Charlotte Baer
Briarcliff High School

Early detection of Parkinson's Disease and the role that Artificial Intelligence and Machine Learning Play

Table of Contents

This research looks at the role that machine learning plays in the detection of Parkinson's Disease. Machine learning (ML) is a method of data analysis that automates analytical model building. It is a branch of artificial intelligence based on the idea that systems can learn from data, identify patterns and make decisions with minimal human intervention. In the past decade there has been a focus on the application of Machine Learning and other AI techniques in the medical sciences. Studies have shown a correlation between PD and activities of daily living such as keystrokes and gait as early signs of the disease. PD patients have significantly slower typing speeds compared to healthy patients. Tracking typing characteristics of at-risk population could help in the early detection of PD. Two datasets were studied, the first data set consisted of 31 subjects, 13 healthy controls and 18 patients with Parkinson's Disease. It looked at a variety of different finger tapping tests including typing speed. The second data set consisted of 93 patients with idiopathic PD and 73 healthy controls both with a mean age of 66.3 years. Gait tests were conducted with subjects walking at their usual self-selected pace for about 2 minutes.

R code was used to categorize the patients and a multivariate data analysis was used to identify what characteristics of keystrokes and gait are important in PD patients. We then created 2 predictive models for each data set, with and without clinical data for each. We found that gait with clinical data had an accuracy of 1.000, gait without clinical data had an accuracy of 1.000, keystroke with clinical data had an accuracy of 0.989, and keystroke without clinical data had an accuracy of 0.989. These findings show us that gait is an important factor in the detection and diagnosis of Parkinson's Disease. This is important because the earlier that the disease is diagnosed, the more effective and inexpensive treatment will be, and patients will have a subsequent better quality of life. Gait can be easily tracked for at risk populations and can therefore be a vital indicator of this disease.

1.0 Review of Literature

Machine learning (ML) is a statistical technique of data analysis that automates analytical model building. It is a branch of artificial intelligence based on the idea that systems can learn from data, identify patterns and make decisions with minimal human intervention. Machine Learning was created from pattern recognition, and it stemmed from the theory that computers can learn without being specifically programmed to perform a simple task. The iterative aspect of machine learning is important because as models are exposed to new data, they are able to independently adapt. They learn from previous computations to produce reliable, repeatable decisions and results. While many machine learning algorithms have been around for a long time, the ability to automatically apply complex mathematical calculations to big data – over and over, faster and faster – is a recent development.

Parkinson's disease (PD) belongs to a group of conditions called motor system disorders, which occur due to the loss of dopamine-producing brain cells in the elderly. The four primary symptoms of PD are tremor, rigidity, bradykinesia, or slowness of movement: and postural instability. As the disease progresses it becomes more difficult for patients to walk, talk or do simple tasks. Other symptoms may include depression and other emotional changes; difficulty in swallowing, chewing, and speaking; urinary problems or constipation; skin problems; and sleep disruptions. There are currently no blood or laboratory tests that can diagnose sporadic PD. The diagnosis is based on medical history and a neurological examination. Doctors may sometimes request brain scans or laboratory tests in order to rule out other diseases. The only concrete way to diagnose PD is posthumous. PD is both chronic, meaning it persists over a long period of time, and progressive, meaning its symptoms grow worse over time. Although some people become severely disabled, others experience only minor motor disruptions. Tremor is the major symptom for some individuals, while for others tremor is only a minor complaint and other symptoms are more troublesome. It is currently not possible to predict which symptoms will affect an individual, and the intensity of the symptoms also varies from person to person. Although there is no cure for Parkinson's disease, medicines, surgical treatment, and other therapies can often relieve some symptoms. The most common treatments are drugs that increase the level of

dopamine in the brain, drugs that affect other brain chemicals such as Catechol O-Methyltransferase and Monoamine Oxidase, and drugs that help control nonmotor symptoms

A number of disorders can cause symptoms similar to those of Parkinson's disease. People with Parkinson's-like symptoms that result from other causes are sometimes said to have parkinsonism. While these disorders initially may be misdiagnosed as Parkinson's, certain medical tests, as well as response to drug treatment, may help to distinguish them from Parkinson's. However, there is currently no definitive test for PD by non-specialist clinicians, especially in the early disease stages where the symptoms may be subtle and poorly characterized. This results in a high misdiagnosis rate (up to 25% by non-specialists) and people can have the disease for many years before diagnosis. There is a need for a more accurate, objective means of early detection, ideally one which can be used by individuals in their home setting. In the last decade, focus of PD symptoms has shifted to the pre-motor symptoms, those non-motor symptoms that present years before the motor onset of the disease. The main premotor symptoms include rapid eye movement sleep behavior disorder, hyposmia, constipation and depression. Subjects with these symptoms usually are not initially seen by a neurologist, and by the time they are consulted neuronal loss in the substantia nigra is over 50%.

