Descriptive Symbolic Models of Gaits from Parkinson's Disease Patients

James Alexander Hughes
St. Francis Xavier University
Department of Computer Science
Antigonish, Nova Scotia, Canada
jhughes@stfx.ca

Sheridan Houghten
Brock University
Computer Science
St. Catharines, Ontario, Canada
shoughten@brocku.ca

Joseph Alexander Brown
Innopolis University
AI in Games Development Lab
Innopolis, Republic of Tatarstan, Russia
j.brown@innopolis.ru

***Abstract*—Parkinson's disease (PD) is a degenerative disorder of the central nervous system that has many debilitating symptoms which affect the patient's motor system and can cause significant changes in their gait. By using genetic programming, we aim to develop descriptive symbolic nonlinear models of PD patient gait from time series data recorded from pressure sensors under subjects' feet. When compared to popular types of linear regression (OLS and LASSO), the nonlinear models fit their data better and generalize to unseen data significantly better. It was found that models developed for healthy control subjects generalized to other control subjects well, however the models trained on subjects with PD did not generalize well to other PD patients, which complicates the issue of being able to detect the progression of the disease. It is suspected that health care professionals can have difficulty classifying PD due to a lack of accurate data from patient reports; having individually trained models for active monitoring of patients would help in effectively diagnosing PD.**

***Index Terms*—Gait; Genetic Programming; Linear Regression; Parkinson's Disease; Symbolic Regression; Time series.**

I. Introduction

Parkinson's disease (PD) is a neurodegenerative disorder affecting over 4 million people globally. Roughly 1% of the population over the age of 60 and 3% over the age of 80 have PD [16]. PD symptoms include tremors, depression, hallucinations, cognitive decline, falls, and changes to gait [3]. As PD progresses, symptoms become worse and individuals may lose their independence. This reduces quality of life for patients and their caregivers. The estimated annual cost of PD in 2002 was $22 billion and is expected to rise beyond $50 billion by 2040 [3]. Most costs are related to inpatient care and other indirect costs (caregivers) [3].

PD is known to be problematic to diagnose [15]. Many of the motor symptoms, currently the primary means of diagnosis, are also present in other diseases [14]. There are several genes believed to be associated with the disease and these, along with environmental risk factors, interact in a complex manner [6], [8], [13]. It is argued in [15] that accurate diagnosis of PD requires multiple types of biomarkers.

Changes in gait are commonly noted in individuals with PD. Such changes include speed, stride, and rhythmicity [4], [5], [7], [17], [20]. PD patients under a cognitive load also have a higher difficulty performing motor tasks [20].

978-1-7281-1462-0/19/$31.00

In a previous study [11], Genetic Programming (GP) was used to develop models of human gait via data from a smartphone. By creating models of subjects walking, the technology was capable of identifying individuals. In followup work, the same GP system was used to model data recorded from a smartwatch. These models were capable of differentiating individuals and tasks (running, etc.) and provided insight into the underlying complex system [12].

In the current study, we aim to use GP to develop similar models for PD patients. To the best of our knowledge, this is the first such analysis to date. We use GP to perform *symbolic regression*, a type of nonlinear regression analysis. With symbolic regression the programs being evolved by GP are mathematical expressions regressing to data. Brief details on the GP system used are presented in Section III.

In [9] it is stressed that there is a need for *objective* biomarkers, as opposed to information from diaries, questionnaires, etc. that are subjective and rely on memory. As such, a long-term goal of this project is to use the generated models for classifying PD and non-PD patients and to actively monitor at risk individuals. As an early step towards this goal, our objective in the current study is to develop *descriptive* and *symbolic* models of the data. Unlike black box models, GP produces models that can be analyzed with relative ease, thus providing a way to learn more about the underlying complex and dynamic system being modelled. An analysis of the results can be found in Section IV; however, performing a thorough analysis of the descriptive symbolic models for the purposes of understanding the mechanics of the underlying system is, at this stage, beyond the scope of the work.

II. Data

Data were obtained from *PhysioNet*, an online collection of physiologic data. For this study we focus on the *Gait in Parkinson's Disease* project [1] containing data from studies by Yogev *et al.* [20], Hausdorff *et al.* [7], and Frenkel-Toledo *et al.* [4], [5]. There are recordings from 93 PD patients (mean age of 66.3 years old; 63% male) and 73 healthy control subjects (mean age of 66.3 years old; 55% male). Individuals were instructed to walk in their usual way at a self-selected pace for two minutes. An *Ultraflex Computer Dyno Graphy* device with eight sensors was placed under each foot (16

TABLE I: Sensor locations

Sensor	x	y	Sensor	x	y
L1	-500	-800	R1	500	-800
L2	-700	-400	R2	700	-400
L3	-300	-400	R3	300	-400
L4	-700	0	R4	700	0
L5	-300	0	R5	300	0
L6	-700	400	R6	700	400
L7	-300	400	R7	300	400
L8	-500	800	R8	500	800

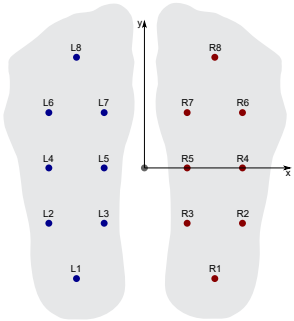

Fig. 1: Approximate Sensor Locations Relative to Foot Outline

in total) and a time series of vertical ground reaction force was recorded in Newtons. Sensors recorded at a frequency of 100Hz. Data was stored in files with rows representing time points and columns representing sensors. In addition to the 16 sensors, the files include two additional columns representing the sum of the sensors' values for each foot.

