Diagnosing Parkinson’s Disease using Machine Learning

Jatin Patel (1229894), Hitarth Panchal (1214537), Dr. Jinan Fiaidhi

*Abstract*—There are currently no blood, imaging, or lab tests to diagnose Parkinson’s Disease (PD). The process of diagnosing Parkinson’s Disease is thorough and consists of six steps. The criteria and standards of the same are set by the International Parkinson and Movement Disorder Society (MDS). Neurologists specializing in movement-related disorders usually assess the severity of PD by following the aforementioned norms and documenting each and every step of the exam. The clinician’s level of expertise, experience, and views introduces an inherent bias in this process, which might hinder the accuracy of their decision. Two clinicians might rate the same patient differently. In this study, we propose a rather data-driven approach to tackle this problem. Our work focuses on the early detection of PD using audio measurements and analyses the suitability of using wearable sensors to assess the acuteness of motor impairment in PD patients. We achieved a 93% F1 score using the Support Vector Classifier in the first task. Pertaining to the second objective, an experiment was performed in which we achieved an 85% F1-score in detecting PD given a small, labeled wearable sensor dataset. Such proves that there is a scope for adopting the data-driven approach as it can lead to better results given more data.

*Index Terms*—Hoehn-Yahr Scale (HYS), Machine Learning (ML), Movement Disorder Society (MDS), Parkinson’s Disease (PD), Unified Parkinson's Disease Rating Scale (UPDRS), Vertical Ground Reaction Force (VGRF)

# INTRODUCTION

T

HE Parkinson’s is the second most common age-related degenerative brain disease with an astounding 103,300 cases in Canada in the year 2019-20 as reported by Canadian Chronic Disease Surveillance System and more than 6 million world-wide[1,2]. It has several issues associated with it but the most prominent are Bradykinesia, Tremors, Dysphonia, Muscle Stiffness, difficulty with balance and co-ordination [1,2]. Parkinson’s Disease is caused due to a significant loss of neurons in the ‘Substantia Nigra’ part of the brain [3]. These neurons produce a chemical messenger known as ‘Dopamine’ to control and co-ordinate body movements, so due to a loss in the number of neurons the amount of dopamine in the body also decreases causing sluggish movements [3]. The most likely reason for the loss of neurons seems to be a combination of genetic and environmental factors [3]. The symptoms can mostly be observed after the age of 60, but sometimes the onset can be early. The only way to diagnose is a neurological examination conducted by trained clinicians and doctors, and its accuracy depends on the examiner’s experience.

Some of the assessment criteria that are used are as follows:

* “Non-motor aspects of experiences of daily living like dementia, depression, anxiety, pain, constipation, incontinence, fatigue, etc.” [4]
* “Motor aspects of experiences of daily living, which includes ability to speak, eat, chew and swallow, dress and bathe yourself.” [4]
* “Motor Examination: healthcare provider uses this section to determine the movement-related issues. It depends on how one speaks, facial expressions, stiffness, and rigidity, walking gait and speed, balance, movement speed, tremors, etc.” [4]
* “Motor complications aspect measures how much of an impact the symptoms of Parkinson’s disease are affecting one’s life.” [4]

The course of treatment is highly dependent on the patient’s assessment, so a highly reliable method of assessment is required.

# early detection using Dysphonia measurements

According to Max et al. Recently there is a lot of focus on the speech impairment in Parkinson’s patients. The primary reason for this is because approximately 90% of Parkinson’s patient exhibit some sort of speech disorder/impairment. And it is one of the few symptoms which can be observed at a very early stage of the disease. Also, it is significantly easier to collect vocal data of the patients due to its non-invasive nature as there is no need of complex devices to be attached to the subject’s body, and hence it is relatively easy to administer these methods [5]. We plan on providing a data-driven solution to accurately identify patients with Parkinson’s Disease. Its main advantages include, no need of any sort of blood samples or lab tests required, which would significantly reduce the amount of money spent on the diagnosis of PD, also most of the people experiencing these symptoms would have a hard time going for clinical appointments. And the results would be much faster than what would usually take at least a few days to accurately say if the patient has Parkinson.

