December 3, 2023

A. J. Thibeault

RESEARCH FIELD: Medical Research - Parkinson's Disease

PROJECT TITLE: An Investigation into the Properties of Freezing of Gait: Assessing Classification Methods to Predict FOG Events via Wearable Technologies

EXECUTIVE SUMMARY

There are estimated to be upwards of 10 million people worldwide affected by Parkinson's disease. While the disease itself remains relatively poorly known, Freezing of Gait is one of the least understood symptoms of Parkinson's disease. Freezing of Gait (FOG) can be described as "an episodic gait pattern characterized by the inability to step that occurs on initiation or turning while walking."[1] These FOG events can severely impair mobility (and mobility confidence) and are largely disruptive to the sufferer's life and daily function.

Using data provided by the Michael J. Fox Foundation, this report examines the relationships between the accelerometer data collected from APDM Wearable Technologies' Opal sensors and observed FOG events to recognize patterns and develop a mechanism by which predict these events. The analysis was conducted by evaluating four machine learning classifier techniques against the three FOG types (Start Hesitation, Turn, and Walking) both individually and in conjunction in Python to determine the highest performer using F_1 and accuracy scores. Of the Extreme Gradient Boosting, Light Gradient-Boosting Machine, Random Forest, and Histogram Gradient Boosting methods evaluated, Extreme Gradient Boosting was the highest performer across the Start Hesitation and Turn FOG types and Light Gradient-Boosting Machine was the highest performer for the Walking FOG type for single-type predictions through evaluation of both performance scores. Of the multiclass classification models, Light Gradient-Boosting Machine yielded the highest performance scores for predicting all three FOG types. Going forward, researchers are recommended this approach to further fine-tune prediction efforts to improve evaluation, monitoring, and prevention of FOG events.

1.0 – PROJECT DESCRIPTION

The neurodegenerative disorder is characterized by a gradual degeneration of an individual's nerve cells in the substantia nigra, the movement control center of the brain, which is vital in the production of the neurotransmitter dopamine (Appendix Figure A). As the dopamine levels decrease, the affected individual then experiences a range of symptoms: tremors, muscular rigidity, slowness of movement, and impaired balance and coordination.

Improved understanding and treatment of this symptom could be liberating for millions of people worldwide and their communities. This study aims to add value to researchers attempting to expand current knowledge of the symptom and finely tune treatments for individuals with Parkinson's Disease.

1.1 - RESEARCH QUESTIONS

Can we detect FOG events using data from a 3D accelerometer on the individual's low back (Opals by APDM Wearable Technologies) and their respective demographic information? If so, which method produces optimal prediction results for each FOG type? Is there a viable prediction model that incorporates all FOG events?

1.2 - STATISTICAL QUESTIONS

Is there enough of a relationship between data collected from APDM's Opal sensors and confirmed FOG event markers to predict when these events will occur with a statistical model?

Identifying relationships between variables will be evaluated through exploratory data analysis, mainly evaluating Opal and demographic data in relation to each other. Additionally, strong potential predictors will be identified through Principal Component Analysis for each FOG type. The question of prediction will be answered via classifier comparison between four machine learning techniques: Extreme Gradient Boosting, Light Gradient-Boosting Machine, Random Forest, and Histogram Gradient Boosting. The four classification types will be evaluated using two prediction performance metrics: F_1 score and accuracy. Lastly, prediction legitimacy will be evaluated on a multiclass classification level to identify the optimal model for predicting all three

FOG types through the four machine learning model techniques and similarly evaluated on their F_1 and accuracy scores.

1.3 - VARIABLES OF INTEREST

This analysis focused on the tDCS FOG (tdcsfog) dataset, the series of lab collections as subjects performed a FOG-provoking protocol, provided by the studies Reches T, Dagan M. et al., 2021[2] and Manor B, Dagan M et al. 2021.[3] This data was collected from three research groups: The Center for the Study of Movement, Cognition and Mobility (CMCM) at the Neurological Institute in Tel Aviv Sourasky Medical Center, Israel, The Neurorehabilitation Research Group at Katholieke Universiteit Leuven in Belgium, and Mobility and Falls Translational Research Center at the Hinda and Arthur Marcus Institute for Aging, affiliated with Harvard Medical School in Boston. After being screened for complete study-inclusion classification, subjects recruited from CMCM and Harvard were recorded performing a developed FOG-provoking sequence and their measurements recorded. Trials were videotaped and annotated by expert reviewers documenting the FOG episodes. A schematic of the FOG-provoking sequence is provided in Figure 1.