Approximate sensor locations are found in Table I and Figure 1 presents an illustration of these locations. The origin $(0, 0)$ is located between the legs with the positive direction of the y-axis extended out in front of the individual. Note that coordinates x and y represent relative positions of the sensors within each insole, according to an arbitrary coordinate system. As specified in the note on data formatting in [1], this coordinate system is designed to calculate a proxy for the location of each foot's centre of pressure.

Metadata, such as age, weight, height, sex, speed, and measures of PD severity (*Hoehn & Yahr* or *Unified Parkinson's Disease Rating Scale*) is also included for many subjects.

A. Preprocessing

Here we focus on the data recorded by Yogev *et al.* [20], the first set of data recorded from the collection. This set contains data from 64 subjects (29 control and 35 PD patients). Data was processed by 1) keeping only the data from the 16 sensors, 2) z-score normalizing the data, and 3) breaking each subject's recording into five subsets of data.

Although the summed force values for each foot included in the data could easily be incorporated into the symbolic models, they were eliminated from our regression to ensure no violation of the multicolinearity assumption of regression analysis. Given the nature of the regression, a perfect fitness could be obtained by simply including the summed feature in the model with a subtraction of all sensors except the dependent variable (see Section III-B).

Data was z-score normalized to enable an easier between-subject comparison of the models being generated, since the recorded data may not have been scaled in a meaningful way.

Each subject's two-minute recording was divided into five subsets of data, each 24s in length. This was done both to reduce the runtimes required to generate a given model and to provide unseen data for simple testing and validation. Since there were sixty-four subjects studied, a total of 320 datasets were available after the subjects' data was split.

III. Methods

A. Genetic Programming Implementation

A specialized GP system for symbolic regression was used in this work [10]. With the exception of the system parameters (discussed in Section III-C), the GP system was identical to the one used in [12] and was ultimately based on Schmidt *et al.*'s work. The GP system incorporated many improvements, most notably *fitness predictors* and an *acyclic graph representation.*

Fitness predictors reduce overfitting and the amount of computation required to evaluate the fitness of a population by approximating the local search gradient [19]. This is done by *evolving* a relatively small subset of data for fitness evaluation alongside the candidate solutions that both generalize the whole dataset and create disagreement among candidate solutions — the idea being that data points that create disagreement between the candidate solutions need more attention.

There are many alternatives to the traditional tree representations used in GP systems. In this work we use an acyclic graph representation which has been shown to reduce overfitting and bloat, scales well, increases quality of results, and has the ability to use possibly useful subexpressions [18].

B. Modelling Strategy

Our modelling strategy is to regress to some dependent variable y to generate a function $\hat{y} = f(X)$ where $\hat{y}$ is a predictor for y and X are independent variables. Here we use sensor $R8$ (front of the right foot) as our y and all other sensors as X. The choice of R8 was ultimately arbitrary, however our expectation is that similar results would be obtained with other sensors being selected as the dependent variable.

Although we are building a model that *predicts* y, this is not the real goal. The long-term aim is to find a symbolic and interpretable expression that defines a trajectory of the metastable complex system through the high-dimensional space. Assuming the generated models are representative of the underlying system, data from a recording of a subject walking can then be applied to a model and the difference between what is expected (trajectory defined by the model) and what was observed can be calculated. This enables a simple mechanism to measure differences between certain individuals or cohorts. For example, one might expect data recorded from a healthy subject to match the trajectory defined by a model fit to a healthy subject, while data from a PD patient might not match the trajectory defined by a model fit to a healthy subject.

This was seen in previous work modelling subjects performing various tasks: almost all data recorded from running-like tasks would match the trajectory of models fit to other running-like tasks, but not models fit to walking-like tasks, and *vice versa* [12]. These symbolic models can also be analyzed relatively easily (when compared to other black-box modelling) to learn something about the underlying system.

Ordinary Least Squares (OLS) and *Least Absolute Shrinkage and Selection Operator* (LASSO) regression are used for linear regression and GP is used to perform nonlinear symbolic regression. OLS was selected as it is the quintessential linear regression strategy and LASSO was selected as it also performs feature selection, one of the major benefits of using GP for symbolic regression.

Linear models generated are of the form $\hat{y} = \beta_1 \cdot X_1 + \beta_2 \cdot X_2 + \cdots + \beta_m \cdot X_m + C$, where X are independent variables, β are coefficients, m is the number of independent variables, and C is some constant. Since symbolic regression is not restricted to this form, the resulting models are $\hat{y} = f(X_1, X_2, \ldots, X_m)$, where the function f is some combination of the independent variables and linear and nonlinear basis functions defined by the GP system's language (discussed in section III-C).

C. Algorithm Execution

Table II contains a summary of the system settings for the GP system. These values were determined empirically and based on previous work, however no significant parameter sweep was done. The mutation rate was set higher than what is typical as a consequence of the representation used — the representation allows for many noncoding genes and mutations to these noncoding genes may have no immediate impact.

In addition to the typical parameters, *Max # Graph Nodes* refers to the number of unique nodes allowed in the acyclic graph representation [18], and the details on *Predictors* and *Trainers* refer to important settings for fitness predictors [19]. These values are included for reproducibility. Full details on these settings can be found in the original sources [18], [19].