## Dataset

#### Data Collection

The dataset we are using was created by Max Little of the University of Oxford, while collaborating with the National Centre for Voice and Speech, Colorado. The dataset is a collection of speech feature measurements from 23 PD (Parkinson Disease) patients and 8 healthy individuals consisting of a total of 195 voice measures corresponding to the 195 speech recordings which were recorded using a C420 AKG-US microphone. In the dataset each individual has contributed 5 recordings and more in some cases. Which included articulation of five sequences of phonations and one sustained phonation of a vowel. The first column is an ASCII subject name and recording number. Each column afterwards is one of the voice attributes associated with that recording of the individual.

#### Dataset Description

The dataset is imbalanced as there are 23 PD patients and only 8 healthy subjects. There are a total of 21 distinct features and most of them were calculated using a software called ‘Praat’ [5]. Most of the features were Traditional measures like Average vocal fundamental frequency, Maximum vocal fundamental frequency, Minimum vocal fundamental frequency, Jitter, Shimmer, Noise to Harmonic ratio and more [5]. But there were also a few Non-Traditional measures as well, like RPDE (Recurrence Period Density Entropy) which is a measure of how random or organized a person’s speech is where random can be associated with some sort of speech impairment, D2 (Correlation Dimension), DFA (Detrended Fluctuation Analysis) which is usually used to determine how the sound of the breath of a person speaking changes over time, because rigid changes can signify problems in controlling the Epiglottis (air valve). And lastly, a relatively new non-linear measure known as PPE (Pitch Period Entropy) which helps us determine if the person’s speech being jumpy, is at all related to some underlying PD conditions or not [5]. Due to the non-linear nature of the human vocal cords, the Non-linear measures (like Spread1, Spread2 and PPE) are calculated using some complex Non-linear mathematical operations like phase space reconstruction, non-linear transformation of the slope of the amplitude variation, and more. Lastly, we have a ‘status’ column which is ‘zero’ for the healthy subjects and ‘one’ for the PD patients. We will decide what features to include and what to discard based on the performance of the various models with the different sets of features [5].

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Name | MDVP:Fo | Jitter% | HNR | RPDE | DFA | Spread1 | PPE |
| phon\_R01\_S01\_1 | 119.992 | 0.00784 | 21.033 | 0.414783 | 0.81528 | -4.8130 | 0.28465 |
| phon\_R01\_S01\_2 | 122.400 | 0.00968 | 19.085 | 0.458359 | 0.81952 | -4.0751 | 0.36867 |

## Data Preprocessing

Firstly, we split the dataset into two parts, 80% as training data and 20% as testing data. And assign class weights to compensate for the unbalanced dataset. Then, we compared the use of three different approaches one is feature selection using ANOVA (Analysis of Variance) score, the other is feature compression using PCA (Principal Component Analysis) and the last one is using all the features as they are without any pre-processing.

##### Analysis of Variance for Dimensionality Reduction

In this we used the f-stat score of all the 21 features,

Between group variability,



[Image credit](https://www.analyticsvidhya.com/blog/2018/01/anova-analysis-of-variance)[6]

Within group variability,



[Image credit](https://www.analyticsvidhya.com/blog/2018/01/anova-analysis-of-variance)[6]

This helps us determine the most prominent features with the maximum score, for comparison purposes we only use the top 2 features which in our case were the PPE and Spread1.

1. Principal Component Analysis for Dimensionality Reduction

Principal Component Analysis is mainly used for analyzing large datasets by compressing them into fewer components while maintaining maximum information. Each of the original features are compressed into any number of principle components as required, in our case there are 2 such principal components.