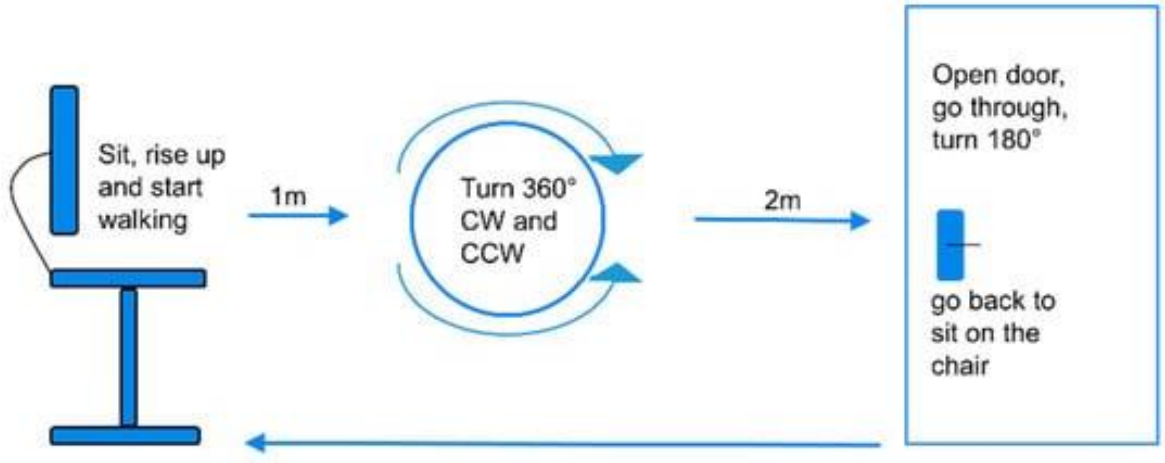

Figure 1: A representation of the FOG-provoking protocol that subjects were recorded performing during their lab visits. Protocol variations included clockwise and counterclockwise rotations. The protocol was repeated three times, each with different levels of difficulty. [6]

From this dataset, there are 7,062,672 retained accelerometer measurements representing 65 individual subjects across 310 visits and 839 total tests. The number of measurements per patient

ranged between 9,500 and 221,504, with one significant contributor of 1,052,819 measurements. All observations included were complete, that is, there were no observations missing details persistent in the analysis as they were removed before consideration. In terms of FOG prediction, the variables of interest are Opal's vertical, mediolateral, and anteroposterior acceleration measurements, the respective timestamps for all events, and the difficulty level of each performed protocol. The acceleration measurements were recorded from the APDM sensors located on each patient's lower back in units of m/s^2. The integer timestamps for time elapsed during protocol were also recorded by the APDM sensors at 128Hz, or 128 timestamps per second. The indicator variables for each FOG type were measured during the video review stage by professionals and difficulty levels were confirmed, with 1 being the least difficult and 3 being the most.

Additional information was collected from the tDCS FOG metadata set and the Subject files. The three datasets were merged along the subject Id's to create an expanded set of acceleration and patient demographic data. Variables of interest from these datasets are largely descriptive: the visit number, subject's medication status, age, sex, and the number of years since the subject's diagnosis. From prior research, we know that sex[4] and age[5] are significant risk factors for developing Parkinson's Disease. Mainly, males ages 60 and older are the highest risk group for the disease, so we want to understand the weight of these demographics in relation to the FOG events. Additionally, the years since each patient's Parkinson's diagnosis is considered in this analysis to determine its relation to FOG events. Demographic information was collected by researchers at the time of participant recruitment.

2.0 - EXPLORATORY DATA ANALYSIS

As a starting point, our analysis begins with an exploration of data distributions. Figure 2 displays a sample distribution of all participants' ages separated by sex, with Male = 0 and Female = 1. As anticipated, the number of male participants significantly outnumbers the female participants. Additionally, the study's distribution of subjects' medication status from each visit is included in Figure 3. Since participants in the study were collected from medical centers for Parkinson's Disease treatment, the higher on-medication count is anticipated.

