Characterisation of Movement Disorder in Parkinson's Disease using Evolutionary Algorithms

Stuart E. Lacy
sl561@york.ac.uk

Michael A. Lones
michael.lones@york.ac.uk

Stephen L. Smith
stephen.smith.@york.ac.uk

Department of Electronics
University of York, Heslington
York YO10 5DD, UK

ABSTRACT

Parkinson's Disease is a devastating illness with no currently available cure. As the population ages, the disease becomes more common with a large financial cost to society. A rapid and accurate diagnosis, as well as practical monitoring methods are essential for managing the disease as best as possible. This paper discusses two approaches to discriminating movement data between healthy controls or Parkinson's Disease patients. One is a standard statistical analysis, influenced by prior work into classifying patients. The other is a programmatic expression evolved using genetic programming, which is trained to observe differences in specific motion segments, rather than using arbitrary windows of a full data series. The performance of the statistical analysis method is relatively high, but it still cannot discriminate as well as the evolved classifier. This study compares favourably to previous work, highlighting the usefulness of analysing a successful classifier to influence design decisions for future work. Examination of the evolved programmatic expressions that had high discriminatory ability provided useful insight into how Parkinson's Disease patients and healthy subjects have differing movement characteristics. This could be used to inform future research into the physiology of repetitive motions in Parkinson's Disease patients.

Categories and Subject Descriptors

I.2.1 [**Artificial Intelligence**]: Applications and Expert Systems—*Medicine and science*; I.2.8 [**Artificial Intelligence**]: Problem Solving, Control Methods, and Search; J.3 [**Computer Applications**]: Life and Medical Sciences

General Terms

Experimentation

Keywords

Classification, Genetic Programming, Parkinson's Disease

GECCO'13 Companion, July 6–10, 2013, Amsterdam, The Netherlands.

1. INTRODUCTION

Parkinson's Disease is a debilitating neurodegenerative disease, currently estimated to effect as many as 1 in 500 in the UK [1]. While the cause of the disease is yet to be known, it manifests itself by the gradual deterioration of dopamine carrying neurons in an affected person's brain. This has the effect of causing severe movement disorders, such as bradykinesia, rigidity, and tremor. There is no cure for Parkinson's Disease, only treatment to help deal with the symptoms. However, the main medication, Levodopa, has a side effect whereby patients who have been taking the drug long term start to display movement disorders, termed "Levodopa induced dyskinesias" [6, 10]. Managing the trade off between these and Parkinsonian symptoms can become very challenging and time consuming. As a result, an early and accurate diagnosis is key to the successful monitoring of the disease. Unfortunately, it has been estimated that up to 25% of early Parkinson's Disease diagnoses by primary care practitioners are inaccurate [5], due to the common similarity in the presentation of symptoms to other neurodegenerative diseases or physical conditions. Thus, an objective and accurate system to diagnose and monitor the disease would be extremely beneficial.

We have previously shown that it is possible to distinguish Parkinson's Disease patients from healthy controls by using a variant of genetic programming (GP) termed Cartesian Genetic Programming (CGP), from data obtained from a simple standard clinical exam, finger tapping [8, 7, 9]. This work highlighted several key features of movement present in Parkinson's Disease patients that can help to differentiate it from healthy motion, in particular the double peak of acceleration; and closing tap velocity. This paper details two contrasting approaches to use this knowledge to improve classifier accuracy. The two methods were a standard statistical analysis, and a second genetic programming based classifier, this time using segmented tap data.

2. METHODS

2.1 Patients

49 patients were assessed at Leeds General Infirmary (UK) between August 2009 and October 2010. They represented varying stages of the disease, (median Hoehn and Yahr scale score of 2.5). 41 age-matched healthy controls were assessed during the same time frame, all with no prior clinical diagnosis of a neurodegenerative condition. The gender ratio (male:female) for the Parkinson's Disease patients was

32:17, representing the higher incidence rate amongst men, and 14:27 for the control group. The mean ages of the two groups were 67 (±9) and 64 (±10) for the patient and control group respectively. All patients were assessed while in the ON stage of Levodopa treatment.