Changes in gait and cognition precede a diagnosis of idiopathic Parkinson’s disease, and may occur earlier than typical non-motor symptoms. Motor symptoms in idiopathic Parkinson’s disease (IPD) are identified relatively late in the disease course, taking away possible neuroprotective benefits of early treatment. Identifying individuals during the early period that precedes motor symptoms could be of great use for clinical studies seeking new therapies to prevent or delay disease progression. Compared to control subjects, IPD patients had no changes in upper-limb motor function, no depression, no sleep disturbances, no urinary symptoms, and no orthostatic hypotension (when blood pressure suddenly drops when standing up quickly). Researchers concluded that the observed changes might serve as markers to improve the early detection of IPD patients, who could then benefit from pharmacological neuroprotection trials and/or prevention trials of lifestyle-related interventions in order to delay, or even prevent, clinical manifestations. (Krohn et al. 2019)

The changes in gait in advanced PD is well studied, however, there is far less evidence when looking at early onset Parkinson's Disease. Among patients with early onset PD, the memory in gait "breaks down" and the stride-to-stride fluctuations in gait now become very similar to white noise or random fluctuations. When analyzing the data, visually, there is no difference between the original time series and the randomly shuffled time series. The DFA scaling exponent becomes close to 0.5 (the value for white noise; an absence of long-range correlations). Similar results were observed for a group of patients with PD62 demonstrating that the long-range scaling and fractal-like behavior are reduced, and the stride-to-stride fluctuations become more random. (Hausdorffa et al. 2018) While the qualitative nature of the Parkinsonian Gait is known, because of sensors we can now measure the quantitative characteristics of gait. Abnormal gait is one of the hallmarks of Parkinson's Disease, specifically Parkinsonian Gait is characterized by small shuffling movements and in some cases is present before any other symptom.

Another factor that could help in the early detection of Parkinson's Disease is keystroke data. The use of keystrokes as a means of identification has a long history. (Das et al. 2018) used keystroke dynamics while typing a computer login string to identify users with 90% to 99% accuracy. Their technique involved key hold times and latency, using a Gaussian mixture model and a neural network. This, along with many other studies show us that keystroke characteristics can be used to accurately classify the features of particular users. Important keystroke data can be found using the timing information of the key press and release events. The hold time of individual keys and the latency between keys (the time interval between pressing one key and a succeeding key) are typically used. One study found that over an extended period of time, PD affects various characteristics of hand and finger movement that can be detected early in life. A novel methodology was used to classify the subjects' disease status, by utilizing a combination of many keystroke features which were analyzed by an ensemble of machine learning classification models. When applied to two separate participant groups, this approach was able to successfully discriminate between early-PD subjects and controls with 96% sensitivity, 97% specificity and an AUC of 0.98. However, in this model, no other cardinal symptom was studied, so it may not be able to differentiate from PD and other similar movement-related disorders.

Since reliable diagnostics and early-stage detection are one of the top priorities in medical practice, much effort is being invested in the development of new methods for diagnostic support, which would increase the accuracy and minimize required time and resources. Information on the patient's condition (ON/OFF states, the occurrence of dyskinesias, falls) and symptoms (tremor, bradykinesia, freezing of gait) outside of clinical settings could provide physicians with deeper insight into patients' disease severity and progress, setting the path towards fully personalized treatment. Support tools based on the use of ML have great potential for clinical practice increasing accuracy, reliability and efficiency of clinical decision making and assessment. Machine learning approaches are frequently used for outputting a diagnostic suggestion, using data collected from various media. In addition, ML has found its application in real-time, remote monitoring and detection of PD severity, symptoms and response to therapy. Artificial intelligence, specifically machine learning, has found numerous applications in computer-aided diagnostics, monitoring and management of neurodegenerative movement disorders of parkinsonian type. These tasks are not trivial due to high inter-subject variability and similarity of clinical presentations of different neurodegenerative disorders in the early stages.

2.0 Research Question

This study will look at data collected from keystrokes and patterns as well as the gait of patients with Parkinson's Disease and patients without Parkinson's Disease to see if machine learning can pick up on any differences. Our objective is to identify large differences in patterns that can help to earlier diagnose Parkinson's Disease in the future.