Principal component analysis is a statistical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components. This transformation is defined in such a way that the first principal component has the largest possible variance, and each succeeding component in turn has the highest variance possible under the constraint that it is orthogonal to the preceding components. The resulting vectors are an uncorrelated orthogonal basis set. PCA is sensitive to the relative scaling of the original variables. (Principal Component Analysis, n.d.)

## Modelling

We use the standard scaler to scale the data and make its values homogenous. We used the f1 metrics to determine the results and for comparison. For training we used a number of different Machine Learning models, listed below:

### Naive Bayes (GaussianNB):

### Description: A probabilistic classifier based on the Bayes' theorem. It assumes independence between features.

### Implementation: ‘GaussianNB()’

### Decision Tree:

### Description: A tree-structured model that makes decisions based on feature values. It is constructed to optimize information gain or Gini impurity.

### Implementation: DecisionTreeClassifier(random\_state=RANDOM\_SEED, class\_weight=class\_weight)

### Random Forest:

### Description: An ensemble of decision trees that improves predictive accuracy and controls overfitting.

### Implementation: RandomForestClassifier(n\_estimators=20, random\_state=RANDOM\_SEED, class\_weight=class\_weight)

### K-Nearest Neighbor (KNN):

### Description: A non-parametric, instance-based learning algorithm that classifies data points based on the majority class of their k-nearest neighbors.

### Implementation: KNeighborsClassifier(n\_neighbors=5)

### Logistic Regression:

### Description: A linear model for binary and multiclass classification problems. It estimates the probability of a given class.

### Implementation: LogisticRegression(solver='liblinear', multi\_class='ovr', random\_state=RANDOM\_SEED, max\_iter=5000)

### Support Vector Classifier (SVC):

### Description: A discriminative classifier that separates classes with a hyperplane. It can handle linear and non-linear classification.

### Implementation: SVC(class\_weight=class\_weight, random\_state=RANDOM\_SEED, probability=True)

### MultiLayer Perceptron (MLP):

### Description: A type of artificial neural network with multiple layers of nodes. It is capable of learning complex relationships in data.

### Implementation: MLPClassifier(solver='adam', activation='relu', hidden\_layer\_sizes=(5, 10), random\_state=RANDOM\_SEED, max\_iter=5000)

### These models were selected based on their diverse characteristics and suitability for the research objectives of this project.

## Results

Figure on the right represent the different model’s performance after training on PCA.

A graph of different colored bars

Description automatically generated with medium confidence

This is a comparison of the results of the 3 approaches used (PCA, ANOVA, ALL) and their F1 score with each model.

We can conclude that using PCA for feature reduction is very great for our use-case. As it can be observed that PCA is more accurate than using all the features together in all models except the KNN (K Nearest Neighbors) and Multilayer Perceptron. With ANOVA feature reduction being a close next approach.

# Diagnosing Parkinson’s using wearable sensor

## Dataset Collection

The wearable here is the shoes called Computer Dyno Graphy System Infotronic, Netherlands. It is shown below in Figure 1. Such is widely used to assess gait impairment in Parkinson’s patients.

A pair of shoes with wires attached to them

Description automatically generated

(Figure 1) [8]

The bottom leather sole has a total of eight sensors, which are shown as squares in Figure 1. The relative positioning of these sensors is shown in Figure 2. The blue dots refer to the sensors fitted on the left shoe, and similarly, the orange dots refer to the sensors fitted on the right shoe. These coordinates are inferred considering a situation where the person stands still at the origin, facing the positive Y axis while keeping both feet parallel. L1 to L8 refers to eight sensors fixated on the left shoe, and similarly, R1 to R8 refers to eight sensors located on the right shoe.A graph with numbers and dots

Description automatically generated

(Figure 2)

Each sensor independently measures the vertical ground reaction force (VGRF), measured in Newtons. The sampling rate of the same is 100 kHz. The dataset used here is a combination of three different research datasets published by [9], [10], and [11]. It is available publicly on Kaggle [12]. The test subjects were instructed to walk normally for approximately two minutes wearing these shoes as part of the data collection process. The data attributes to the VGRF values captured by each of the sixteen sensors. Here, the test subject refers to PD patients or control subjects who do not have PD.