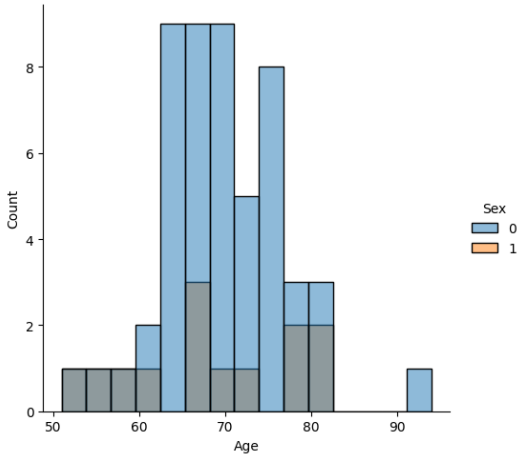

Figure 2: Age Distribution by Sex

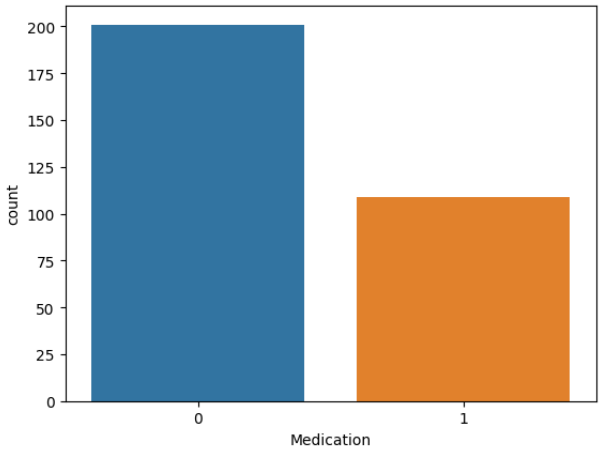

Figure 3: Medication Status Count

From Figure 4, it is clear a significant majority of subjects participated in the study through the second lab visit and slowly tended to drop off after the fifth, with few participants lasting through the twentieth.

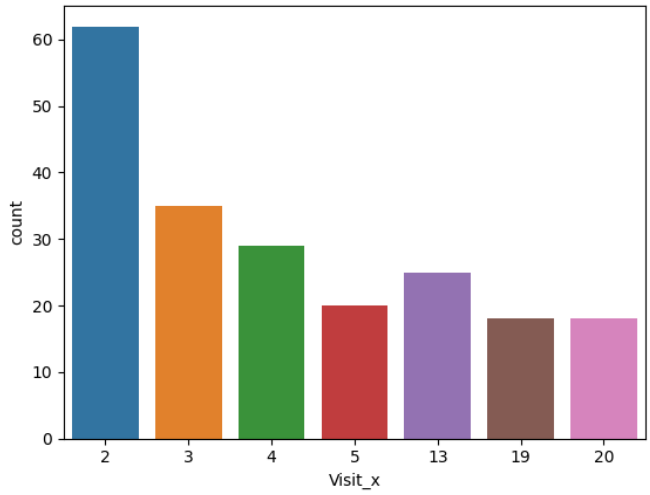

Figure 4: Distribution of Recorded Visits

Subjects' medication status is plotted against the three FOG types in Figure 5. While the impact of medication is nonexistent in FOG types Start Hesitation and Walking, there is a difference in the Turn type, where subjects not on anti-Parkinsonian medication were recorded to experience less Turn FOG events than their medicated counterparts.

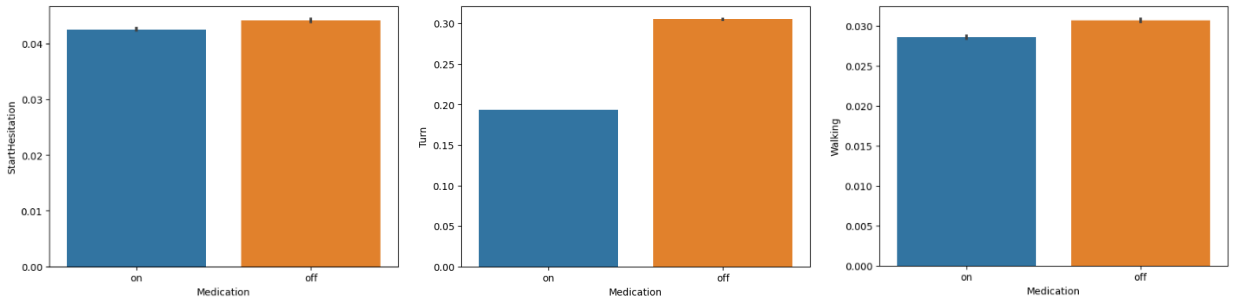

Figure 5: A comparison of medication status across the experienced FOG types.