Is it possible, using data collected from the activities of daily living in Parkinson's Disease patients in combination with genomic testing, to potentially aid in the diagnosis and early detection of Parkinson's Disease? How can we use computer science techniques such as supervised and unsupervised analysis and predictive analytics to help identify signatures or profiles that may help in the diagnosis and early detection of Parkinson's Disease?

3.0 Hypothesis

- H0 There are no differences in gait between patients with and without Parkinson's Disease
- H0 There are no differences in keystroke patterns between patients with and without Parkinson's Disease
- H1 PD patients will show changes in gait and can therefore be used in an accurate predictive model
- H2 PD patients will show changes in keystroke patterns and can therefore be used in an accurate predictive model

4.0 Materials and Methods

Data acquisition

Two previously defined datasets that are of public use from MIT were used in this study. The first data set looks at keystrokes in 31 subjects, 13 healthy controls and 18 patients with Parkinson's Disease. The second data set looks at gait and has 54 subjects, 30 healthy controls and 24 patients with Parkinson's Disease. In the first data set, along with the raw typing collections, clinical evaluations were also performed on each subject, including Unified Parkinson's Disease Rating Scale (UPDRS) and finger tapping tests. Each data contains a subject summary csv file which lists for each subject: pID - Patient ID, gt - Ground truth label of whether or not they had PD, updrs108 - Unified Parkinson's Disease Rating Scale part III (UPDRS-III), afTap - Alternating finger tapping result, sTap - Single key tapping result, nqScore - neuroQWERTY index (nQi), Typing speed, and file_n - The csv file(s) containing the patient's typing data. Each keystroke data csv file has four columns which give: the key pressed, the hold duration in seconds, the key release time in seconds from time 0, and the key press time in seconds from time 0.

Our next data set looks at gait. This dataset contains measures of gait from 93 patients with idiopathic PD (mean age: 66.3 years; 63% men), and 73 healthy controls (mean age: 66.3 years; 55% men). The data includes the vertical ground reaction force records of subjects as they walked at their usual, self-selected pace for approximately 2 minutes on level ground. Underneath each foot were 8 sensors (Ultraflex Computer Dyno Graphy, Infotronic Inc.) that measure force (in Newtons) as a function of time. The output of each of these 16 sensors has been digitized and recorded at 100 samples per second, and the records also include two signals that reflect the sum of the 8 sensor outputs for each foot. A subset of the database includes measures recorded as

subjects performed a second task while walking. Under usual walking conditions, variability is larger in the patient with PD (Coefficient of Variation = 2.7%), compared to the control subject (CV = 1.3%). Variability increases during dual tasking in the subject with PD (CV = 6.5%), but not in the control subject (CV = 1.2%). Using a variety of R packages such as DBI, tidyverse, dplyr, ggplot, and leaflet, we were able to upload, analyze, and visualize the data.

Machine Learning Modeling

Briefly, we used each of the datasets to evaluate different machine learning approaches that could predict the presence of Parkinson's disease based on the measurements described above. Using IBM Watson Studio, we designed a set of experiments for each of the datasets. Using AutoAI binary classification and after analysis of the data for completeness each one was loaded into the system. For each experiment we defined the cross-validation to a holdout split of 85-15. The models were tested against that 15% hold out data. The system then scanned a series of Machine Learning (ML) algorithms that adapt automatically to the type of variable and 8 different pipelines were tested and ranked for accuracy. In our case the system selected a different set of algorithms for each experiment, for those regarding gait with clinical features UPDRS, UPDRSM, and HoehnYahr score, gait without clinical features UPDRS and Hoehn Yahr score but with UPDRSM, keystrokes with clinical features UPDRS, Levodopa, MAOB (an MAO-B inhibitor), and DA and keystrokes without these clinical features. Once the algorithms were selected the data goes through a standard ML data processing pipeline. The first of that being hyperparameter optimization to tune the models to best fit the data then feature engineering to define those features that will best support the prediction and finally a second hyperparameter optimization to decide what pipeline is the most accurate. The system then evaluated the model against the holdout data. To test the validity of the prediction we measured accuracy, specificity, sensitivity, F1 score and Log Loss. The model was visualized and evaluated for overall performance using the receiver operating characteristic curve (ROC curve). Finally, each of the features was evaluated to determine how much they contribute to the classification of the samples; this was the average of 9 measures, namely, Linear Correlation (f_regression) metric, Maximal Information Coefficient (MIC) metric, Linear Regression (LR) metric, L1 regularization metric (Lasso), Ridge metric, RF metric, Stability Selection, Recursive Feature Elimination (RFE) and Recursive Feature Elimination plus selection of best number of features.