2.2 Test

The data collected for this study was undertaken as one part of the Unified Parkinson's Disease Rating Scale (UPDRS), the current gold standard for staging Parkinson's Disease [3]. It consists of five sections, four of which are multiple choice questions while the remaining part is a motor exam scored by a trained clinician. During the application of the motor exam 18 different activities are assessed, one of which is finger tapping. For this the patient is instructed to extend their index finger and thumb on their dominant hand, and repeatedly bring them together in a tapping motion, using movements as fast and with as large amplitude as possible. The test finishes after 10 taps have been completed. This is then repeated for the non-dominant hand. This is typically assessed, along with the other sections of the UPDRS, by a trained clinician. The patient then receives a single score from 0 to 4. The patients were also assessed on the UPDRS finger tapping scale during data collection, with the median score being 1.

However, there are several issues with this process, which is partly why we have decided to investigate the finger tapping test and see whether assessment can be improved. Firstly, having the test be assessed by a single person introduces an element of subjectivity to the marking procedure. This is crucially important, as the aim of finger tapping is to gain an understanding of one of the cardinal features of Parkinson's Disease, bradykinesia. This is a commonly misused term for several closely related movement disorders;

- Akinesia - hesitations in initiating movement
- Hypokinesia - reduced amplitude of movement
- 'True' Bradykinesia - slowness of movement

The single score of the finger tapping test is intended to take all of these factors into account, however it has been shown that clinicians have a tendency to assign disproportionately greater weight to amplitude than other measures [4]. This is a concern, as amplitude has been shown to be one of the variables of movement least affected by Parkinsonian medication [2]. As such, an objective analysis of the finger tapping exam would allow for a more accurate understanding of exactly what constitutes bradykinesia and how it manifests itself in simple repetitive motor tasks.

2.3 Equipment

A Polhemus Patriot position sensor was used to obtain real time information about the subject's movements. The Patriot consists of an electromagnetic source, with two small lightweight sensor units. These receivers were attached to the subject's index finger and thumb while they performed the finger tapping test. The sensors provided the position of the digit at a sampling frequency of 60Hz and with 6 degrees of freedom, while being unobtrusive for the subject.

2.4 Data Processing

The raw coordinate positional values were obtained from the two sensors and passed through a Butterworth Low Pass Filter with cut-off frequency 5Hz. The difference between the two sensors was taken to provide values relative to the thumb. Then, the Euclidean distance (Equation 1) was calculated at each time step to be used as a measure of separation between the digits.

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

To help reduce the effects of varying hand sizes, the separation values were normalised by dividing by the maximum separation for each test. This produced a continuous, normalised, view of the subject's motion throughout the 30 second test period. Taps were defined as intervals between two consecutive local minima. Figure 1 shows a trace of separation values, sampled over a 3 second period, with taps described as the period of separation between two successive minima. The maxima of the waveform indicate the index finger and thumb are at maximum distance from each other, and touching when the separation is at a minimum. The taps can be further sub-divided into the opening and closing phases of the movement. The 'max sep' label demonstrates how the value of maximum separation was derived.

2.5 Statistical Analysis

Various measurement variables that could prove useful in distinguishing the performance of healthy controls and Parkinson's Disease patients were derived, with knowledge gained from analysing the earlier automated classification system. These consisted of separation, speed (the magnitude of velocity was calculated as we were not concerned with direction) and acceleration metrics. Nearly all of the variables consisted of a single measure taken for each tap, then an average taken to generate a single value for each test. While in most cases this average was the mean, the coefficient of variation was also used. The calculated variables are as follows:

- Mean tap separation: The maximum separation for each tap was taken as that tap's separation, with the mean over the whole test then taken to obtain a single value for each test.
- Separation fatigue: The ratio of the last 5 tap separations to the first 5 (representing decrementing amplitude over the test duration).
- Separation rhythm: To get a measure for the rhythm of the tapping separation, the coefficient of variation was calculated over the tap separation values.
- Mean speed: The mean speed value from the entire trace.
- Mean opening speed: The opening speed was defined as the maximum speed during the first half of the tap motion. The mean was then calculated to obtain a single output for each test.
- Mean closing speed: The closing speed was defined as the maximum speed during the second half of the tap motion. The mean was then calculated to obtain a single output for each test.
- Mean opening acceleration: The peak of positive acceleration during the first half of the tap motion. The mean was then calculated to obtain a single output for each test.