A time series plot in Figure 6 shows the movement fluctuations over time during the lab sessions, where peaks along the graph indicate significant movement from each of the three accelerometer measurements. Our initial interest with this relationship was to identify if there was any difference between FOG events, or spikes in the accelerometer measurements, over time. There are clear disruptions noted about the 1600-time mark, but otherwise, the distribution appears relatively uniform over time.

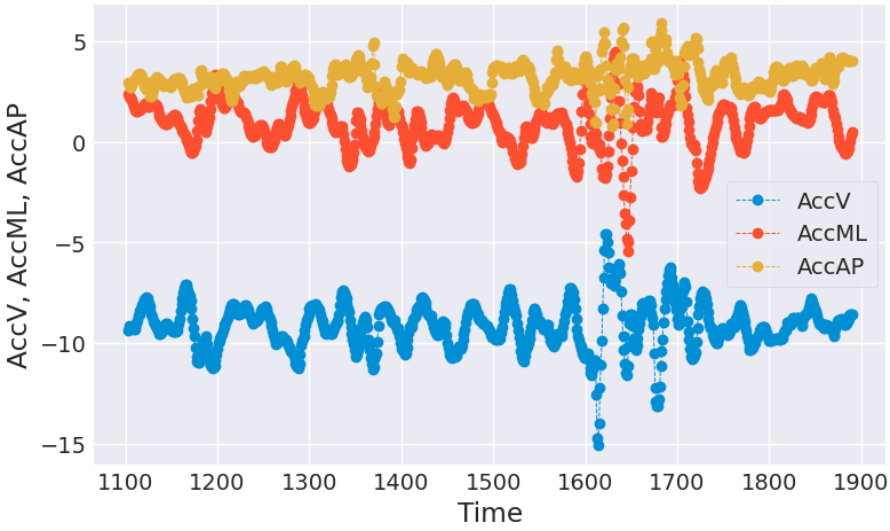

Figure 6: Time Series Illustration of the Acceleration Measurements Recorded.

To examine the relationships between the accelerometer measurements (presence and absence of FOG events) and the subjects' demographic information (collectively, the potential predictors), Figure 7 summarizes the correlation between these factors, with brighter boxes representing more positive relations between the two variables and darker ones indicate a more negative

relationship. Time shows the most significant correlation between AccAP and Visit_x, while all other pairings show very low to no correlation.

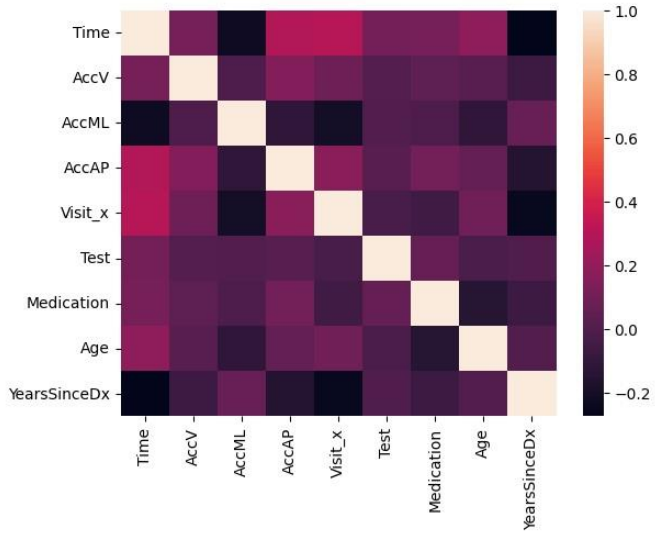

Figure 7: Correlation Heat plot of Potential Predictors

A numeric summary of the data is presented in Appendix Table A, with variable descriptions in Appendix Table B.