## Dataset Description

There are a total of 306 samples, among which 214 belong to PD patients and the remaining 92 to healthy subjects. This indicates a severe class imbalance, 70:30 for Positive to Negative classes. Each sample has its text file, and the file is named as such to indicate the study it was taken from. As discussed previously, these studies refer to [9], [10], and [11]. Apart from these, each sample also contains gender, age, weight, height, Hoehn-Yahr Scale (HYS), Unified Parkinson's Disease Rating Scale (UPDRS), and the result of the Timed Up and Go Test (TUAG). TUAG refers to “the time a person takes to stand up from a chair, walk three meters, turn around 180 degrees, walk back to the chair, and sit down while turning 180 degrees” [13].

HYS and UPDRS categorize the severity of PD into seven and six stages, respectively. Around 18% of samples lack HYS value, and even more lack UPDRS scores. Available UPDRS values have two types: one belonging to the pre-revision of UPDRS and the other to post-revision. UPDRS was originally developed in the 1980s. Later, in 2008, the Movement Disorder Society (MDS) revised it and named it MDS-UPDRS. The class imbalance issue is also present in HYS and UPDRS columns; there are more cases of some stages than others. Moreover, only 306 samples are available for analysis. These issues hinder us from assessing motor impairment on a finer level due to the insufficient amount of clean labeled data. Circumventing these issues using data imputation or upsampling techniques might introduce false information, resulting in machine learning algorithms/models learning incorrect representations. Such might compromise the health and safety of individuals. Therefore, it is not performed. Considering these challenges, we have only conducted binary classification on the dataset, predicting whether a person has PD or not.

A random sample from the dataset is plotted below in Figure 3. The Y-axis shows VGRF, measured in Newtons, and the X-axis indicates time. Such is trimmed to the first three seconds for demonstration. The data is captured at 100kHz, as mentioned previously. This graph displays a time series of 8 sensors belonging to the left leg; each is color-coded. As one walks, they put their heel on the ground first and subsequently pull up the toes from the ground to take the next step. Such a pattern is clearly visible in the data as sensor-1 (L1 or R1), which is affixed where the person’s heel is located, reaches the peak first, and sensor-8 (L8 or R8), which is affixed where the central forefoot is located, is last to peak out. Here on, we will refer to the left leg’s sensor-1 as L1, the left leg’s sensor-8 as the L8, the right leg’s sensor-1 as R1, and the right leg’s sensor-8 as R8 for ease.

A graph of different colored lines

Description automatically generated

(Figure 3)

Each sample has a variable length ranging from approximately 4s to 4m. There are a total of 16 sensors, 8 for each foot. Each sensor generates 100 readings in one second. Hence, the number of values ranges from 6400 to 384000.

## Feature Extraction

Since the length of the time series is variable for each sample, modeling this binary classification task using machine learning algorithms is difficult as those take a fixed-size feature vector as input. Length varies from 6400 to 384000, as mentioned earlier. One must sum up this time series to a small set of features to model this task. The number of features must be less, considering the size of the dataset, which is 306. We have manually engineered eight features to represent the time series data. Stride time or gait cycle duration is the time elapsed between the first contact of two consecutive footsteps of the same foot [14]. Swing time is the time elapsed between when the person picks up their foot from the ground (or toe off) and putting the same foot down (or heel strike) on the ground. A visual representation of stride time and swing time is shown in Figure 4.

A diagram of a person's body

Description automatically generated

(Figure 4) [15]

In the dataset, in each sample, the test subject walks multiple steps. For each step, one can calculate the stride and swing time. Such is done for both feet, left and right.