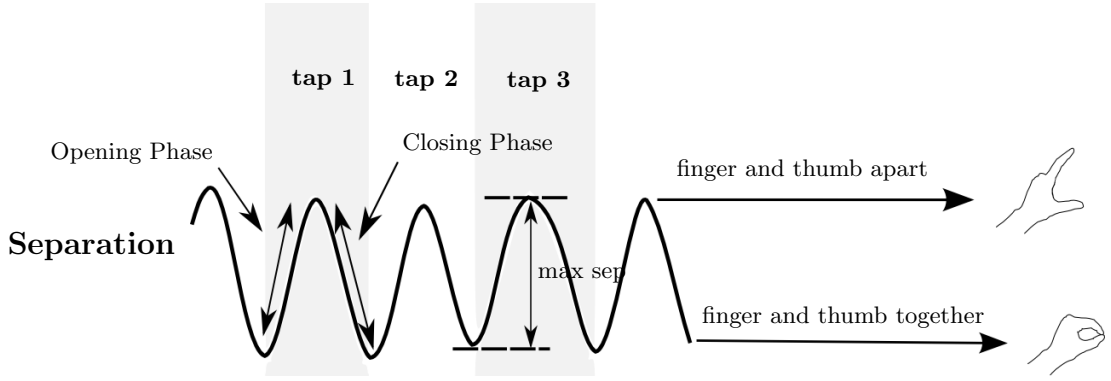

Figure 1: Obtaining the tap separation values.

- Mean opening deceleration: The peak of negative acceleration during the first half of the tap motion. The mean was then calculated to obtain a single output for each test.
- Mean closing acceleration: The peak of positive acceleration during the second half of the tap motion. The mean was then calculated to obtain a single output for each test.
- Mean closing deceleration: The peak of negative acceleration during the second half of the tap motion. The mean was then calculated to obtain a single output for each test.

Figure 2 shows how these measures were obtained.

2.6 Classification

2.6.1 Genetic Program

The classifier was a standard syntax tree representation of a programmatic expression evolved using genetic programming. Functions available were standard arithmetic operations $(+, -, *, MAX)$. The terminal set comprised of two groups: constants and inputs. Constants were floating point numbers in a pre-determined range (1.0, 10.0). The input set consisted of floating point values in the half open interval $[0, 1)$, representing a fraction of the current input tap sequence. The measure of the data used for the input tap sequence was the normalised separation, from only the dominant hands, as these had previously shown to be most discriminatory [8]. The programs were evolved over 100 generations, with a population of 200. Standard subtree crossover and point mutation (with mutation rate of 4%) were used to guide the evolutionary process, with parents being selected by tournaments of size 4. Elitism of size 1 was employed to ensure the fittest individual from each generation was brought forward to the subsequent one. To obtain statistical significance 50 runs of the program were completed.

2.6.2 Evaluation

Receiver Operating Characteristics (ROC) graphs were used to assess an individual's ability to discriminate between the two classes, with the Area under the ROC Curve (AUC) used as the fitness function. The data was split into three groups: the training set; the validation set; and a test set in the ratio 4:1:1 respectively. The individuals in the population had their fitness evaluated on the training set at each generation, with the fittest individuals being those most likely to influence the next generation. At each generation, the fittest individual had their fitness tested on the validation set to obtain an idea of how well generalised the classification was. The fittest individual kept at the end of the run was the best performing individual on the validation set. After all runs had been completed, the individual that performed best on the validation set, across all runs, was re-evaluated on the test set to obtain an unbiased measure of generality. Finally, the best performing classifier also had its ability tested all the data together, to provide a measure of accuracy when confronted with a large data set.