3.0 - STATISTICAL ANALYSIS

To identify relationships between our predictors (recorded timesteps, the three accelerometer measurements, the visit number, test sequence, the subject's medication status, age, sex, and the number of years since the subject's diagnosis) and the three FOG types, we utilized Principal Component Analysis (PCA). PCA is a dimension reduction machine learning technique that transforms data into a new coordinate system where each principal component describes the characteristics of most variation in the remaining data. When plotted, this allows us to visually identify dimensions of the data that are linearly uncorrelated and thus independent contributors. While performing the analysis, six principal components were determined to be sufficient to represent the data as they collectively covered approximately 72% of the variation expressed across the data set. Further examining the predictors' contributions to each component, all predictors were found to be significant contributors for each FOG type and were included in further analysis.

The four machine learning classifier methods (Extreme Gradient Boosting, Light Gradient-Boosting Machine, Random Forest, and Histogram Gradient Boosting) were constructed, repeated 5-fold each with variations to the randomized input data, and their resulting models were evaluated on F_1 and accuracy scores to determine prediction power for the three FOG events. The models chosen are all decision tree-based classifiers utilizing different methods by which to build the learned classifier (Appendix Table C). Decision trees are hierarchical models that utilize rules to make decisions and consequences about the classification of each corresponding set of data points based on observed patterns (Appendix Figure B). A completed tree will have built a structure of decisions such that every possible case for data values is mapped down to the most likely class.

F_1 evaluates the models' predictions based on the precision and recall scores (Figure 8), where precision measures how many of the "positive" predictions made were correct and recall measures how many positive class samples present were correctly identified. Accuracy is calculated as the ratio of the number of correct predictions over the total number of predictions. A realistic model aims to produce both F_1 and Accuracy values of between 0.7 – 1.0. While high values for both metrics are preferred, for cases where the two metrics greatly disagree, we will rule in preference of the F_1 score as its calculation is more robust for imbalanced class distributions.

$$F_1 = \frac{no.of\ true\ positives}{no.of\ true\ positives + \frac{1}{2}(no.of\ false\ positives + no.of\ false\ negatives)}$$

Figure 8: F_1 Score Schema

The classifier production procedure involved model creation using the respective classifier to predict the occurrence of a FOG event, applying the built model to the set of predictions desired, and calculating the performance metrics at each iteration. The procedure was then repeated for each FOG type. The dataset provided was split into a random 70/30 training and testing datasets where the models in question were built using the training set and evaluated on their predictions for the remaining testing set. The results of the three procedures are summarized in Table 1.

Start Hesitation

	F1 score	Accuracy
XGBClassifier	0.7660	0.9810
LGBMClassifier	0.4877	0.9665
RandomForest	0.5042	0.9568
HistGradientBoostingClassifier	0.4731	0.9660

Turn

	F1 score	Accuracy
XGBClassifier	0.7337	0.8777
LGBMClassifier	0.7087	0.8690
RandomForest	0.7074	0.8607
HistGradientBoostingClassifier	0.7083	0.8686

Walking

	F1 score	Accuracy
XGBClassifier	0.4338	0.9749
LGBMClassifier	0.3756	0.9738
RandomForest	0.4450	0.9670
HistGradientBoostingClassifier	0.3426	0.9733

Table 1: Binary Classification Modeling Results

Of the four types of classifiers tested, we can see the Extreme Gradient Boosting classifier outperformed the others in accuracy in all three FOG categories and yielded the best F_1 score for Start Hesitation and Turn, while Random Forest slightly outperformed the F_1 score for prediction Walking FOG. The results held least merit in the Walking FOG evaluation, where the F_1 score achieved a maximum of 44.50%, but maintained an 'at least 70%' benchmark in all other areas.

Separately, we can see significantly stronger F_1 score results for the Turn FOG than for Start Hesitation or Walking FOG. Due to the nature of the gyroscopic, acceleration, and magnetic measurements recorded, we anticipated this result as the change in multidirectional movement is easier to identify.

To round out the analysis, the models were further expanded to incorporate all three FOG events and evaluated on their ability to predict the presence of all three types using the significant predictors. To accommodate this, the F_1 score calculation was adjusted to use the weighted average of each prediction class's contribution. The results are condensed in Table 2. Of the four multiclass models evaluated, we can see the Light Gradient-Boosting Machine model outperformed the others in both F_1 score and prediction accuracy, achieving 85.33% and 85.98%, respectively.