The GP language was selected to be *at least* as expressive as linear regression. It contains the operators linear regression uses and a collection of nonlinear operators: e for natural exponentiation, absolute value for point nonlinearities, and trigonometric functions since any periodic function can be expressed as a sum of *sine* functions. Originally *ln* was included, however it was excluded since it caused domain errors. Not only are nonlinear basis functions included, but the model is not restricted to the ridged form of linear regression with sums of scaled independent variables; the generated nonlinear models can be any combination of the basis functions allowed by the language.

Unlike OLS and LASSO which are deterministic, symbolic regression is stochastic and will produce different, but typically similarly effective models every time it is executed. For this reason, 100 nonlinear models were generated for each dataset for both statistical significance and to increase the chances of producing a collection of high-quality models. This does however mean that we have 100 nonlinear models for every 1 linear model generated of each type (OLS and LASSO) which can make for an unfair comparison between the nonlinear and linear models. The authors take care to acknowledge this where appropriate.

TABLE II: Summary of the system parameters

Elitism	1 (Single top candidate solution)
Population	101
Subpopulations	7
Generations	100,000 (1,000 per migration)
Migrations	100
Crossover	80%
Mutation	10% (x2 chances)
Fitness Metric	Mean Squared Error: $\frac{1}{n}\sum_{i=1}^{n}(\hat{y_i} - y_i)^2$
Language	$+, -, *, /, exp, abs, sin, cos, tan$
Max # Graph Nodes	128
Predictors	10
Predictor Pop. Size	10% of whole dataset
Trainers	8

Given the 320 datasets (5 subsets for each of the 64 subjects) and the 100 nonlinear models generated for each, a total of 32,000 nonlinear models, 320 OLS linear models, and 320 LASSO linear models were generated.

The generation of each nonlinear model took one hour on average while running threaded on 8 cores. Multiple clusters were used for computation and the processors on the clusters were *Intel Broadwel* (2.1GHz) and *Intel Skylake* (2.4GHz).

IV. Results and Discussion

Figure 2 shows the recorded signal from sensor R8 from a subject alongside nonlinear and linear models fit to the first 24s of that subject's data. The large top plot is the section of the signal that the models were fit to (training data) and the smaller four plots are segments of the signal that the model had not seen before (testing data — the other four subsets from the recording). Visually it can be seen that all models fit the signal well, however the nonlinear models match the signal better than the linear models, especially where the signal flat-lines in the cycle. The *mean absolute error* (MAE) of the models from the signal are included in the legends in the figure. Given that there were 100 nonlinear models, we naively selected the one top model of the 100 based on model error from the signal the models were fit to — no rigorous model selection was performed. Overfitting is likely occurring with this model selection strategy, and indeed the MAE values included do suggest this is occurring; as one would expect, there is a much smaller error on the training data. Despite this, not only did the nonlinear models fit the data they were trained on better than the linear models, but they generalized to the unseen data better than the linear models. It is expected that with a more meticulous model selection strategy the nonlinear models may generalize even better to unseen data from the same subject.

Table III includes summary statistics of the models. This table includes median, interquartile range (IQR), and a Mann-Whitney U test probability values comparing the errors obtained by the nonlinear models and the respective column's

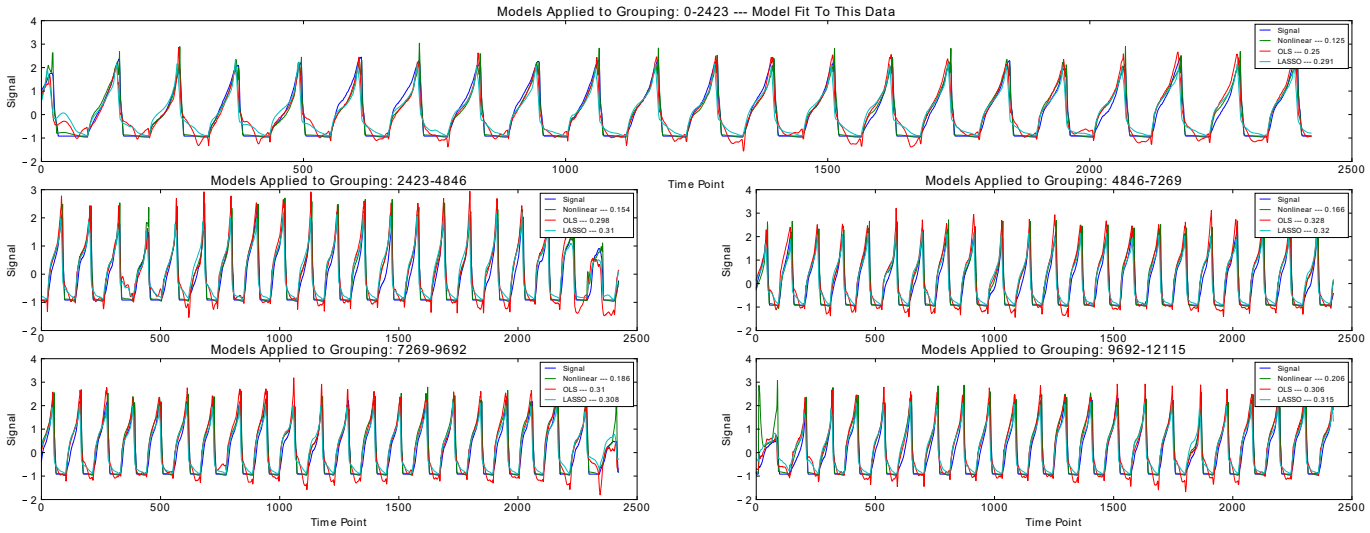

Fig. 2: Two linear and one nonlinear model plotted against the recorded signal. Mean absolute errors are presented in each plot's legend. The top plot shows the models applied to the data they were fit to. The bottom 4 plots show the models when applied to unseen data from the same subject.