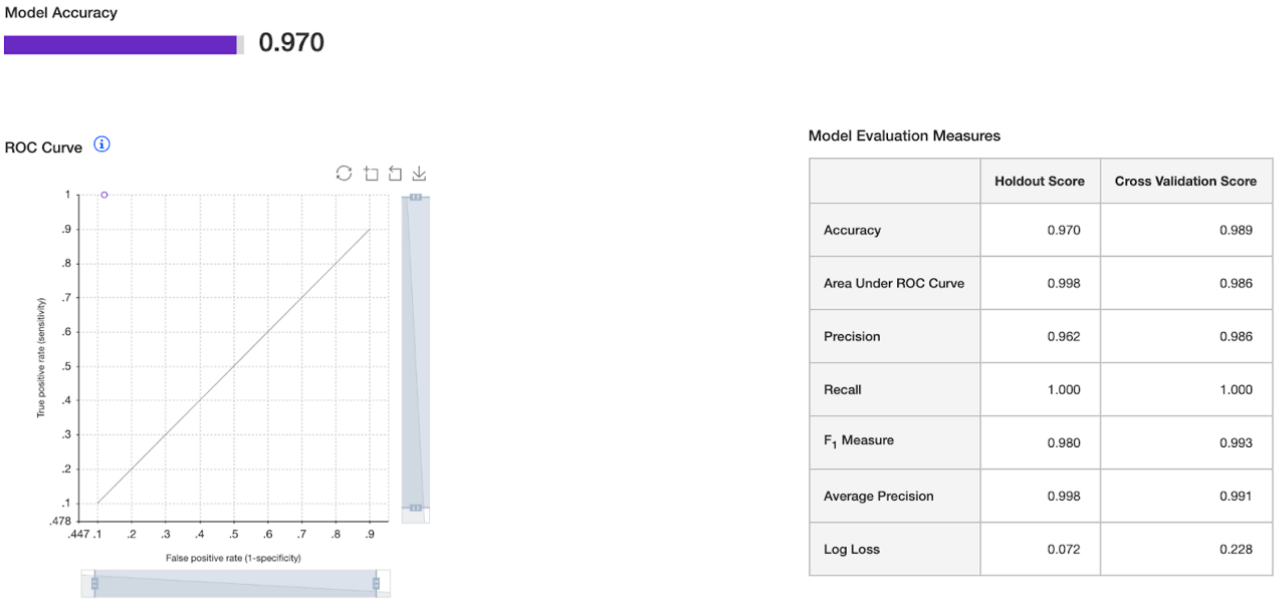
5.0 Results

The first data set that I analyzed using R code was the typing data. There were three factors that I was looking at for significance: average flight time which is the time between the release of the previous key and the press of the current key in milliseconds, average latency time which is the time between pressing the previous key and pressing the current key in milliseconds, and average hold time which is the time between the press and the release of the current key in milliseconds. There wasn't much statistical significance of any of the three factors with all p values being above .05. The p value for average flight time was 0.3144, the p value for average latency time was 0.4199, and the p value for average hold time was 0.6373. I found the p value by using a t test which is a type of inferential statistic used to determine if there is a significant difference between the means of two groups, the two groups being patients with Parkinson's Disease and healthy controls. This may be due to the fact that many of the patients with Parkinson's Disease were on a variety of drugs and treatments such as Levodopa or other dopamine enhancing medicines, which could lessen the severity of the symptoms detectable by keystroke data. Because of this, I then evaluated the P value after eliminating patients that were on Levodopa. Although none of the P values were below .05 they were substantially lower with the p value for average flight time being 0.07583, the p value for average latency time was 0.1346, and the p value for average hold time was 0.3621.

The second data set I analyzed was Gait data of patients with and without Parkinson's Disease. There were two factors that I was looking at for significance, speed in meters per second, and TUAG which is the timed up and go test, a measure of gait. There was great statistical significance with p values less than 0.05 and very close to zero. The p value for speed in meters per second is 1.788e-11. The p value for TUAG is 2.339e-08. Both of these p values are extremely close to zero and show us that these two factors of gait are extremely significant to the detection of Parkinson's Disease.

