2019 5th International Conference for Convergence in Technology (I2CT)
Pune, India. Mar 29-31, 2019

Detecting Parkinsonian Symptoms using Data Analysis

Aaditya Naik
B. Tech., Computer Engineering
NMIMS MPSTME
Mumbai, India
aadityanaik.nmims@gmail.com

Utkarsh Pant
B. *Tech., Computer Engineering*
NMIMS MPSTME
Mumbai, India
utkarshpant.nmims@gmail.com

Aditya Pansare
B. Tech., Computer Engineering
NMIMS MPSTME
Mumbai, India
adityapansare.nmims@gmail.com

Krishna Samdani
Computer Engineering
NMIMS MPSTME
Mumbai, India
krishna.samdani@nmims.edu

***Abstract*— Parkinson's Disease (PD) is one of the most common neurodegenerative diseases. In this paper, we have researched the various methods used to detect major symptoms of Parkinson's Disease (PD), namely resting tremors, impaired gait and vocal impairments. We have conducted an extensive literature review on the automated detection of each of the aforementioned symptoms. Each symptom is discussed in a separate section which covers the data mining and classification techniques for the respective symptom and weighs the advantages and disadvantages of each method. These techniques are then compared to find the most useful, convenient and easy-to-use process which could be easily implemented in a normal medical checkup. A comparison between these processes has been presented with respect to their accuracies, precisions and other relevant metrics. This comparison reveals that testing the extent of tremors can detect PD with an accuracy of 92.19% while examining gait results in 92.71% accuracy and voice-based analysis provides 97.7% accuracy.**

Keywords*— *Parkinson's Disease, Parkinsonian Symptoms, Data Analysis, Data Mining, Tremors, Gait, Voice

I. INTRODUCTION

Parkinson's Disease (PD) is a neurodegenerative disorder that causes problems in a patient's motor and mental faculties, such as the loss of proper balance, difficulty in spoken communication and other physical difficulties. It is caused by the deficiency of the "dopamine" neurotransmitter, resulting in impaired movements, muscle stiffness, tremors and reduced vocal ability [19].

PD is globally distributed, affecting an estimated 6.3 million people. PD tends to affect more men than women. While the 60-year plus age demographic is shown to be susceptible to the disease, younger age patients are developing PD at a high rate. It has been estimated that approximately 1% of the population over the age of 60 and 4% over the age of 80 are expected to develop PD. The most common form of PD is adult onset which normally sets in after 50 years, followed by young onset PD (between the age group 21-40) and juvenile onset PD (relatively rare; <21 years).

The motivation of this paper is to identify some non-invasive methods that may be effective in identifying the presence of PD in a subject and grant them some ability to improve their quality of life post-diagnosis, both physically and mentally. The physical symptoms selected for review are tremors, freeze and shuffling of gait and noticeable impairment of vocal abilities.

II. TREMORS

A. *Introduction*

Tremors are a major symptom when a person has PD. This symptom is considered as involuntary, rhythmic and alternating movements of one or more muscles or parts [16]. It usually emerges when the muscle is in the state of rest. It is a symptom found in almost 80% of the patients suffering from the disease [17]. It is asymmetric in nature and if often manifested as twitches in the fingers, tremors in the jaw, foot and tongue. These tremors can be stabilized or the conditions can be alleviated by the administration of dopamine.

B. *Experimental Setup and Procedure*

Gupta et. al. used a dataset which recorded data for 62 Patients with Parkinson's (PWP) and 15 healthy patients [1, 2]. For all the subjects, three tests were conducted viz. SST (Static Spiral Test), DST (Dynamic Spiral Test) and STCP (Stability Test on Certain Point). Regression can be applied to all of these dataset values since they are recorded as images. They implemented the Optimized Cuttlefish Algorithm (OCFA), derived from the Traditional Cuttlefish Algorithm (CFA) for feature selection, applied over the datasets used [1,2]. Minimizing the number of features, and narrowing it down to the most important features is the goal of this algorithm. It has been applied to the above datasets to optimize the problem of feature selection and detect the disease at a preliminary stage. The performance was evaluated on the basis of a training set and a test set. The OCFA model was trained using the training set and later, the test set was used and applied to receive and detect results.

Wang et. al. used the dataset where 956 recordings of spiral test drawings were extracted from the tablet inputs [4]. Sampling time was 0.023 s and the length of the tracking was

15 seconds. Five subjects volunteered for the spiral drawing tests out of which, two were normal subjects, and others were patients with PD. Each subject was right-handed. Using the Polar Coordinate System with varied origin. The rigidity & tremors in the hands affect the origin or the centre of the spiral as time progresses. Hence, using coordinate geometry and using mathematical equations they derived a mathematical model. This model collects data for both the types of subjects and hence we can compare and analyze. The proposed algorithm has selected lesser number of features and has higher accuracy of 92.19% as opposed to the average accuracy of 87.12% and 84.49% of other machine learning algorithms such as KNN and decision trees respectively.