linear models (p-value). On occasion the nonlinear models would return *Not a Number* (NaN) values due to computational limitations. Ithese situations the NaN values were simply ignored. The training data was the data to which the models were regressed, and the testing data consisted of the other four datasets from the same subject. These results show that the nonlinear models fit the data significantly better than the linear models, however the authors do acknowledge that the nonlinear models benefit from the stochastic nature of the generation process and that there was a large pool of models to select from. This is important to emphasize as this stochastic process can sometimes be seen as a hindrance, and at other times it is a benefit as it allows for the easy generation of many different but similarly effective models.

Figure 3 contains three matrices showing similar data, but with important differences. The first two matrices show the beta weights from the linear models generated for each dataset. The bottom matrix shows the percentage of times a given feature appeared in the 100 nonlinear models generated by symbolic regression. These matrices can be used as a proxy for feature importance in the underlying system when the models are being regressed to sensor R8 (the bottom sensor in the matrices). Each matrix is split in half. The left sides are the features from the models fit to control subjects and the right sides are the features from the models fit to the PD patients in order of increasing PD severity based on the *Unified Parkinson's Disease Rating Scale* (UPDRS).

Based on the differences in the features observed between the different tasks in previous work modelling data recorded from individuals performing different walking and running tasks [12], it was expected that we would observe consistency in the features within the control and PD cohorts and differences between them; however, in all three matrices there are no obvious differences. What is observed is a consistency between almost all subjects' models making use of sensors L1–L3 and R5–R8 while not consistently using the other sensors.

When referring to Table I and Figure 1 containing the sensor locations, we can see that sensors L1–L3 are the back part of the left foot and R5–R8 are in the front of the right foot. Although at first this was surprising, when considering the actual physical manifestation of walking, and the fact that we chose R8 as the dependent variable to regress to, the front of the left foot will not likely be in contact with the ground at the same time as the front of the right foot, and similarly for the heels of the subject's feet. Sensor R8 was arbitrarily chosen as the dependent variable, and although we expect similarly effective models to be generated with a regression to a different sensor, we would likely obtain models containing different features that provide additional insights.

Having the generated models reflect the physical manifestation of the task provides some assurance that the models are representative of the underlying system and can be used to understand some of the system's mechanics. The authors do however emphasize that their original expectation of seeing differences between the cohorts was not realized.

Figure 4 contains matrices of MAE values when a given model (column) is applied to some dataset (row). Error values above 1 were truncated to help with visualization (error values of 1 should be understood to simply mean a poor fit). The top two matrices were generated with the OLS linear models and the bottom two with the nonlinear models. The diagonal represents the case when the models were applied to the data they were fit to. The solid black vertical and horizontal lines divide control subjects (CO) from the PD patients (Pt). PD patients are ordered based on an increasing PD severity based on the UPDRS. Left matrices show the resolution of all models applied to all data (320x320 matrix) and for each of these, the corresponding right matrix shows the median error when

TABLE III: Summary statistics for the nonlinear and linear models.

	Nonlinear		OLS			LASSO		
	Median	IQR	Median	IQR	p-value	Median	IQR	p-value
Training	0.108	$\pm$ 0.036	0.220	$\pm$ 0.112	$1.625{*}10^{-78}$	0.274	$\pm$ 0.112	$8.577{*}10^{-104}$
Testing	0.132	$\pm$ 0.052	0.250	$\pm$ 0.121	$5.540{*}10^{-274}$	0.284	$\pm$ 0.112	0.00

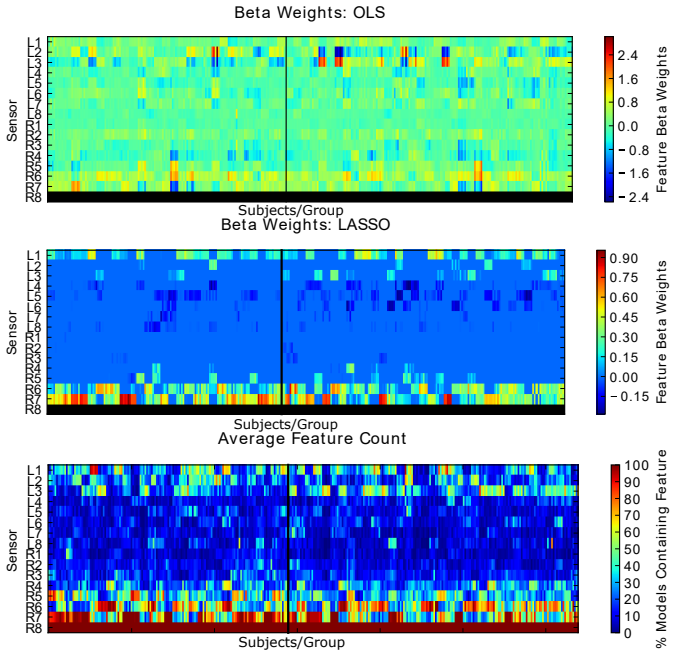

Fig. 3: The top matrix shows the beta weights from all OLS linear models generated for each dataset. The middle shows the same as above but for LASSO linear regression. The bottom matrix shows the number of times the respective row's feature appeared in the 100 models generated for each dataset.

applying all models from the same subject to all data from the same subject (64x64). LASSO linear models were excluded here since their results were similar to those obtained by OLS.