Table 1: Data set splits

	All Data	Training	Validation	Test
Controls	37	25	6	6
Patients	45	31	7	7

3. RESULTS

3.1 Statistical Measures

The performance of the standard statistical measures in discriminating between tests taken by controls and Parkinson's Disease patients can be seen in Table 2. While certain variables are particularly strong, notably mean separation, the closing acceleration values, and both the opening and closing speed, others, such as fatigue perform less well. This indicates that the classifiers still can observe differences missed by our statistical approach. This demonstrates that for complex classification tasks, where the underlying mechanisms which distinguish the two groups are not well understood, but where there is a lot of available data, an evolutionary algorithm approach has considerable potential.

3.2 Classifier Results

The results of the genetic program runs were promising, many classifiers were produced with AUCs of over 0.9. The classifier taken as the best of the runs, was the genetic program whose syntax tree produced the highest accuracy on the validation set. The structure of this individual is seen in Figure 3. The floating point values represent fractions of the input tap data. The values are mostly less than 0.5, indicating that this particular classifier was generally look-

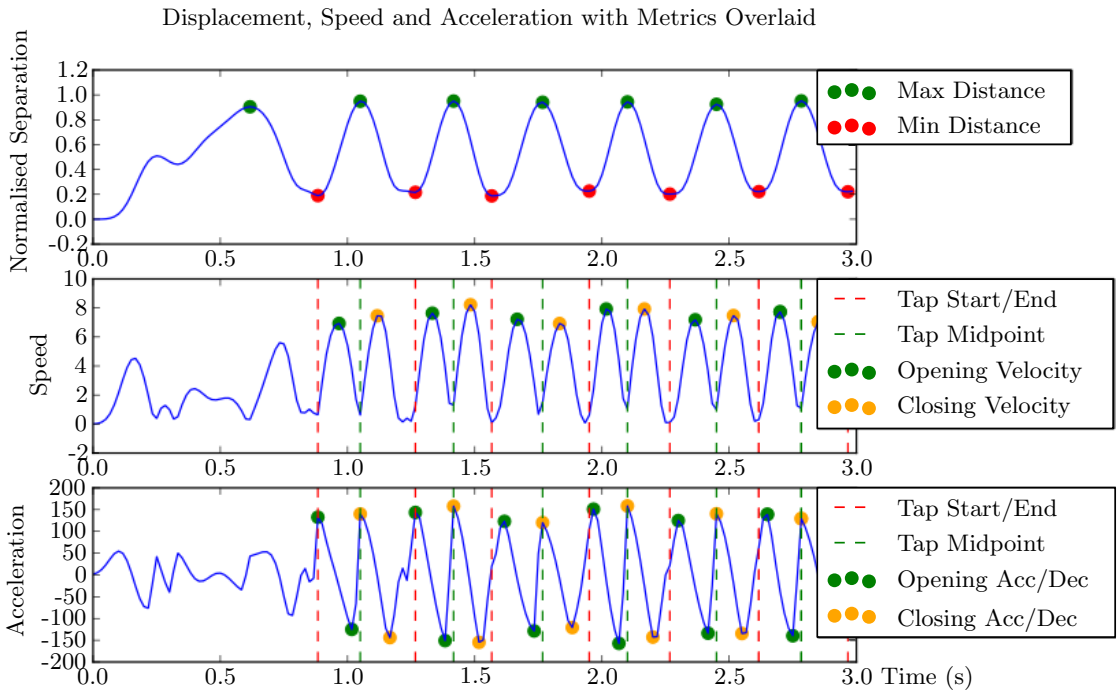

Figure 2: Calculating variables from separation, speed, and acceleration.