	F1 score	Accuracy
Multiclass XGBClassifier	0.6796	0.8575
Multiclass LGBMClassifier	0.8533	0.8598
Multiclass RandomForest	0.6547	0.8434
Multiclass HistGradientBoostingClassifier	0.8472	0.8547

Table 2: Multiclass Classification Modeling Results

4.0 - RECOMMENDATIONS

To further additional research into the cause and recognition of Freezing of Gait (FOG) events, researchers should implement Extreme Gradient Boosting classifiers to identify incidences. The classifier is particularly effective for predicting Turn events with both performance metrics within the acceptable range.

For more wholistic results, it is recommended that researchers explore the combined dynamic of the data collected from not only lab measurements but home measurements as well, as both have the added layer of data verification from professional review. Significant computing power is required for a full combined analysis, but additional information upon which to train the classifiers may yield different or yet stronger results.

Scientists are encouraged to apply these findings to stronger data collection methods and implement the results in tandem with other treatments. With further context as to when and why Freezing of Gait events occur, additional measurements can be collected during these events to expand current knowledge of the symptom.

5.0 – CONSIDERATIONS

Trials from the tdcsfog dataset were videotaped and annotated by expert reviewers who documented the Freezing of Gait episodes. Having no control of the data collection process, we're assuming the collection of measurements and determination of events is accurate and the machinery is properly calibrated. In actuality, the observation of a FOG incident could reflect the situation delayed from its conception.

Additionally, the off- and on-medication trials were conducted at separate times, with the on-medication procedure following the first. There could be an underlying relationship to the timeline influencing the results of the later test that should be acknowledged during interpretation. For example, participants whose symptoms are aggravated with exertion may experience higher symptoms in the second procedure purely due to the circumstance.

The study data provided included approximately double the observations for subjects on anti-Parkinsonian medication versus without any medication, so any conjectures made about demographic representation need to have this adjusted for to have similarly represented outcomes. Likewise, as the unique sample size came from only 65 subjects, an expanded or meta-study is recommended to make stronger conclusions about the investigated properties for the population of individuals Parkinson's disease. Due to the restricted sample size, only nonparametric analysis techniques were considered. These results may prove different with a larger sample size or applied transformations for normality of distribution to allow for additional testing and modeling procedures.

A significant limiting factor for this analysis was system memory, and as a result, only one data set was able to be analyzed of the two possible. A combined analysis exceeded the technical limits of this study, so likely memory compression techniques will be needed to advance these research methods and allow for stronger results. Techniques such as a memory profiler have proven effective in some cases.[7]

Thank you for the opportunity to contribute to this research. If there are any additional questions or comments, please contact me.

APPENDIX:

Figure A: Visualization of the Dopamine Receptor Channel

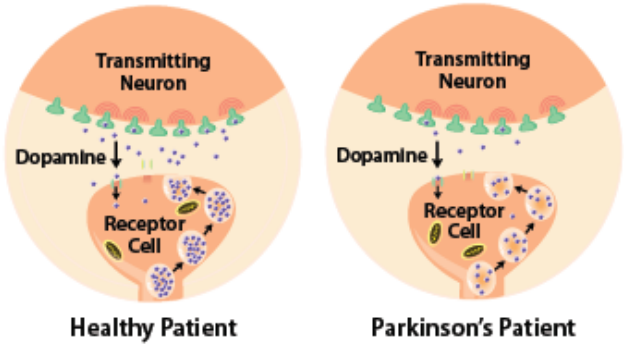

Figure B: A Generalized Decision Tree Schematic[8]