Similar to the expectation discussed above regarding a difference in which features would be present in the models, based on previous work [12] we expected a difference between the MAEs when applying models fit to control subjects to other control subjects' data versus models fit to control subjects and applied to PD patients' data (and *vice versa*); however, this was not easily observed. In the previous work the differences between the tasks could be observed visually, but here there appears to be little consistency.

There is some consistency between certain collections of subjects. For example, nonlinear models fit to Pt39 and Pt31 do not fit other subjects well, which suggests that their physical manifestation of walking may be markedly different. Additionally, nonlinear models for Pt19, Pt16, Pt10, Pt13, Pt35, Pt12, Pt27, Pt16, and Pt14 fit each others' data similarly well, which is perhaps reasonable as these subjects had similar UPDRS values; however, this trend cannot be generalized.

The area with the highest relative consistency in these matrices is the upper left sections, which corresponds to the MAEs obtained when applying models fit to control subjects to data from control subjects. Given the nature of PD symptoms and assuming the models are representative, this could be explained by a relative consistency in the data between control subjects and high inconsistency between all PD subjects. This would result in little consistency between PD subjects' models to each others' data and to control subjects' data. Further, control subjects' models would not fit the inconsistent PD subjects' data.

The MAE values from each of the four segments (quadrants) were collected into separate groups representing when a cohort's set of models were applied to a cohort's set of data: Control on Control, Control on PD, PD on Control, and PD on PD. For example, PD on Control would be the group of MAEs when all the models fit to PD subjects were applied to control subjects' data (the top right quadrant of the matrices seen in Figure 4). After outliers (any error values above 2.0) were eliminated, Control on Control had 20,708 and 21,025 MAE values for the nonlinear and linear models respectively, Control on PD had 24,914 and 25,375, PD on Control had 24,577 and 25,375, and PD on PD had 29,540 and 30,625.

Figure 5 presents comparisons between the four quadrants' distributions of MAEs for both the OLS linear models and the nonlinear models. The first observation is that the distributions from the nonlinear models are much closer to zero and skewed towards zero relative to the linear models. This shows that not only do the nonlinear models generalize to unseen data from the same subject well (as presented in Table III), but they also generalize to unseen data from the same cohort well.

The important question remains: do the models generalize to unseen data from a *different* cohort well? Our original expectation is that they should *not* generalize to the different cohort's data. However, based on a visual inspection of the distributions for both the OLS linear and nonlinear models, it can be seen that the models *do* seem to generalize reasonably well to data from a different cohort. Further, not only do they generalize well, they generalize similarly well — or better — compared to the models actually fit to that cohort, especially for the nonlinear models. For example, consider the second distribution comparing Control on Control versus PD on Control with the nonlinear models. The PD on Control distribution appears similar to the Control on Control, despite not being fit to the control subjects' data; however, when actually performing a Mann-Whitney U test to compare these distributions a p-value of 0 is obtained.

Tables IV, V, and VI show all p-values comparing all quadrants' distribution of MAE values for the OLS, LASSO, and nonlinear models. In all cases there was a significant difference (although with small effect sizes) between the