A graph of a line graph

Description automatically generated

A graph of a person's foot

Description automatically generated

(Figure 5)

Figure 5 shows time-series plots of VGRF values of sensor-1 (heel sensor); the left and right graphs refer to readings of the left foot and right foot, respectively. Dots represent the timestep at which the foot hits the ground, meaning when the heel touches the floor, that is, when L1 or R1 attains maximum value. Time elapsed between two consecutive dots gives us a measure of stride time.

The location of these dots is calculated using a simple algorithm. First of all, only those timesteps whose VGRF value is less than 2 Newtons are retained. Filtered timesteps are stored in chronological order. Secondly, forward differences are calculated between two consecutive time steps. If this difference is greater than 300 ms, the current timestep marks the start of the gait cycle, or else, the timestep is ignored. Using this algorithm, stride times are calculated for each gait cycle and stored in a list. Values of these thresholds are obtained using trial and error. Below Ti is a list of stride times from ith sample.

A graph of a line graph

Description automatically generated with medium confidence

A graph of a line graph

Description automatically generated with medium confidence

(Figure 6)

Figure 6 shows time-series plots of VGRF values of L1 (colored green) & L8 (colored blue) and R1 (colored green) & R8 (colored blue) for left and right feet, respectively. Red dots represent the timestep from which the foot is picked up from the ground, that is when L8 or R8 reaches the trough. After which, it stays in the air, attributing to swing time. Then, it hits the ground where L1 or R1 peaks; these points are shown as black dots. Time elapsed between two consecutive red and black dots gives us a measure of swing time.

The previously mentioned algorithm to calculate stride time is applied here as well to find timesteps at which L8 or R8 reaches the trough. There is a small difference when we find the location of trough points, the same algorithm is run on the reversed sequence. Once we get timesteps when L8 or R8 reaches the trough and timesteps when L1 or R1 reaches the peak, we can subtract two of these consecutive timesteps to obtain swing time, as shown in Figure 6. Below Wi is a list of swing times from ith sample.

Now, each sample has a list of stride times and a list of swing times. The mean and standard deviation of stride times and swing times for both feet from our final set of eight features. These features are listed below. The superscript indicates foot, left, or right. Subscript has two values, t for stride and w for swing.

## Machine Learning Methodology

1. Data Split

As mentioned earlier, there is a significant class imbalance as there are more positive samples than negative samples. Here, the positive class refers to PD patients, and the negative indicates healthy individuals. Firstly, the data is randomly divided into train and test sets using an 80/20 split. There is no logical or proven explanation for choosing these numbers. However, one has to allocate a sufficient number of samples to the training set for ML algorithms to be able to learn patterns from data. One can argue that when the size of the test set is larger than the train set, there is more to uncover or learn from examples of test dataset, and the size of the training set is insufficient. The test set is used to evaluate the trained ML models. The data is split such that the ratio of the Positive class to the negative class is the same in both the train set and the test set. If it is done completely randomly, minority class samples might get distributed such that an insufficient number of examples are allocated to the train set, and ML algorithms fail to learn their pattern. On the other hand, if an insufficient number of samples get allocated to test set, one might get an inaccurate estimate of the ML algorithms’ performance.

1. Feature Preprocessing

A simple Z-Score Normalization technique is applied to the train set. In this method, the features are scaled such that they have zero mean and unit variance. This technique aids in the convergence of machine learning models. Updated feature value z = (x - µ)/σ, where μ and σ are the mean and the standard deviation of the feature, respectively. The same is done on the test set as well, using the mean and standard deviation of the train set.