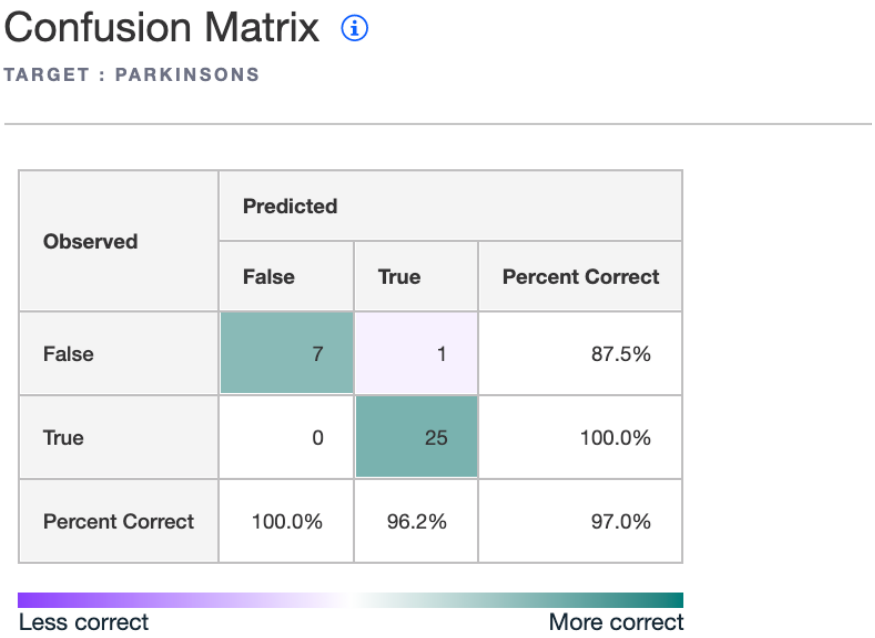
We then created predictive models using IBM Watson Studio/IBM Cloud Pak and a combination of R code and Python. The accuracy of the data set is out of 1 with 1 being 100% accurate. There are multiple pipelines that use different algorithms and are ranked based on their accuracy. For keystroke data without the clinical features taken into account, pipeline 1 was the most accurate with an accuracy of 0.989 using a cross validation score and an accuracy of 0.970

using a holdout data score. Pipeline 1 used the Random Forest Classifier algorithm and used no enhancements. The largest feature of importance was impact, with its importance being about 94%, then average hold time with 5% importance, then average flight time with 1% importance, then average latency time with 0% importance. The next predictive model used the same keystroke data but did take into account the clinical data. With the clinical data taken into account, the most accurate pipeline was pipeline 1 with an accuracy of 0.989 using a cross validation score and an accuracy of 0.970 using a holdout data score. Pipeline 1 used the Random Forest Classifier algorithm and used no enhancements. The largest feature of importance was impact with its importance being 62%, then sided at 8%, then average hold time at 7%, then whether or not the patient was on levodopa at 6%, then whether the patient was on other PD medication at 6%, and whether or not they had tremors at 5%.



	Holdout Score	Cross Validation Score
Accuracy	0.970	0.989
Area Under ROC Curve	0.998	0.986
Precision	0.962	0.986
Recall	1.000	1.000
F_1 Measure	0.980	0.993
Average Precision	0.998	0.991
Log Loss	0.072	0.228