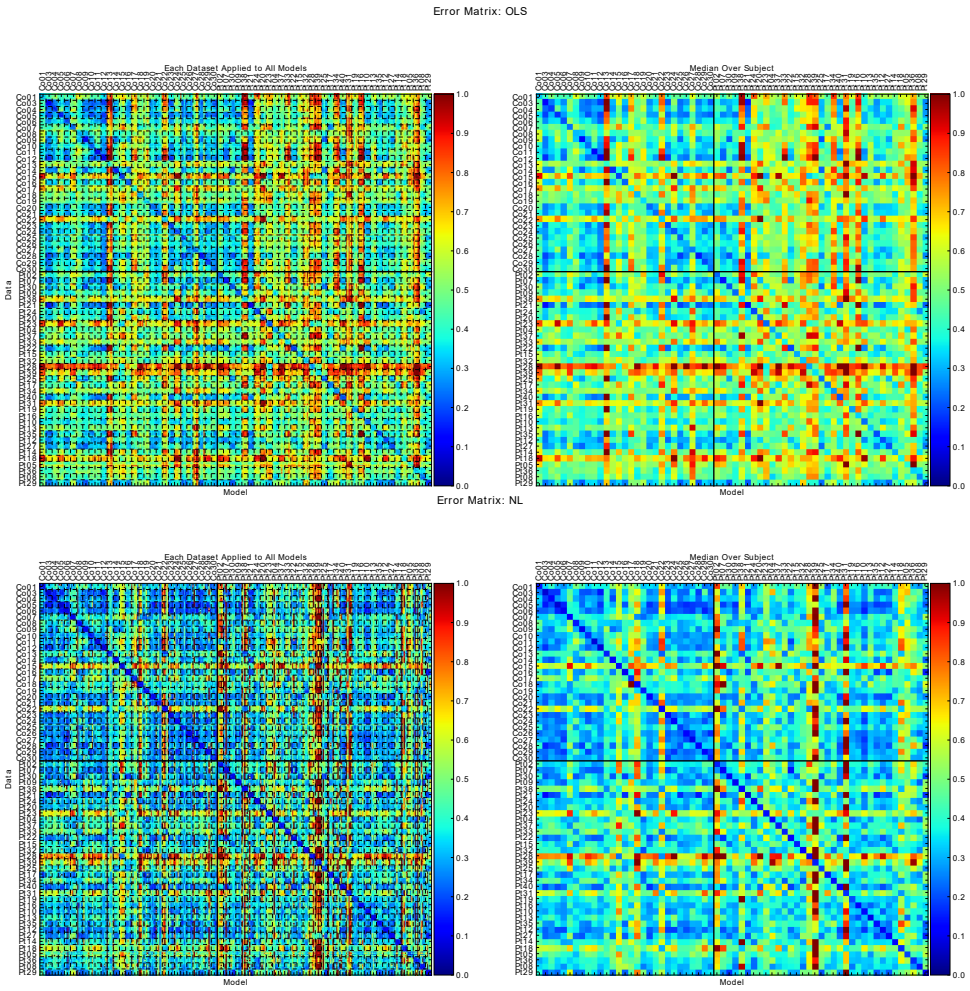

Fig. 4: Mean absolute error matrix generated by applying all models generated to all datasets. The top pair were generated with OLS linear models and the bottom were from nonlinear models.

distribution of errors except for when comparing PD on PD versus PD on Control for the nonlinear models, which had a p-value of approximately 0.07. Consider the meaning of this result: the most similar distributions, which also happen to be the distributions with the highest (worst) median MAE, are those generated by applying PD subjects' models to PD subjects' data and applying PD subjects' models to control subjects' data. This shows that the models fit to PD subjects' data are typically less capable of fitting any subject's data, regardless of their cohort. This result is not unreasonable since it is already well documented that PD patients have different gaits when compared to otherwise healthy individuals [3] and the physical manifestations of symptoms (tremors, slower movements, rigid, poor posture) would impact the recorded data. In the end, the most unreliable models were those fit to PD subjects' data, which we would expect to be inconsistent.

Further, consider the last pair of distributions comparing PD on PD and Control on PD. The median MAE value is *lower* for the models fit to control subjects when applied to PD subject data relative to models actually fit to PD subjects. Although control models applied to control data had better MAEs when compared to the control models applied to the PD data (third distribution), the models fit to the more consistent control data were better in general, even when applied to inconsistent PD data. This further emphasizes the previous observation that the models fit to inconsistent data (PD subjects' data) are less effective regardless of which cohort the data came from. A similar observation can be made for models fit to PD subjects' data and applied to the control data — models obtain a lower error when given more consistent data.

In the linear models there is a statistical difference between the MAEs from each of the quadrants, however the actual MAEs obtained by the linear models are much worse than the nonlinear models. Perhaps the linear models are not as capable as the nonlinear models in picking up on important nuances in the data (Figure 2).

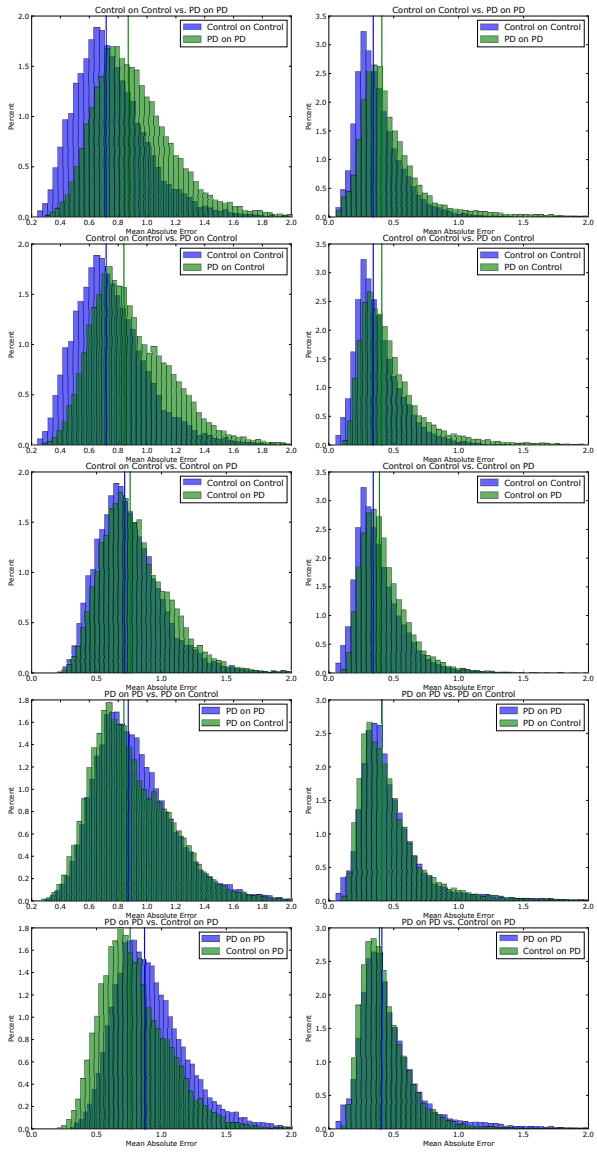

Fig. 5: Distributions of mean absolute error values obtained when applying groups of models to groups of data. Histograms on the left are from the OLS linear models and those on the right are from the nonlinear models. The distributions' median MAE values are solid vertical lines.