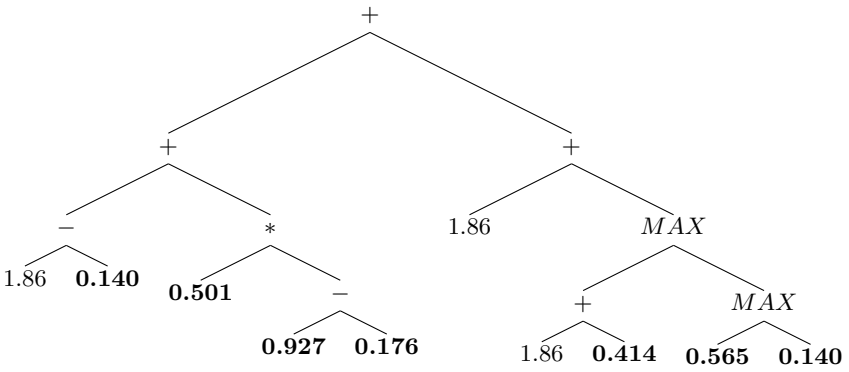

Figure 3: Syntax tree of the fittest individual over all the runs.

Table 2: Discriminatory ability of statistical variables

Measure	AUC
Mean Separation	0.836
Fatigue Separation	0.616
Separation Rhythm	0.836
Mean Speed	0.780
Mean Opening Speed	0.838
Mean Closing Speed	0.848
Mean Opening Acceleration	0.798
Mean Opening Deceleration	0.775
Mean Closing Acceleration	0.819
Mean Closing Deceleration	0.839

Table 3: Classifier performance on the data sets.

Set	AUC
Training	0.881
Validation	1.000
Test	0.976
All	0.904

ing at the first half of the tap, at values associated with the opening of the index finger and thumb.

Table 3 presents the results of this evolved classifier on all of the four data sets. Due to being selected for its validation set AUC, it performs well on this set. Its strong classification ability on the test set data further highlights this classifier's ability to generalise to many different data sets. While it does perform well on the training set, the AUC is not as high as it might be expected. Generally, classifiers perform better on the training set than on the test set, the opposite of this individual. One explanation for this behaviour could be the small data set sizes used to assess the classifier. As seen in Table 1, there are only 6 healthy controls and 7 patients in the validation and test sets. It is possible that these 7 patients happened to be in the later stages of Parkinson's Disease and thus displayed more of the characteristic traits that the classifier can pick up on. Likewise this could mean that the data sets from patients in the training set were mostly from patients who did not display many symptoms and so were harder to classify. This would explain why AUC on all the data is lower than for the validation and test sets as well, as the training set data makes up the majority of all the data. This is a plausible explanation as the median UPDRS Finger Tapping score was 1 out of 4, indicating a majority of the Parkinson's Disease patients completed the task with only minor impairment. Another factor that could affect the results, is the fact that the patients were measured while on medication, which could be assumed to dampen the bradykinesia characteristics that the classifier locates.

The Receiver Operating Characteristics graphs are shown in Figure 4. They visually demonstrate the differing performance of this classifier on the various data sets.

3.3 Comparison to Previous Classifier

The AUC scores obtained compare favourably to those calculated by the previous evolved classifier (Table 4). This demonstrates that using tap separation values as the inputs to the genetic program results in a classifier that is at least as discriminatory, if not more so than using windows of data.

The GP classifier performs slightly worse on the data set

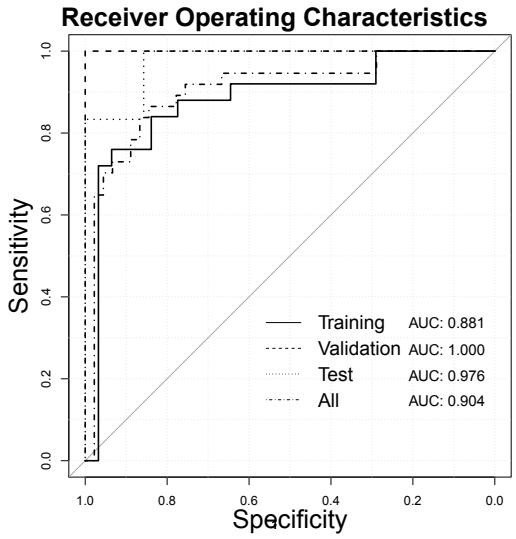

Figure 4: ROC plots.