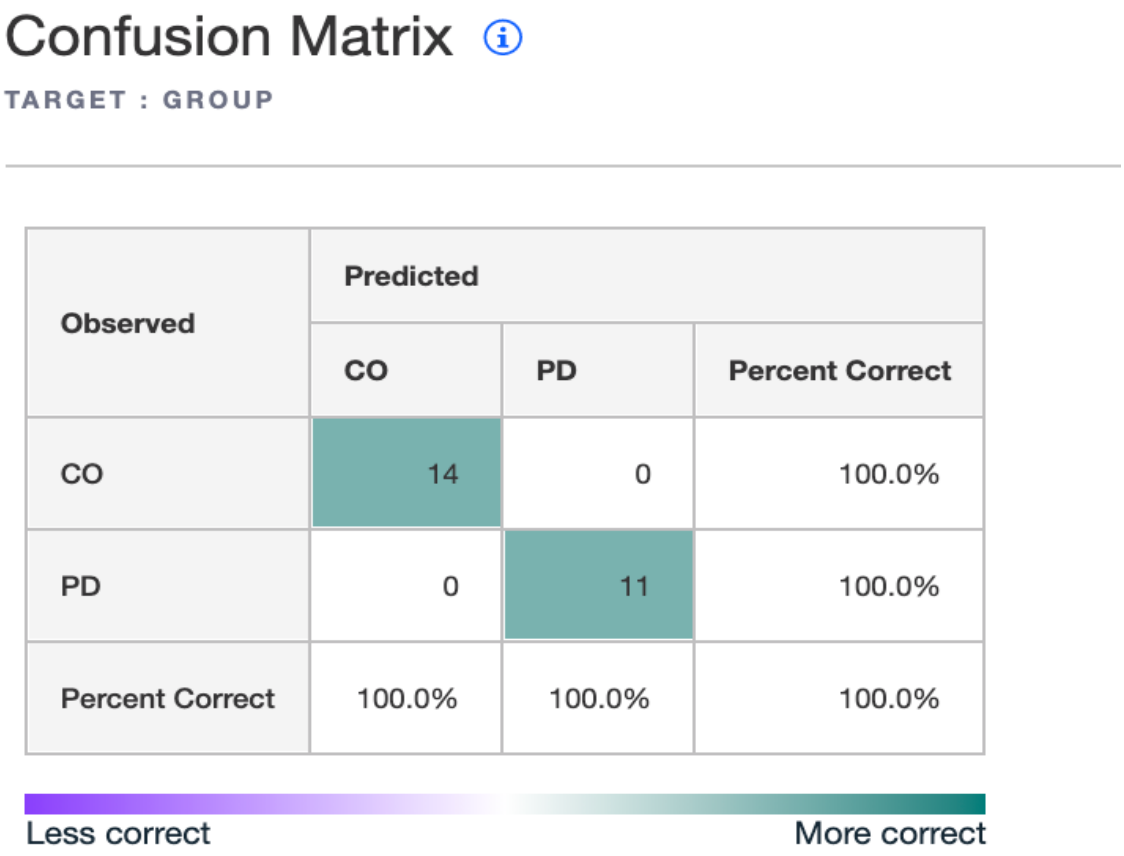
(Figure 1: Model accuracy, keystroke data w/ clinical features)



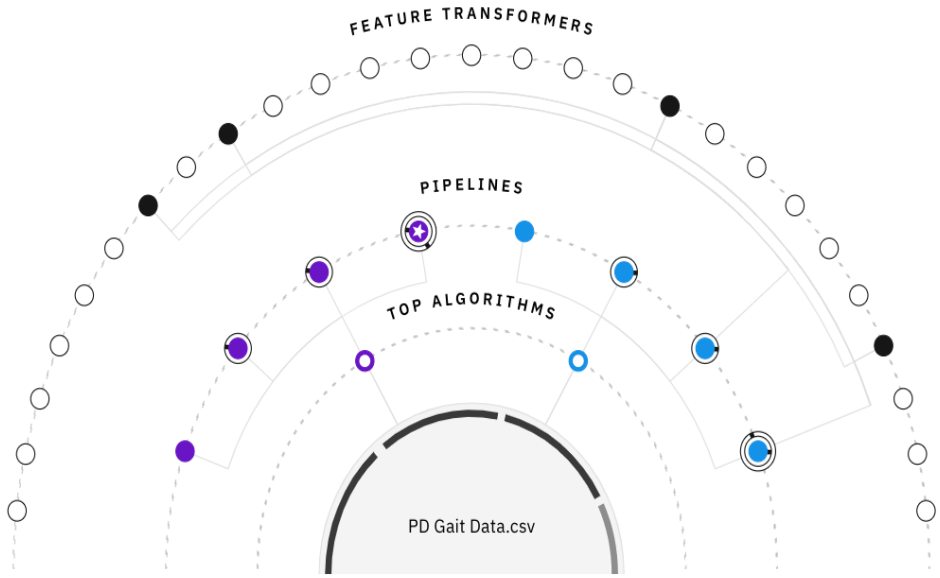
(Figure 2: Confusion matrix, keystroke data w/o clinical features)

The other predictive model used the gait data. The first gait model was made without taking clinical data into consideration. Using the cross validation score the most accurate pipeline was pipeline four which used a gradient boosting classifier algorithm and had an accuracy of 0.957. Using holdout score the most accurate pipeline was pipeline two which also used gradient boosting classifier and had an accuracy of 1.000. Pipeline two also used the enhancement HPO-1 or 1st hyperparameter optimization. Looking at pipeline two, the most important features were the UPDRSM score, which is a score of the severity of the degradation of motor functions of Parkinson's Disease, which was 86% of the feature importance. TUAG was 9% of the importance and speed in meters per second was 5%. For the predictive model that did use clinical data, the most accurate pipeline using cross validation score was pipeline two with an accuracy of 0.986 and the enhancement HPO-1. Pipeline two used a gradient boosting classifier algorithm. Using the holdout score the most accurate pipeline was pipeline five with no enhancements and an accuracy of 1.000. Pipeline five used an LGBM algorithm. Looking at pipeline 5, the most important

features were UPDRS at 44%, UPDRSM at 15%, Speed_01 in meters per second at 10%, age at 9%, weight in kg at 8%, and TUAG at 7%.