Given the p-values presented in Tables IV, V, and VI, we see that there is a difference between the Control and PD models in terms of their effectiveness on each others' cohorts. Although the differences were less visual than expected and had a small effect size, there is a statistically significant difference.

V. Conclusions and Future Work

Symbolic descriptive models of PD and control subjects' gait were generated with linear (OLS and LASSO) and nonlinear regression analysis (symbolic regression with GP). All models were capable of fitting the data well with low MAE values, however the nonlinear models generated with GP fit the data the models were fit to the best. Further, with a naive model selection strategy, the nonlinear models generalized to unseen data from the same subjects significantly better than the linear models. The authors acknowledge the advantage the nonlinear models have over the deterministically calculated linear models as GP can easily be used to generate a collection of high-quality and similarly effective nonlinear models to choose from. For this reason, no rigorous model selection was done for the nonlinear models.

Since the models are symbolic it was possible to analyze them. A relationship in the models that exists within the actual underlying system was noted, namely, only sensors from locations on subjects' feet that would be in contact with the ground at the same time were related. The authors expected to see consistency with respect to which sensors were within the models for the control cohort and PD cohort, however no consistency was observed.

Models were applied to all sets of data from all subjects and the MAEs were plotted in matrices to visualize relationships between the models and the data. These matrices were broken down into four quadrants representing when the models fit to the two cohorts (control and PD) were applied to data from the two cohorts. It was expected that there would be differences between the MAEs obtained by the models from the cohorts when applied to data from the different cohorts. Although no visual differences were observed, calculated p-values indicated that there were statistically significant differences. In the end, models fit to control subjects were capable of fitting data from other control subjects the best. Models fit to PD subjects' data did not generalize well to other PD subjects' data, which one should expect given PD symptoms and that PD patients have inconsistent gaits [3].

This long term project has many opportunities for future

TABLE IV: Mann-Whitney U test p-values comparing the distributions of MAEs from the four cohorts when using OLS linear models.

	Control on Control	PD on Control	Control on PD	PD on PD
Control on Control	0.5	0	$3.943*10^{-84}$	0
PD on Control	0	0.5	0	$3.902*10^{-24}$
Control on PD	$3.943*10^{-84}$	0	0.5	0
PD on PD	0	$3.902*10^{-24}$	0	0.5

TABLE V: Mann-Whitney U test p-values comparing the distributions of MAEs from the four quadrants when using LASSO linear models.

	Control on Control	PD on Control	Control on PD	PD on PD
Control on Control	0.5	$8.945*10^{-219}$	$1.017*10^{-23}$	0
PD on Control	$8.945*10^{-219}$	0.5	$3.797*10^{-131}$	$2.588*10^{-6}$
Control on PD	$1.017*10^{-23}$	$3.797*10^{-131}$	0.5	$3.155*10^{-206}$
PD on PD	0	$2.588*10^{-6}$	$3.155*10^{-206}$	0.5

TABLE VI: Mann-Whitney U test p-values comparing the distributions of MAEs from the four cohorts when using the Nonlinear models.

	Control on Control	PD on Control	Control on PD	PD on PD
Control on Control	0.5	0	$5.265*10^{-231}$	0
PD on Control	0	0.5	$1.959*10^{-32}$	$\mathbf{7.057}*10^{-2}$
Control on PD	$5.265*10^{-231}$	$1.959*10^{-32}$	0.5	$4.100*10^{-45}$
PD on PD	0	$\mathbf{7.057}*10^{-2}$	$4.100*10^{-45}$	0.5

work. Expanding the data used to include all 166 subjects from the database will allow for a much more significant comparison between the cohorts and will verify if our conclusions generalize further. Models could be fit to the whole time series from a subject or to data from multiple subjects at the same time in an effort to generate models more effective at intersubject generalization. Selecting a different sensor to regress to, such as L8, or focusing on data from only one foot might provide different insights into the system.

In this work the metadata was used to order the PD subjects' data based on the subject's UPDRS, however there is more metadata that could be incorporated into the analysis. In the Yogev *et al.* experiment [20], data was recorded from subjects as they walked while performing a serial subtract seven cognitive load and it is observed that PD patents have a higher difficulty performing motor tasks when doing such a cognitive task. This data could be modelled and compared to our current models to see how they differ.

The current study used data from eight sensors on each foot. Using information from such sensors, one can determine measures such as stride time and length. However, the sensors may have difficulty noticing information relating to other PD motor symptoms, for example, decreased arm swing, and as a result such symptoms are unlikely to be reflected in our models. It is possible that a smartwatch or other wearable sensors may provide a wider range of data, of which our models could take advantage. Furthermore, PD tends to progress asymmetrically [14]; given data on which side is most affected for a given patient, our models could see considerably greater accuracy as we could choose which sensors (left or right) to regress to based on that information. Our models would likely also become more accurate when given information as to which patients are undergoing treatment – for example, it is shown in [9] that there are significant differences in stability of stance for patients who have untreated PD or who have undergone different treatments.