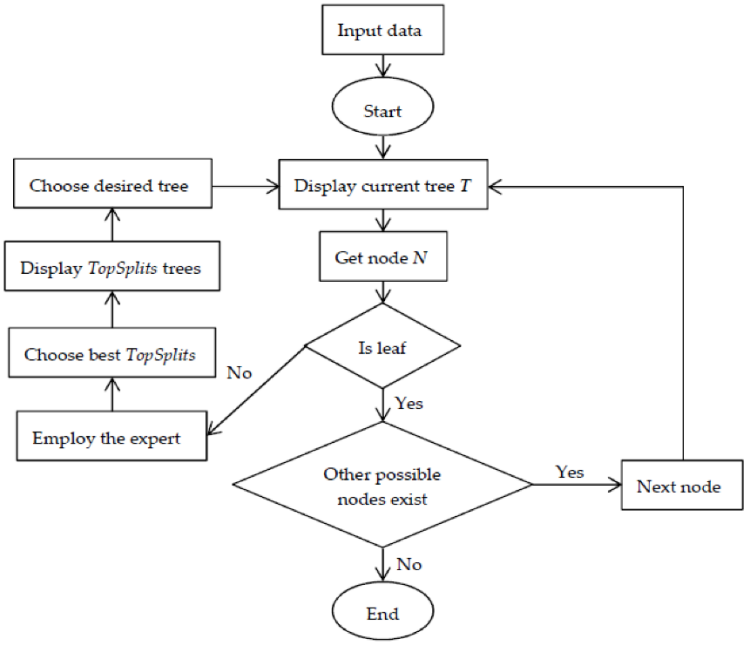

Table A: Numeric Analysis Summary

	Time	AccV	AccML	AccAP	StartHesitation	Turn	Walking	Visit_x	Test	Medication	Age	YearsSinceDx
count	7.062672e+06	7.062672e+06	7.062672e+06	7.062672e+06	7.062672e+06	7.062672e+06	7.062672e+06	7.062672e+06	7.062672e+06	7.062672e+06	7.062672e+06	7.062672e+06
mean	9.289467e+03	-9.306317e+00	-2.012513e-01	1.808524e+00	4.315506e-02	2.376979e-01	2.942767e-02	7.352151e+00	2.093533e+00	3.971565e-01	6.972636e+01	9.119993e+00
std	1.399893e+04	1.080174e+00	1.269525e+00	2.285849e+00	2.032061e-01	4.256731e-01	1.690020e-01	6.779498e+00	8.301861e-01	4.893089e-01	7.717013e+00	5.789700e+00
min	0.000000e+00	-3.552112e+01	-2.616440e+01	-4.782964e+01	0.000000e+00	0.000000e+00	0.000000e+00	2.000000e+00	1.000000e+00	0.000000e+00	5.100000e+01	1.000000e+00
25%	2.119000e+03	-9.762402e+00	-9.295446e-01	5.672254e-01	0.000000e+00	0.000000e+00	0.000000e+00	2.000000e+00	1.000000e+00	0.000000e+00	6.600000e+01	4.000000e+00
50%	4.310000e+03	-9.363524e+00	-1.722245e-01	1.987101e+00	0.000000e+00	0.000000e+00	0.000000e+00	4.000000e+00	2.000000e+00	0.000000e+00	7.100000e+01	8.000000e+00
75%	8.433000e+03	-8.776814e+00	5.752114e-01	3.449026e+00	0.000000e+00	0.000000e+00	0.000000e+00	1.300000e+01	3.000000e+00	1.000000e+00	7.400000e+01	1.200000e+01
max	9.707600e+04	2.090695e+01	2.748472e+01	3.033769e+01	1.000000e+00	1.000000e+00	1.000000e+00	2.000000e+01	3.000000e+00	1.000000e+00	9.400000e+01	2.300000e+01

Table B: Description Summary of Usable Data

Name	Description
tDCS FOG (tdcsfog) dataset	Comprising data series collected in the lab, as subjects completed a FOG-provoking protocol
Time	An integer timestep. Series from the tdcsfog dataset are recorded at 128Hz (128 timesteps per second), while series from the defog and daily series are recorded at 100Hz (100 timesteps per second)
AccV, AccML, and AccAP	Acceleration from a lower-back sensor on three axes: V - vertical, ML - mediolateral, AP - anteroposterior. Data is in units of m/s^2
StartHesitation, Turn, Walking	Indicator variables for the occurrence of each of the event types
Visit_x	Lab visits consist of a baseline assessment, two post-treatment assessments for different treatment stages, and one follow-up assessment.
Test	Which of three test types was performed, with 3 the most challenging
Medication	Subjects may have been either off or on anti-parkinsonian medication during the recording
Age	Subject's age
YearsSinceDx	Years since Parkinson's diagnosis