1. Modeling and Performance Evaluation

We have trained the following set of classifiers on the train data:

* 1. Gaussian Naïve Bayes
  2. Decision Tree
  3. Random Forest
  4. K Nearest Neighbor
  5. Logistic Regression
  6. Support Vector Classifier
  7. Multilayer Perceptron

Once trained, we evaluated their performance on the test set. Evaluation metrics like accuracy, precision, recall, and F1 score are calculated for the techniques, and they are displayed in Table T1. These quantities are calculated using the formulas shown below.

|  |  |  |
| --- | --- | --- |
| Actual | Prediction | |
| Negative | Positive |
| Negative | True Negative (TN) | False Positive (FP) |
| Positive | False Negative (FN) | True Positive (TP) |

Accuracy = (TN + TP) / (TN + TP + FP + FN)

Precision = TP / (TP + FP)

Recall = TP / (TP + FN)

F1 score = (2 x Precision x Recall) / (Precision + Recall)

Gaussian Naive Bayes, Decision Tree, Random Forest, K-Nearest Neighbors, Logistic regression, Support Vector Classifier, and Multilayer Perceptron achieved 49.18%, 65.57%, 78.69%, 75.41%, 77.05%, 67.21% and 77.05% accuracy respectively. The precision scores were obtained as follows: 77.27%, 78.95%, 84.09%, 81.82%, 77.36%, 87.09%, and 85.36% for Gaussian Naive Bayes, Decision Tree, Random Forest, K-Nearest Neighbors, Logistic regression, Support Vector Classifier, and Multilayer Perceptron, respectively. Similarly, Gaussian Naive Bayes, Decision Tree, Random Forest, K-Nearest Neighbors, Logistic regression, Support Vector Classifier, and Multilayer Perceptron achieved 39.53%, 69.77%, 86.05%, 83.72%, 95.35%, 62.79% and 81.39% recall respectively.

Here, we are considering the F1 score as the prime performance metric due to the presence of class imbalance. Precision and recall give a flawed measure of the algorithm’s performance when there are more positive samples in the train set compared to negative. F1 scores were obtained as follows: 52.31%, 74.07%, 85.06%, 82.76%, 85.42%, 72.97%, and 83.33% for Gaussian Naive Bayes, Decision Tree, Random Forest, K-Nearest Neighbors, Logistic regression, Support Vector Classifier, and Multilayer Perceptron, respectively.

# Conclusion

From the results of the first objective, we can conclude that using PCA for feature reduction is very great for our use-case. As it can be observed that PCA is more accurate than using all the features together in all models except the KNN and MLP. With ANOVA feature reduction being a close next approach. Using Support Vector Machine model with PCA gives staggering 93.1% F1 score, which can be used to accurately identify PD patients in the early stages of the disease.

The logistic regression model from wearable sensor task, despite being so simple, achieved the highest F1 score, which is 85.42%. However, such a value is still unacceptable considering the healthcare domain. The health and safety standards set out to meet certain quality requirements are critical as a decision about the individual’s health shall never be taken by algorithms that are barely good at what they do. We believe the insufficient amount of high-quality clean labeled data is a prime reason for average performance, and it can be improved further by collecting more samples. Sometimes, when the motor impairment exam of a patient takes place, a video of the same is also recorded. During the exam, a healthcare professional evaluates the patient’s symptoms and decides upon the accurate UPDRS value, which signifies the severity of the disease. Our approach can be altered by including a video analysis system along with the wearable sensor module. Both data can be synchronized and fused together to provide a much more accurate prediction. The video analysis task would involve extracting 3D key points of the patient’s joints from each frame and fusing these time series with wearable sensors’ measurements to classify the acuteness of motor impairment on a finer level. Accurate prediction is vital as it affects the type of treatment the patient receives and the prognosis report.