Smekal et al. used a dataset which has samples obtained from tests such as handwriting tests, Archimedean Spiral Test and Ellipse Illustration Test [5]. The method revolves around a non-invasive, automated analysis of symptoms of neurological disorder. The authors considered several main features that characterize the movement of the hand while the exercise is performed and the task quality is given by the other features which examine the smoothness or density of the spiral trajectory. This test gives us a rough idea about the advancement of the disease and the intensity of the tremors.

The ellipse test is an important exercise in this experiment. In this test, parameters like velocity, acceleration and jerk are analyzed. These ellipses, which are 2D in nature, can be digitally processed by their transformation into 1D periodic signals and hence can be analyzed. Micrography is noted as an important marker. The Czech sentence, “Tram will no longer go” was used in the handwriting tests.

Graça et al. reviewed a smartphone application “Parkdetect” [6]. There are two phases in the application, the “Spiral Test” (aided with a stylus) and the “Tap Games” [6]. We can compare reaction times, pressure and hold times for verifying asymmetry. In “Tap Games” the frequencies of ‘taps’ from both the hands were registered. Speed and frequency are the major features of these tests. For the tests, 18 healthy subjects and 17 affected subjects volunteered. Feature extraction was then performed and a subset of features was chosen. The most accurate results in this application are obtained using the Bayesian Networks, giving us an accuracy of 87.5% ± 23.05, the precision of 86.67% ± 30.55 and a recall of 85% ± 32.02.

C. *Advantages and Disadvantages*

Gupta et al. implemented a complete model in which the OCFA is used to give an increased accuracy as opposed to that obtained in the traditional variant of the same algorithm [3]. This procedure requires a heavy use of equipment. Any kind of discrepancy in the datasets can lead to huge changes in the training set, which result in inaccurate results. Also, an implementation flaw can lead to varying accuracies.

A Proper and specific hardware specification layout was provided for this test conducted by Wang et al. [4]. Test results were easily reproducible and could be formatted conveniently. Data representation was easily possible using various tools on the test inputs. In the end, a concrete mathematical and implementable system for data collection, preprocessing and representation. Despite its advantages, there was no permanent and standalone data analysis and prediction model proposed.

Smekal et al. implemented a system to test for resting tremors and analyze the intensity of the same [5]. It was an elaborate way of representing data which also picks up on otherwise undetected markers. However, like Wang et al. their implementation lacks a prediction model and a permanent framework for data analysis and conclusive predictions.

Graça et al. reviewed the existing multimodal system of “Parkdetect” which provides a framework that allows us to make an informed prediction [6]. Detection, data processing and data collection all use the same equipment and is highly economical when it comes to feasibility. Data is collected efficiently and the interface provided is very interactive and user-friendly. However, if a sensor were to fail, all the data would be erroneously collected and negatively affect the results. Also, system compatibility issues may limit the number of subjects.

III. FREEZING GAIT

A. *Introduction*

Freeze of Gait (FoG) is a symptom where the PWP hesitates while stepping, or faces difficulties when initiating walking [19]. The FoG events may be frequent and short. The Freezing may occur in specific situations and in specific places. FoG events are potentially dangerous may result in injuries if the patient falls over. It is, therefore, necessary to detect FoG events to avoid mishaps.

B. *Collection of Data*

Bonato et al. demonstrate a system to mine data to detect FoG events [7]. The data was collected during tests on two PWPs by means of accelerometers (ACC) attached to specific body parts and electromyographs (EMG) attached to select muscles. The time series from the ACC and EMG were used to calculate linear and non-linear features. The analysis of this data was performed using data visualization techniques. The authors show that the change of patterns during motor fluctuations is in a specific and distinct manner, making it possible to distinguish between the different motor states.

C. *Detection of FoG events*

Bachlin, et al. presents a wearable assistant for PWPs which would detect FoG events and sound an audio cue to help patients resume walking [8]. The wearable assistant was a tiny computer attached to the patient’s waist, capable of recording and processing data from sensors attached to various body parts. Sensors were attached to the thigh, the shank and the waist, at the belt used to secure the computer to the patient’s body.