Table C: Classifier Method Descriptions collected from Python Documentation

Classifier Method	Process Description
Extreme Gradient Boosting	"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."[9]
Light Gradient-Boosting Machine	"LightGBM is a gradient boosting framework that uses tree-based learning algorithms. It is designed to be distributed and

	efficient with the following advantages: faster training speed and higher efficiency; lower memory usage; better accuracy; support of parallel, distributed, and GPU learning; capable of handling large-scale data."[10]
Random Forest	"A random forest is a meta estimator that fits a number of decision tree classifiers on various sub-samples of the dataset and uses averaging to improve the predictive accuracy and control over-fitting. The sub-sample size is controlled with the max_samples parameter if bootstrap=True (default), otherwise the whole dataset is used to build each tree."[11]
Histogram Gradient Boosting	"This estimator has native support for missing values (NaNs). During training, the tree grower learns at each split point whether samples with missing values should go to the left or right child, based on the potential gain. When predicting, samples with missing values are assigned to the left or right child consequently. If no missing values were encountered for a given feature during training, then samples with missing values are mapped to whichever child has the most samples."[12]

REFERENCES:

1. Rahimpour, S., Gaztanaga, W., Yadav, A. P., Chang, S. J., Krucoff, M. O., Cajigas, I., Turner, D. A., & Wang, D. D. (2020, December 26). *Freezing of gait in parkinson's disease: Invasive and noninvasive neuromodulation*. Neuromodulation: journal of the International Neuromodulation Society. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8233405/#:~:text=Freezing%20of%20gait%20(FoG)%20is,with%20perception%20of%20tight%20surroundings.
2. Reches T, Dagan M, Herman T, Gazit E, Gouskova NA, Giladi N, Manor B, Hausdorff JM. Using Wearable Sensors and Machine Learning to Automatically Detect Freezing of

Gait during a FOG-Provoking Test. Sensors (Basel). 2020 Aug 10;20(16):4474. doi: 10.3390/s20164474. PMID: 32785163; PMCID: PMC7472497.

3. Manor B, Dagan M, Herman T, Gouskova NA, Vanderhorst VG, Giladi N, Travison TG, Pascual-Leone A, Lipsitz LA, Hausdorff JM. Multitarget Transcranial Electrical Stimulation for Freezing of Gait: A Randomized Controlled Trial. Mov Disord. 2021 Nov;36(11):2693-2698. doi: 10.1002/mds.28759. Epub 2021 Aug 18. PMID: 34406695.
4. Cerri S, Mus L, Blandini F. Parkinson's Disease in Women and Men: What's the Difference? J Parkinsons Dis. 2019;9(3):501-515. doi: 10.3233/JPD-191683. PMID: 31282427; PMCID: PMC6700650.
5. Reeve A, Simcox E, Turnbull D. Ageing and Parkinson's disease: why is advancing age the biggest risk factor? Ageing Res Rev. 2014 Mar;14(100):19-30. doi: 10.1016/j.arr.2014.01.004. Epub 2014 Feb 3. PMID: 24503004; PMCID: PMC3989046.
6. Reches, T.; Dagan, M.; Herman, T.; Gazit, E.; Gouskova, N.A.; Giladi, N.; Manor, B.; Hausdorff, J.M. Using Wearable Sensors and Machine Learning to Automatically Detect Freezing of Gait during a FOG-Provoking Test. Sensors 2020, 20, 4474. https://doi.org/10.3390/s20164474
7. Hoang, Nhu. "Optimize Memory Tips in Python." Medium, Towards Data Science, 1 Sept. 2021, towardsdatascience.com/optimize-memory-tips-in-python-3bbb44512937.
8. 8 Gajowniczek K, Ząbkowski T. Interactive Decision Tree Learning and Decision Rule Extraction Based on the *ImbTreeEntropy* and *ImbTreeAUC* Packages. *Processes*. 2021; 9(7):1107. https://doi.org/10.3390/pr9071107
9. XGBoost Documentation — xgboost 2.0.2 documentation. (n.d.). https://xgboost.readthedocs.io/en/stable/
10. Welcome to LightGBM's documentation! — LightGBM 4.0.0 documentation. (n.d.). https://lightgbm.readthedocs.io/en/stable/
11. sklearn.ensemble.RandomForestClassifier. (n.d.). Scikit-learn. https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html#sklearn.ensemble.RandomForestClassifier

12. *sklearn.ensemble.HistGradientBoostingClassifier*. (n.d.). Scikit-learn. https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.HistGradientBoostingClassifier.html