Table 4: Previously obtained AUC scores

Set	AUC
Training	0.912
Validation	0.911
Test	1.000
All	0.918

it has been trained on than the previous CGP classifier. Possible explanations for this have already been given, with the main concern being the size of the data sets. The earlier classification work used data sets of the same size as those used in this study. Over all the data, the two classifiers achieve similar discriminatory ability, indicating they both generalise well. The validation and test set scores, however, demonstrate the inherent problem with using relatively small data sets. While the new GP classifier has an AUC of 1.0 on the validation set, it has a lower AUC for the test set. This is understandable as the individual was selected by means of its validation set score so it is natural that it performed well on it and less well on the test set. However, the prior CGP classifier had almost the same scores but for the opposite data sets. This is likely due to having data from only 15 subjects in the validation set and 14 in the test set. As previously mentioned, this could result in the Parkinson's Disease patients represented in the validation set being at an early stage of the disease, and are thus harder to distinguish from healthy controls. And the opposite would explain the perfect discrimination of the test set.

3.4 Classifier Patterns

To gain further insight into the underlying physiological differences between the motions of Parkinson's Disease patients and healthy subjects, the taps that were strongly classified as belonging to either group were analysed. The 100 taps with the highest, and lowest output values were selected from the validation set. These represented the taps that were classified most strongly. Figure 5 shows the separation values from these taps. The shape of the taps with the highest output value, indicating strongly healthy motions,

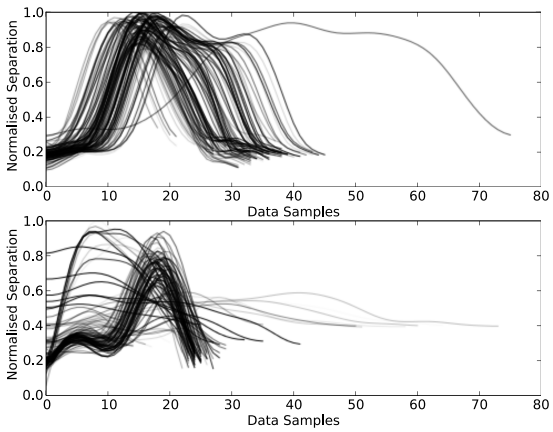

Figure 5: Plots of the 100 highest (top graph) and lowest (bottom graph) output values from the evolved mathematical expression. These represent the tap shapes most likely to be classed as from a healthy control and Parkinson's Disease patient respectively. A darker line indicates a stronger output value.

shows a single clearly defined peak, which is occasionally followed by another slightly smaller peak a short time later. This shape is evident in nearly all of the 100 taps, albeit at different time scales. The taps from the tests correctly classified as patients however are drastically different. While there are a number of taps with low maximum separation values and a very small negative gradient, the majority follow a consistent tap shape. The graph shows a small initial local minimum before the separation peak where the index finger and thumb are at maximum separation. This could perhaps be indicative of akinesia, whereby Parkinson's Disease patients hesitate when initiating movements.

4. CONCLUSIONS

This work presents a novel application of GP to investigate short term patterns of motion in PD patients. It has shown to be comparable to previous uses of GP to classify Parkinson's Disease, and classifies more accurately than statistical measures of movement. Analysis of the syntax trees from the highest performing GP classifier could help to gain an insight into what features of an individual tap motion help to distinguish Parkinson's Disease patients from healthy controls. This would further current understanding of bradykinesia and how it manifests itself in movement tasks. An initial analysis of the differing separation traces between Parkinson's Disease patients and healthy controls highlights a small separation peak before the main peak that is evident in the data from patients, but not from controls. While further analysis is needed, it appears that this could be due to the akinesia component of bradykinesia. Furthermore, healthy controls demonstrate a second local maximum of separation, shortly after the main peak. This could prove to be a useful indicator of normal motor control and the subject of further investigation.

Since this preliminary study has proved the usefulness of having tap sequences input into an evolved classifier, there are several ways to expand on this. A second classifier could be evolved in parallel, but rather than classifying based on individual tap sequences, it could instead search for characteristic patterns at a more global level. This could involve representing the input data as a sequence of values, with one measure per tap. The variable used to represent each tap could be something as simple as the mean tap separation or a more complex measure. This could be guided by an analysis of how this current classifier differentiates data from healthy control subjects to Parkinson's Disease patients. As with the previous work into using evolved classifiers to classify finger tapping data, it might be the case that acceleration sequences could provide more useful than separation values.

