnoses needs proper interpretation of sketching and handwriting. Traditionally the sketching or handwritings
were performed in the papers and interpreted manually by the interpreters specialized inthose fields. With the availability of digital devices, it is easier to perform those tasks digitally as well as the assessments were done by the machine in a more precise and accurate manner compared to the traditional ways. Some of the common features
present inside the sketches could be considered as the potential indicators to
differentiate different group of subjects that includes healthy subjects and PD subjects and those tasks can be used to perform the reliability analysis in the real-time. In recentdays the wisest decision for detecting something in real-time is to make the system automatic so that we can perform the same operation with less time as well as in a more precise way. In this respect machine learning techniques are more effective and shown enough potential to be used in real-life situations .

So,we made an attempt has been made to develop an automated system that
trained with the features extracted from the different sketches performed by the healthygroup of patients as well as PD patients to assess the severity of the PD disease amongdifferent stages as well as between the healthy groups of patients. The investigation performed in this study to differentiate the healthy subjects from PD subjects based onthe spiral sketches by extracting features from the images sketched by the healthy subjects and PD subjects.

Some of the studies related to the implementation of the machine learning techniques for the development of an automated system using different datasetsrelated to Parkinson's disease has been discussed here.

Zham et al. proposed a study that used two criteria such as speed, and pen-pressure while performing the sketches to distinguish PD subjects at different stages. They have extracted features from the sketches and proposed a method that can provide a correlation factor between the features and severity level of the PD. Finally, they performed the Mann-Whitney test to validate the study that these methods can beused to distinguish different stages of the PD. They observed that there was a significant difference in the correlation factor at different stages of PD .

Kotsavasiloglou et al., presented an investigation based on the trajectory of the tip of the pen on the surface of the pad while drawing simple horizontal lines by the healthy subjects and PD subjects. They extracted features from the simple drawings and trained the machine learning algorithms using those features to distinguish the PD subjects from the healthy subjects. They have used different classifiers such as Naïve Bayes, AdaBoost, Logistic Regression, J48, Support Vector Machine (SVM), and Random Forest classifiers to train the features for developing an automated system. The performance metrics used in that study were accuracy, Area under the curve (AUC),True positives (TP), and True negatives (TN). They achieved an accuracy of 91%, and TPof 0.88 and TN of 0.95.

Memedi et al. proposed a study based on the spiral data collected using telemetry touch screen devices in the home environments to distinguish off episodesand peak dose dyskinesia using machine learning algorithms. Several features were extracted from the data and used as input to the machine learning classifiers. They haveused Support Vector Machine, Logistic Regression, Random Forest, and Multi-Layer Perceptron (MLP) to train the features for the development of automated systems andfound MLP performed well among all the classifiers with an accuracy of 84% .

Aich et al. proposed a study that used a voice dataset to distinguish PD patientsfrom others. They have used different feature selection techniques to find the best features that can be used to train different classifiers. The feature selection techniquesused for that study is principal component analysis. Finally, a performance comparison study was performed using two groups of datasets such as original feature sets and PCA based feature sets using nonlinear decision tree-based classifiers. It was foundthat the

random forest classifier (RFC) was the best classifier among them and PCAbased feature set performance is better than the original feature sets. The best accuracy of 96.83% was found with RF classifier and PCA based feature sets .

References

[1] Zham, P., Arjunan, S.P., Raghav, S. and Kumar, D.K., 2017. Efficacy of guided spiraldrawing in the classification of Parkinson's disease. IEEE journal of biomedical and health informatics, 22(5), pp.1648-1652.

[2] Kotsavasiloglou, C., Kostikis, N., Hristu-Varsakelis, D. and Arnaoutoglou, M., 2017. Machine learning-based classification of simple drawing movements in Parkinson's disease. Biomedical Signal Processing and Control, 31, pp.174-180.

[3] Memedi, M., Sadikov, A., Groznik, V., Žabkar, J., Možina, M., Bergquist, F., Johansson, A., Haubenberger, D. and Nyholm, D., 2015. Automatic spiral analysis for objective assessment of motor symptoms in Parkinson's disease. Sensors, 15(9), pp.23727-23744.

[4] Aich, S., Sain, M., Park, J., Choi, K.W. and Kim, H.C., 2017, November. A mixed classification approach for the prediction of Parkinson's disease