(Figure 3: Confusion matrix, gait w/ clinical features)



(Figure 4: feature transformers, gait w/o clinical features)

6.0 Discussion

Looking at both of these data sets, we found that gait is a more predictive factor of Parkinson's Disease and is potentially more useful in aiding in the early detection of it. However, since patients in the keystroke dataset were receiving medication it's hard to remove that confounding factor and we cannot discard it from playing a larger role than seen in this data set. In this small but significant set of patients we show both conceptually and empirically that we can use ML algorithms to predict the presence or absence of PD. The addition of clinical features further refines the predictive power of the gait and keystroke data. This supports the idea that measurements do not happen in a vacuum of context and that providing that context is important to both build better models and aid in the clinical validity and interpretation of those models. Keystroke data is also very prone to human error or interference since typing speed can differ depending on age and socioeconomic status and whether one grew up with access to a computer or typing classes. The gait data shows a stronger significance. It allowed me to reinforce that both UPDRS and UPDRSM are extremely accurate measures of Parkinson's Disease as they both were

very important in the predictive models for gait. The data with gait is extremely promising, showing strong correlations and statistical significance between TUAG and walking speed and Parkinson's Disease.

Limitations

There are several inherent limitations to ML models in clinical applications. Data collection is a limiting factor as recruiting patients into formal studies is a time-consuming process. Machine Learning models by definition require vast amounts of data to be able to discern accurately between populations. And lastly biological systems in general have an unlimited number of confounding factors that make the interpretation of the models, in some cases, speculative at best. These may include age, weight, gender, hormone levels, or other diseases. It should also be said that all models run the risk of bias and the proper design and balance of the dataset id paramount to their utility in a real-world setting.

7.0 Conclusion

In my research I have found that aspects of gait specifically speed and TUAG are significant in the study of Parkinson's Disease. The addition of clinical features to specific measures of activities of daily living enhances the predictive capacity of ML models in PD. One of my largest barriers was a difference in variables and lack of control of these variables in both data sets. In order to continue this research and further examine the role that gait, and keystroke patterns can play in the early detection of Parkinson's Disease, more studies must be done with more regulated controls. These controls must not be on dopamine enhancing drugs such as levodopa and must be of the same age to truly hone in on the role that gait and keystroke patterns play in detecting the disease.

Along with advances in treatment, advances in the accuracy of diagnosis and timing of diagnosis can greatly improve the quality of life of someone living with Parkinson's Disease. Diagnosing PD early is important because treatments such as levodopa/carbidopa are more effective when administered early on in the disease. Non-pharmacologic treatments, such as increased exercise, are also easier to perform in the early stages of PD and may help slow down disease progression. In the future, I hope to collect data (when safe), and continue to use machine learning algorithms to analyze the correlation between gait and keystroke patterns and

Parkinson's Disease. I also hope to be able to work this data along with genomics data to further analyze diagnosis and make it sooner and more accurate.

Acknowledgments:

I would like to thank my mentor Dr. Fernando José Suarez Saiz, Medical Science Lead- IBM for all of his help in my project. He has been a great support throughout my entire project and always was there if I needed any help while allowing me to come up with a unique and personal project and always wanting to hear my insight. I would also like to thank my parents and my teachers Ms. O'Brien, Mr. Inglis, and Ms. Dyer for all of their help and support throughout my three years.

Bibliography

Kelly, J., Moyeed, R., Carroll, C., Albani, D., & Li, X. (2019). *Gene expression meta-analysis of Parkinson's disease and its relationship with Alzheimer's disease. Molecular Brain, 12(1).* doi:10.1186/s13041-019-0436-5 https://molecularbrain.biomedcentral.com/articles/10.1186/s13041-019-0436-5

Dinov, I. D., Heavner, B., Tang, M., Glusman, G., Chard, K., Darcy, M., … Toga, A. W. (2016). *Predictive Big Data Analytics: A Study of Parkinson's Disease Using Large, Complex, Heterogeneous, Incongruent, Multi-Source and Incomplete Observations. PLOS ONE, 11(8), e0157077.* doi:10.1371/journal.pone.0157077 https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0157077