The authors refer to Moore et al. to calculate a Freeze Index, or FI, which is used to find out the Freeze Threshold [9]. If the calculated FI is above the calculated Freeze Threshold, the time series is identified as a FoG event. An energy threshold was defined to distinguish between standing and other states. For online detection, the shank data was sampled at 64 Hz, with a 4s long rectangular window in steps of 0.5 seconds.

The performance of FoG detection was affected by factors like walking style and mannerisms (like foot-drop) and limited mobility in severe conditions. It was concluded that the system must also take the user’s walking style into consideration. Optimizing the Power Threshold and Freeze Threshold for each patient gave an average sensitivity of 88.6% and specificity of 92.4%.

The authors compared the results obtained from each combination of sensor positions and axis orientation. The results obtained are tabulated by Bachlin et al. in [8]. The difference is observable. The best result was given by the knee for the y and z axes as well as n, given by formula (1).

$$n = \sqrt{x^2 + y^2 + z^2} \tag{1}$$

D. Advantages and Disadvantages

A few advantages are apparent from Bonato et al. and Bachlin et al., one of them being the use of computers to detect FoG events. While Bachlin et al. proposed an assistant to help patients, a similar system could be used during medical check-ups to detect the presence of FoG in the patient's gait. However, sensor placement is an issue. Since the knee is a very intrusive and inconvenient place for attaching the sensors, the system could use an accelerometer attached to the patient's ankle while testing for FoG events.

IV. Shuffling Gait

A. Introduction

Shuffling gait is a symptom wherein the patient appears to be shuffling his/her feet instead of actually lifting them up and walking. Shuffling gait is defined as "a gait in which the foot is moving forward at the time of initial contact or during mid-swing, with the foot either flat or at heel strike, usually accompanied by shortened steps, reduced arm swing and forward flexed posture" [20].

B. Identify the Headings

Chang et al. describe the use of machine learning techniques to detect shuffling gait [10]. The data used was drawn from a public dataset maintained by Physionet containing measures of gait from 93 patients with idiopathic PD and 73 healthy patients [18]. The database includes the vertical ground reaction forces (VGRFs) of subjects who walked on level ground. Under each foot were 8 sensors, resulting in a total of 16 per person, that measured force as a function of time.

Features extracted included the mean force of each sensor, the stance and swing times, the centre of pressure and the foot strike profile. The variability in swing times tends to be a significant marker of PD since healthy people tend to have consistent swing times, while PWPs have a higher variance in swing times.

The centre of pressure is calculated to detect the foot strike profile. Healthy people walk by first lifting their heels, then their toes, and land their feet on their heels. However, due to PWPs suffering from the 'shuffling gait' symptom, their foot strike is flat-footed.

C. Classification of Patients

The classifiers chosen to be used were Logistic Regression, Random Forest classifier, Linear Kernel SVM and RBF Kernel SVM with min-max normalization. The imbalance of the classes meant that the accuracy had to be considered using precision, recall and f-score, as well as the AUC under the ROC curve. All models achieved AUC scores more than 90% with the SVM RBF performing the best at 92.71% [10].

D. Advantages and Disadvantages

Chang et al. present a system which gives insight into considering the Shuffling Gait symptom while predicting if a patient is suffering from PD. While 16 sensors are needed to collect the data of the patients, the sensors need not be too intrusive to the patients. The sensors can be conveniently placed in a shoe to get the pressure distribution for each foot while remaining relatively inexpensive to implement compared to a mobile computer chip.

V. Voice

A. Introduction

Among other symptoms, vocal impairments often manifest as dysphonia- difficulty in producing or sustaining sounds (such as those of vowels, for instance), or dysarthria- difficulty in the articulation of regular speech [11]. Sustained phonation is often preferred over running speech in designing testing methods since it is a simpler method to generate vocal samples displaying Parkinsonian symptoms without the complexities of spoken languages.

B. Analyzing Specific Properties Vocal Samples

Usually, the main properties that are extracted from samples and examined using specific speech-processing algorithms include the fundamental frequency F0 of a sample, jitter (the frequency variation from F0 between vocal cycles) and shimmer (amplitude variation between vocal cycles).

The analysis of vocal samples is complicated by the fact that an individual's voice may be affected by factors including age and nervous or muscular impairment. A sample may further be affected by noise introduced during recording. Thus, while there exist several methods to accurately assess PD in a patient, these methods cannot be utilized when a test is administered remotely. For this reason, we must weigh the contribution of a statistical quantity in the decision process against the effort we must invest to obtain it.

To account and adjust for the various factors that can introduce noise in the measurement, Little et al. propose a new quantity called Pitch Period Entropy in [11]. In [11] and [12], measurements extracted from the samples of each patient are placed in a feature vector that forms the input to our system. A support vector machine aims to find a boundary in the feature space formed by the samples in our input.