As is mentioned in Section I, accurate diagnosis of PD should use multiple different types of information. It has been demonstrated in this paper that an accurate model of gait could provide an important piece of the puzzle. Many genes are believed to be involved in PD in some manner, with some such as PARK1 (SNCA) being known monogenic forms while others are suspected of some role in PD or may be risk factors. It is of interest to note that there are distinct subtypes of the postural instability/gait disturbance present in PD [2], with postural instability with falling (PIF) related to one gene and freezing of gait (FOG) to another. With continued successful generation of these symbolic descriptive models, a system based on these models could be used to aid in the diagnosis of PD and a real-time system used to monitor at risk individuals. Such a system could also be used to detect anomalies in an individual's gait to predict falls and prevent injuries.

VI. Acknowledgments

This research was supported by ACEnet (https://www.ace-net.ca/), Compute Ontario (www.computeontario.ca), WestGrid (www.westgrid.ca), Compute Canada (www.computecanada.ca), and the Natural Sciences and Engineering Research Council of Canada (NSERC).

The authors thank Dr. Matea Drljepan for her artwork.

References

[1] Gait in parkinson's disease. https://physionet.org/pn3/gaitpdb/. Accessed: September 10, 2018.

[2] Stewart A Factor, N Kyle Steenland, Donald S Higgins, Eric S Molho, Denise M Kay, Jennifer Montimurro, Ami R Rosen, Cyrus P Zabetian, and Haydeh Payami. Postural instability/gait disturbance in parkinson's disease has distinct subtypes: an exploratory analysis. *Journal of Neurology, Neurosurgery & Psychiatry*, 82(5):564–568, 2011.

[3] Leslie J Findley. The economic impact of parkinson's disease. *Parkinsonism & related disorders*, 13:S8–S12, 2007.

[4] Silvi Frenkel-Toledo, Nir Giladi, Chava Peretz, Talia Herman, Leor Gruendlinger, and Jeffrey M Hausdorff. 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*, 2(1):23, 2005.

[5] Silvi Frenkel-Toledo, Nir Giladi, Chava Peretz, Talia Herman, Leor Gruendlinger, and Jeffrey M Hausdorff. Treadmill walking as an external pacemaker to improve gait rhythm and stability in parkinson's disease. *Movement disorders: official journal of the Movement Disorder Society*, 20(9):1109–1114, 2005.

[6] Thomas Gasser. Genetics of parkinson's disease. *Current opinion in neurology*, 18(4):363–369, 2005.

[7] Jeffrey M Hausdorff, Justine Lowenthal, Talia Herman, Leor Gruendlinger, Chava Peretz, and Nir Giladi. Rhythmic auditory stimulation modulates gait variability in parkinson's disease. *European Journal of Neuroscience*, 26(8):2369–2375, 2007.

[8] Ashkan Entezari Heravi, Koosha Tahmasebipour, and Sheridan Houghten. Evolutionary computation for disease gene association. In *Computational Intelligence in Bioinformatics and Computational Biology (CIBCB), 2015 IEEE Conference on*, pages 1–8. IEEE, 2015.

[9] Fay B Horak and Martina Mancini. Objective biomarkers of balance and gait for parkinson's disease using body-worn sensors. *Movement Disorders*, 28(11):1544–1551, 2013.

[10] James Alexander Hughes. jGP. https://github.com/jameshughes89/jGP, March 2015. Accessed: September 21, 2018.

[11] James Alexander Hughes, Joseph Alexander Brown, and Adil Mehmood Khan. Smartphone gait fingerprinting models via genetic programming. In *2016 IEEE Congress on Evolutionary Computation (CEC)*, pages 408–415. IEEE, 2016.

[12] James Alexander Hughes, Joseph Alexander Brown, Adil Mehmood Khan, Asad Masood Khattak, and Mark Daley. Analysis of symbolic models of biometric data and their use for action and user identification. In *2018 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology*. IEEE, 2018.

[13] Christine Klein and Ana Westenberger. Genetics of parkinsons disease. *Cold Spring Harbor perspectives in medicine*, 2(1):a008888, 2012.

[14] João Massano and Kailash P Bhatia. Clinical approach to parkinson's disease: features, diagnosis, and principles of management. *Cold Spring Harbor perspectives in medicine*, 2(6):a008870, 2012.

[15] Diane B Miller and James P O'Callaghan. Biomarkers of parkinson's disease: present and future. *Metabolism*, 64(3):S40–S46, 2015.

[16] Robert L Nussbaum and Christopher E Ellis. Alzheimer's disease and parkinson's disease. *New england journal of medicine*, 348(14):1356–1364, 2003.

[17] Meir Plotnik, Nir Giladi, and Jeffrey M Hausdorff. A new measure for quantifying the bilateral coordination of human gait: effects of aging and parkinsons disease. *Experimental brain research*, 181(4):561–570, 2007.

[18] Michael Schmidt and Hod Lipson. Comparison of tree and graph encodings as function of problem complexity. In *Proceedings of the 9th annual conference on Genetic and evolutionary computation*, pages 1674–1679. ACM, 2007.

[19] Michael D Schmidt and Hod Lipson. Coevolution of fitness predictors. *IEEE Transactions on Evolutionary Computation*, 12(6):736–749, 2008.

[20] Galit Yogev, Nir Giladi, Chava Peretz, Shmuel Springer, Ely S Simon, and Jeffrey M Hausdorff. Dual tasking, gait rhythmicity, and parkinson's disease: which aspects of gait are attention demanding? *European journal of neuroscience*, 22(5):1248–1256, 2005.