References

1. “Parkinson’s Disease.” *AANS*, American Association of Neurological Surgeons, [www.aans.org/en/Patients/ Neurosurgical-Conditions-and-Treatments/Parkinsons-Disease](https://www.aans.org/en/Patients/Neurosurgical-Conditions-and-Treatments/Parkinsons-Disease).
2. Khan, Mahreen. “Reporting Rates of Parkinson’s in Canada.” Parkinson Canada, 21 Mar. 2023, [www.parkinson.ca/reporting-rates-of-parkinsons-in-canada](http://www.parkinson.ca/reporting-rates-of-parkinsons-in-canada).
3. “Causes of Parkinson’s Disease.” NHS Choices, NHS, [www.nhs.uk/conditions/parkinsons-disease/causes](http://www.nhs.uk/conditions/parkinsons-disease/causes).
4. Cleveland Clinic medical. “Parkinson’s Disease: What It Is, Causes, Symptoms & Treatment.” Cleveland Clinic, 15 Apr. 2022, [www.my.clevelandclinic.org/health/diseases /8525-parkinsons-disease-an-overview](http://www.my.clevelandclinic.org/health/diseases%20/8525-parkinsons-disease-an-overview).
5. Max A. Little, Patrick E. McSharry, Eric J. Hunter, Lorraine O. Ramig (2008), “Suitability of Dysphonia Measurements for Telemonitoring of Parkinson’s Disease.” IEEE Transactions on Bio-Medical Engineering, U.S. National Library of Medicine, Apr. 2009, <www.ncbi.nlm.nih.gov/pmc/articles/PMC3051371>**.**
6. Singh, Gurchetan1000. “ANOVA: Complete Guide to Statistical Analysis & Applications” Analytics Vidhya, 7 July 2023, <www.analyticsvidhya.com/blog/2018/01/anova-analysis-of-variance>
7. “Principal Component Analysis.” Wikipedia, Wikimedia Foundation, 28 Nov. 2023, [en.wikipedia.org/wiki/Principal\_component\_analysis](https://en.wikipedia.org/wiki/Principal_component_analysis)**.**
8. Jeleń, Piotr, et al. “Expressing Gait-line Symmetry in Able-bodied Gait.” Dynamic Medicine, vol. 7, no. 1, Dec. 2008, <https://doi.org/10.1186/1476-5918-7-17>
9. Frenkel‐Toledo, Silvi, et al. “Effect of Gait Speed on Gait Rhythmicity in Parkinson’s Disease: Variability of Stride Time and Swing Time Respond Differently.” Journal of Neuroengineering and Rehabilitation, vol. 2, no. 1, July 2005, <https://doi.org/10.1186/1743-0003-2-23>
10. Yogev, Galit, et al. “Dual Tasking, Gait Rhythmicity, and Parkinson’s Disease: Which Aspects of Gait Are Attention Demanding?” European Journal of Neuroscience, vol. 22, no. 5, Sept. 2005, pp. 1248–56. <https://doi.org/10.1111/j.1460-9568.2005.04298.x>.
11. Hausdorff, Jeffrey M., et al. “Rhythmic Auditory Stimulation Modulates Gait Variability in Parkinson’s Disease.” European Journal of Neuroscience, vol. 26, no. 8, Oct. 2007, pp. 2369–75. <https://doi.org/10.1111/j.1460-9568.2007.05810.x>.
12. “Gait in Parkinson’s Disease” *Kaggle*, 11 June 2022, www.kaggle.com/datasets/zarif98sjs/gait-in-parkinsons-disease/data.
13. Wikipedia contributors. “Timed up and Go Test.” *Wikipedia*, 13 Aug. 2023, en.wikipedia.org/wiki/Timed\_Up\_and\_Go\_test.
14. Beauchet, Olivier, et al. “Walking Speed-related Changes in Stride Time Variability: Effects of Decreased Speed.” Journal of Neuroengineering and Rehabilitation, vol. 6, no. 1, Aug. 2009, <https://doi.org/10.1186/1743-0003-6-32>.
15. Stöckel, Tino & Jacksteit, Robert & Behrens, Martin & Skripitz, Ralf & Bader, Rainer & Mau-Moeller, Anett. (2015). The mental representation of the human gait in young and older adults. Frontiers in Psychology. 6. 943. 10.3389/fpsyg.2015.00943.