Reference [13] compares Mel Frequency Cepstral Coefficients [MFCC] that can be extracted from a given voice sample. The Mel Scale is a logarithmic scale that closely approximates the perception of sound in the human ear. The voice samples collected must be properly pre-processed in order for MFCC's to be extracted properly.

C. Preparing Measured Data for Analysis

In [11], since a smaller set of measurements is collected, Little et al. suggest that some form of correlation be considered between the measurements that are collected since they are measuring similar characteristics of a signal. As such, those measurements in each pair that are very correlated must be removed to reduce redundant information. In [12], the number of measurements collected is increased to 132 dysphonia measures. Feature selection algorithms were applied to select only relevant features.

D. Results

In [11], it is observed that the decision boundaries separating healthy patients from PWP may not be simple curves or hyperplanes. Thus, kernel-SVM formulations that allow smooth, curved decision boundaries are used. The results indicate that the non-rhythmic repetition of vocal samples in PWP indicates irregular positioning of folds in the vocal cords. In [12], a classifier validation scheme was applied and the previous best accuracy of 93% achieved using the same data set and a subset of algorithms was improved to 97.7% when the number of dysphonia measurements was increased to 132. In [13], a Leave-One-Subject-Out-Validation-Scheme was used to train the SVM over successive iterations with varied permutations of MFCC's. The SVM was constructed with different types of kernels – RBF, Linear and Polynomial. It is observed that the RBF kernel and Polynomial kernel resulted in 73.53% accuracy. The Linear classification kernel gave 91.17% accuracy.

E. Advantages and Disadvantages

Several advantages can be identified for using voice as an indicator of PD as explored in [11], [12], [13]. Vocal samples are easy to collect since simple sustained phonations are usually enough to extract almost all relevant data points. The setup is minimal in terms of hardware, and the software used in the analysis is easily and freely available. This implies a reduced financial cost of implementation. Increasing the number of data points also dramatically increases the accuracy of classification. However, an increased number of data points can also present the challenge of noise introduction in the feature space. To remove this possibility, feature selection algorithms must be used, increasing the computational cost of the process.

VI. INFERENCES

TABLE I. TREMOR DETECTION SYSTEMS AND THEIR PERFORMANCES

System	Performance	Considerations
OCFA Implemented System [1]	92.19% Accuracy	Significantly greater than the KNN (87.116%) and the Decision Tree (84.486%)
Parkdetect [6]	86.67% ± 13.54 Accuracy 91.67% ± 17.08 Precision 86.67% ± 20.82 Recall	Decision Trees are used to analyze the data
	80.83% ± 17.10 Accuracy 80.83% ± 20.43 Precision 90% ± 20 Recall	Classification Rules (RipperK) are used to analyze the data
	87.5% ± 23.05 Accuracy 86.67% ± 30.55 Precision 85% ± 32.02 Recall	Bayesian Networks are used to analyze the data

TABLE II. GAIT DETECTION SYSTEMS AND THEIR PERFORMANCES

System	Performance	Considerations
Wearable Assistant [8]	92.4% Specificity 88.6% Sensitivity	Parameters have to be optimized specifically to each patient.
	86.9% Specificity 78.1% Sensitivity	Using a global threshold.
PD Detection [10]	92.71% AUC	SVM with RBF kernel with normalized data.

TABLE III. VOICE ANALYSIS SYSTEMS AND THEIR PERFORMANCES

System	Performance	Considerations
Using Mel Frequency Cepstral Coefficients with SVMs [13]	91.17% accuracy using 12 MFCC's with Linear Kernel SVM	The samples recorded are required to be heavily pre-processed to extract proper MFCC's.
Using novel measurements with feature-selection and SVMs [12]	97.7% accuracy achieved	Proper feature selection must be done in order to ensure that the least amount of noise is introduced into feature space.

ACKNOWLEDGEMENT

The authors would like to thank Professor Krishna Samdani for his guidance in the matter.

REFERENCES

[1] Deepak Gupta, Arnav Julka, Sanchit Jain, Tushar Aggarwal, Ashish Khanna, N. Arunkumar, Victor Hugo C. de Albuquerque, "Optimized cuttlefish algorithm for diagnosis of Parkinson's disease," Cognitive Systems Research, Volume 52, 2018, pp. 36-48, ISSN 1389-0417.

[2] Isenkul, M.E.; Sakar, B.E.; Kursun, O. "Improved spiral test using digitized graphics tablet for monitoring Parkinson's disease," The 2nd International Conference on e-Health and Telemedicine (ICEHTM-2014), pp. 171-175, 2014.