Pang, S. Y.-Y., Ho, P. W.-L., Liu, H.-F., Leung, C.-T., Li, L., Chang, E. E. S., … Ho, S.-L. (2019). *The interplay of aging, genetics and environmental factors in the pathogenesis of Parkinson's disease. Translational Neurodegeneration, 8(1).* doi:10.1186/s40035-019-0165-9 https://translationalneurodegeneration.biomedcentral.com/articles/10.1186/s40035-019-0165-9

Zhan, A., Mohan, S., Tarolli, C., Schneider, R. B., Adams, J. L., Sharma, S., … Saria, S. (2018). *Using Smartphones and Machine Learning to Quantify Parkinson Disease Severity. JAMA Neurology, 75(7), 876.*doi:10.1001/jamaneurol.2018.0809 https://sci-hub.tw/10.1001/jamaneurol.2018.0809

J. Jankovic, "Parkinson's disease: clinical features and diagnosis," J. Neurol. Neurosurg. Psychiatry, vol. 79, no. 4, pp. 368–376, 2008.

Norden, A. D., Dankwa-Mullan, I., Urman, A., Suarez, F., & Rhee, K. (2018). Realizing the Promise of Cognitive Computing in Cancer Care: Ushering in a New Era. JCO Clinical Cancer Informatics,(2), 1-6. doi:10.1200/cci.17.00049

Health Databases and Health Database Organizations: Uses, Benefits, and Concerns. (n.d.). doi:10.3897/bdj.4.e7720.figure2f

Weidlich, V., & Weidlich, G. A. (2018, April 13). Artificial Intelligence in Medicine and Radiation Oncology. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5999390/

Pesapane, F., Codari, M., & Sardanelli, F. (2018, October 24). Artificial intelligence in medical imaging: Threat or opportunity? Radiologists again at the forefront of innovation in medicine. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6199205/

Galimova, R., Buzaev, I., Ramilevich, K., Yuldybaev, L., & Shaykhulova, A. (2019, January 09). Artificial intelligence-Developments in medicine in the last two years.

Retrieved December 16, 2020, from https://www.sciencedirect.com/science/article/pii/S2095882X18300938

Institute of Medicine (US) Committee on Regional Health Data Networks. (1994, January 01). Health Databases and Health Database Organizations: Uses, Benefits, and Concerns. Retrieved December 16, 2020, from https://www.ncbi.nlm.nih.gov/books/NBK236556/

CE;, K. (n.d.). From Images to Actions: Opportunities for Artificial Intelligence in Radiology. Retrieved December 16, 2020, from https://pubmed.ncbi.nlm.nih.gov/29155645/

Saiz F. S. (n.d.). US20180113867A1 - Natural language processing review and override based on confidence analysis. Retrieved December 16, 2020, from https://patents.google.com/patent/US20180113867A1/en

Ramesh, V., Glass, R., & Vessey, I. (2003, September 28). Research in computer science: An empirical study. Retrieved December 16, 2020, from https://www.sciencedirect.com/science/article/abs/pii/S0164121203000153

Malik, S. (2019, April 17). Data Analysis and Visualisations using R. Retrieved December 16, 2020, from https://towardsdatascience.com/data-analysis-and-visualisations-using-r-955a7e90f7dd

Grenn, F., Kim, J., Makarious, M., Iwaki, H., Illarionova, A., Brolin, K., . . . On behalf of the International Parkinson's Disease Genomics Consortium (IPDGC). (2020, January 01). The Parkinson's Disease GWAS Locus Browser. Retrieved December 16, 2020, from https://www.biorxiv.org/content/10.1101/2020.04.01.020404v1

Li, Y., Wong, G., Humphrey, J., & Raj, T. (2019, March 01). Prioritizing Parkinson's disease genes using population-scale transcriptomic data. Retrieved December 16, 2020, from https://www.nature.com/articles/s41467-019-08912-9

Adams, W. (2017, November 30). High-accuracy detection of early Parkinson's Disease using multiple characteristics of finger movement while typing. Retrieved December 16, 2020, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5708704/

Nalls, M., Blauwendraat, C., Vallerga, C., Heilbron, K., Bandres-Ciga, S., Chang, D., . . . Zhang, F. (2019, November 06). Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: A meta-analysis of genome-wide association studies. Retrieved December 16, 2020, from https://www.sciencedirect.com/science/article/abs/pii/S1474442219303205

Catarina Silva, M. (2019, January 10). Changes in Gait, Cognition May Be Early Signs of Idiopathic... Retrieved December 16, 2020, from https://parkinsonsnewstoday.com/2019/01/10/change-gait-early-sign-idiopathic-parkinsons-disease-study/