The results analysis brought up the issue of small data sets distorting results. To counter this it would be preferable to obtain more data to increase the sample sizes. Furthermore, when collecting more data, it would be extremely useful to assess Parkinson's Disease patients not exhibiting signs of medication (termed the 'OFF' stage). It could be assumed that these patients would present more noticeable signs of bradykinesia for a classifier to locate. A final way to improve the robustness of the results would be to test subjects at multiple centres in different locations.

5. ADDITIONAL AUTHORS

Additional authors:
Jane E. Alty (Leeds General Infirmary UK,
email: `jane.alty@hyms.ac.uk`),
D. R. Stuart Jamieson (Leeds General Infirmary UK,
email: `stuart.jamieson@leedsth.nhs.uk`),
Katherine L. Possin (Memory and Aging Center, University of California San Francisco,
email: `katherine.possin@ucsf.edu`) and
Norbert Schuff (Centre of Imaging of Neurodegenerative Diseases,
email: `norbert.schuff@va.gov`).

6. REFERENCES

[1] Website of the parkinsons's disease society. Available at http://www.parkinsons.org.uk, 2013.

[2] A. J. Espay, J. P. Giuffrida, R. Chen, M. Payne, F. Mazzella, E. Dunn, J. E. Vaughan, A. P. Duker, A. Sahay, S. J. Kim, et al. Differential response of speed, amplitude, and rhythm to dopaminergic medications in parkinson's disease. *Movement Disorders*, 26(14):2504–2508, 2011.

[3] C. G. Goetz, B. C. Tilley, S. R. Shaftman, G. T. Stebbins, S. Fahn, P. Martinez-Martin, W. Poewe, C. Sampaio, M. B. Stern, R. Dodel, et al. Movement disorder society-sponsored revision of the unified parkinson's disease rating scale (mds-updrs): Scale presentation and clinimetric testing results. *Movement Disorders*, 23(15):2129–2170, 2008.

[4] D. A. Heldman, J. P. Giuffrida, R. Chen, M. Payne, F. Mazzella, A. P. Duker, A. Sahay, S. J. Kim, F. J. Revilla, and A. J. Espay. The modified bradykinesia rating scale for parkinson's disease: reliability and comparison with kinematic measures. *Movement Disorders*, 26(10):1859–1863, 2011.

[5] J. Jankovic, A. H. Rajput, M. P. McDermott, D. P. Perl, et al. The evolution of diagnosis in early

parkinson disease. *Archives of neurology*, 57(3):369, 2000.

[6] N. L. Keijsers, M. W. Horstink, and S. C. Gielen. Automatic assessment of levodopa-induced dyskinesias in daily life by neural networks. *Movement disorders*, 18(1):70–80, 2003.

[7] M. A. Lones, J. E. Alty, S. Lacy, S. Jamieson, K. Possin, N. Schuff, and S. L. Smith. Evolving classifiers to inform clinical assessment of parkinson's disease. Paper presented at 2013 IEEE Symposium Series on Computational Intelligence (SSCI 2013), Singapore, Singapore, 2013.

[8] M. A. Lones, S. L. Smith, J. E. Alty, S. E. Lacy, K. L. Possin, D. R. S. Jamieson, and A. M. Tyrrell. Evolving classifiers to recognise the movement characteristics of parkinson's disease patients. *IEEE Transactions on Evolutionary Computation*, Accepted for publication, 2013.

[9] M. A. Lones, S. L. Smith, A. M. Tyrrell, J. E. Alty, and D. Jamieson. Characterising neurological time series data using biologically motivated networks of coupled discrete maps. *Biosystems*, 2013.

[10] A. Manson, P. Brown, J. O'sullivan, P. Asselman, D. Buckwell, and A. Lees. An ambulatory dyskinesia monitor. *Journal of Neurology, Neurosurgery & Psychiatry*, 68(2):196–201, 2000.