[3] Erdogdu Sakar, B., Isenkul, M., Sakar, C.O., Sertbas, A., Gurgen, F., Delil, S., Apaydin, H., Kursun, O., "Collection and Analysis of a Parkinson Speech Dataset with Multiple Types of Sound Recordings," IEEE Journal of Biomedical and Health Informatics, vol. 17(4), pp. 828-834, 2013.

[4] M. Wang, B. Wang, J. Zou, J. Zhang and M. Nakamura, "Hand movement compensation for patients with Parkinson's disease based on polar coordination system with varied origin," *2011 4th International Conference on Biomedical Engineering and Informatics (BMEI)*, Shanghai, 2011, pp. 955-959.

[5] Z. Smekal, J. Mekyska, I. Rektorova and M. Faundez-Zanuy, "Analysis of neurological disorders based on digital processing of speech and handwritten text," *International Symposium on Signals, Circuits and Systems ISSCS2013*, Iasi, 2013, pp. 1-6.

[6] R. Graça, R. Sarmento e Castro and J. Cevada, "ParkDetect: Early diagnosing Parkinson's Disease," *2014 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, Lisboa, 2014, pp. 1-6.

[7] P. Bonato, D. Sherrill, D. Standaert, S. Salles, and M. Akay, "Data mining techniques to detect motor fluctuations in Parkinson's disease," in Proc. 26th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (IEMBS 2004), Sep., vol. 2, pp. 4766–4769.

[8] M. Bachlin, M. Plotnik, D. Roggen, I. Maidan, J. Hausdorff, N. Giladi, G. Troster, "Wearable Assistant for Parkinson's Disease Patients With the Freezing of Gait Symptom," IEEE TRANSACTIONS ON INFORMATION TECHNOLOGY IN BIOMEDICINE, vol. 14, no. 2, pp. 436-445, 2010.

[9] S. T. Moore, H. G. MacDougall, and W. G. Ondo, "Ambulatory monitoring of freezing of gait in Parkinson's disease," J. Neurosci. Methods, vol. 167, no. 2, pp. 340–348, 2008.

[10] D. Chang, M. Alban-Hidalgo, K. Hsu, "Diagnosing Parkinson's Disease From Gait".

[11] M. A. Little, P. E. McSharry, E. J. Hunter, J. Spielman and L. O. Ramig, "Suitability of Dysphonia Measurements for Telemonitoring of Parkinson's Disease," in IEEE Transactions on Biomedical Engineering, vol. 56, no. 4, pp. 1015-1022, April 2009.

[12] A. Tsanas, M. A. Little, P. E. McSharry, J. Spielman and L. O. Ramig, "Novel Speech Signal Processing Algorithms for High-Accuracy Classification of Parkinson's Disease," in IEEE Transactions on Biomedical Engineering, vol. 59, no. 5, pp. 1264-1271, May 2012.

[13] Benba, A. Jilbab, A. Hammouch and S. Sandabad, "Voiceprints analysis using MFCC and SVM for detecting patients with Parkinson's disease," 2015 International Conference on Electrical and Information Technologies (ICEIT), Marrakech, 2015, pp. 300-304.

[14] "Cerebral Causes of Parkinsonian Resting Tremor", *Medscape*, 2018. [Online]. Available: https://www.medscape.com/viewarticle/775696_2. [Accessed: 15- Oct- 2018].

[15] "Tremor in Parkinson's | APDA", *APDA*, 2018. [Online]. Available: https://www.apdaparkinson.org/what-is-parkinsons/symptoms/tremor/. [Accessed: 15- Oct- 2018].

[16] "Gait in Parkinson's Disease", *Physionet.org*, 2018. [Online]. Available: https://physionet.org/pn3/gaitpdb/. [Accessed: 04- Oct- 2018].

[17] "Symptoms - Parkinson's Gait | ParkinsonsDisease.net", *ParkinsonsDisease.net*, 2018. [Online]. Available: https://parkinsonsdisease.net/symptoms/parkinsons-gait/. [Accessed: 07- Oct- 2018].

[18] "shuffling gait", *TheFreeDictionary.com*, 2018. [Online]. Available: http://medical-dictionary.thefreedictionary.com/shuffling+gait. [Accessed: 04- Oct- 2018].

[19] "What is Parkinson’s? | Parkinson's Disease And Movement Disorder Society", *Parkinsonssocietyindia.com*, 2018. [Online]. Available: http://www.parkinsonssocietyindia.com/knowing-parkinsons/what-is-parkinsons/. [Accessed: 15- Oct- 2018].