Classification and Characterisation of Movement Patterns During Levodopa Therapy for Parkinson's Disease

Michael A. Lones*
Dept. of Computer Science
Heriot-Watt University
Edinburgh, UK

Jane E. Alty
Department of Neurology
Leeds General Infirmary
Leeds, UK

Phillipa Duggan-Carter
Department of Neurology
Leeds General Infirmary
Leeds, UK

Andrew J. Turner
Department of Electronics
University of York
York, UK

D. R. Stuart Jamieson
Department of Neurology
Leeds General Infirmary
Leeds, UK

Stephen L. Smith
Department of Electronics
University of York
York, UK

ABSTRACT

Parkinson's disease is a chronic neurodegenerative condition that manifests clinically with various movement disorders. These are often treated with the dopamine-replacement drug levodopa. However, the dosage of levodopa must be kept as low as possible in order to avoid the drug's side effects, such as the involuntary, and often violent, muscle spasms called dyskinesia, or levodopa-induced dyskinesia. In this paper, we investigate the use of genetic programming for training classifiers that can monitor the effectiveness of levodopa therapy. In particular, we evolve classifiers that can recognise tremor and dyskinesia, movement states that are indicative of insufficient or excessive doses of levodopa, respectively. The evolved classifiers achieve clinically useful rates of discrimination, with AUC>0.9. We also find that temporal classifiers generally out-perform spectral classifiers. By using classifiers that respond to low-level features of the data, we identify the conserved patterns of movement that are used as a basis for classification, showing how this approach can be used to characterise as well as classify abnormal movement.

Categories and Subject Descriptors

I.2.8 [**Artificial Intelligence**]: Problem Solving, Control Methods, and Search; I.5.2 [**Pattern Recognition**]: Design Methodology; I.5.4 [**Pattern Recognition**]: Applications—*Signal processing*; J.3 [**Computer Applications**]: Life and Medical Sciences

General Terms

Experimentation

*Corresponding author. Email: M.Lones@hw.ac.uk.

GECCO'14, July 12–16, 2014, Vancouver, BC, Canada.

ACM 978-1-4503-2881-4/14/07 ...$15.00.
http://dx.doi.org/10.1145/2598394.2609852.

Keywords

Classification, Genetic programming, Parkinson's disease, Time series analysis, Fourier analysis, Pattern discovery

1. INTRODUCTION

In previous work, we have focused on evolving classifiers to diagnose Parkinson's disease (PD) and other neurological conditions [12, 11, 13]. Diagnosis is, however, just the first stage of clinical management. Following diagnosis, a patient's condition needs to be carefully monitored in order to choose an appropriate medication regimen. Since many of the movement disorders associated with PD, such as tremor and bradykinesia (slowed movements), are caused by the loss of dopamine-producing neurones in the brain, patients are often prescribed dopamine replacements such as the drug levodopa. However, doses of these drugs must be carefully monitored in order to maintain a balance between treatment of the patient's symptoms and avoidance of side-effects. One of the most debilitating side-effects is levodopa-induced dyskinesia (LID), which causes involuntary spasms of the limbs, trunk, neck and face. Once LID becomes established, it is difficult to treat. A common approach to managing its occurrence is to reduce the dose of levodopa just prior to periods of severe LID, but this is difficult to monitor, as the presence and severity of LID tends to fluctuate throughout the day. Hence, there is a need for accurate measurements of the patient's symptoms, and a means to perform these measurements frequently, non-invasively, and inexpensively.

Ideally, monitoring of patients should take place during the course of their normal day-to-day activities, and recently there has been growing interest in the use of lightweight wearable devices for carrying out monitoring in this fashion. The idea is that the patient wears these devices at appropriate locations around their body. The data from the accelerometers is then passed to a portable device, which periodically checks for the presence of movement disorders, and sends this information to a clinician who can then adapt the patient's medication as appropriate. However, dyskinesias can be difficult to distinguish both from normal movement and from other types of movement disorder. Hence, the development of computational techniques that can discriminate between different classes of movement, and differ-

ent severities within these classes, has become an important area of research.

In this paper, we describe the use of genetic programming for training classifiers that can identify the occurrence of tremor and LID within accelerometry data collected from wearable devices. We compare the performance of classifier models based on temporal and spectral representations of the data, and carry out analysis to identify the movement patterns used by evolved classifiers. Notably we find that classifiers trained directly on time series data are more effective than those trained on Fourier series coefficients for discriminating both tremor and LID.

2. RELATED WORK

Previous work on applying machine learning and data mining techniques to movement data collected from PD patients is reviewed in [1]. Particularly relevant are previous studies that have looked at predicting specific motor symptoms. This includes methods for measuring tremor whilst at rest [2], whilst performing specified movements [16], and during unscripted activities [15]. Several studies have also considered methods for identifying LID in movement data, whilst at rest [7], whilst performing specified movements [3], and during unscripted activities [9, 17]. In [6], the authors presented a method for detecting both tremor and dyskinesia during unscripted activities.

The work reported in this paper differs from earlier studies in our use of evolutionary algorithms and symbolic classifier models based upon low-level features of the data (e.g. raw acceleration values and Fourier coefficients), in comparison to the use of neural networks and higher-level features (such as signal energy and spectral powers over frequency ranges) in comparable studies [9, 6, 17, 15]. This is justified by our earlier work on PD diagnosis, which showed benefits in terms of interpretability and novelty discovery [12, 11], and which motivates our interest in characterising as well as classifying movement patterns. Characterisation is important, since the patterns of movement in PD are only partially understood. Better understanding of these movements may provide insight into their underlying causes, which in turn may contribute to the development of more effective treatments.

3. MATERIALS AND METHODS

3.1 Data Collection

We recruited six patients with confirmed PD for the clinical study. Lightweight devices containing integrated accelerometers and gyroscopes were fitted to the patients' legs, arms, torso, head and trunk, each able to record movement data in the three spatial and three rotational planes at a sample rate of 100Hz. An infrared camera was used to record video footage of the patients' movements. Each patient was recorded continuously for a period of 6 hours. Other than being situated in a hospital day case unit, the patients were unconstrained in their movements. Following the recording session, the video footage was analysed by two trained clinicians, who used the standard UPDRS (Unified Parkinson's Disease Rating Scale) and UDysRS (Unified Dyskinesia Rating Scale) scoring systems to mark up periods of tremor and LID within each body part with quantitative values (ranging from 0, indicating normal movements, to 4, indicating serious abnormality). Data sequences corresponding to these time periods were then extracted from the respective devices, resulting in sets of movement data corresponding to normal movement, and to tremor and LID at different clinical grades. Table 1 shows the number of data sequences in each movement class.

Tremor		Dyskinesia	
UPDRS	Count	UDysRS	Count
0	6661	0	2933
1	33	1	1227
2	61	2	1688
3	57	3	681
4	11	4	64

Table 1: Movement Data

3.2 Evolutionary Algorithm

We used implicit context representation Cartesian GP (IRCGP) to search for classifiers. IRCGP is a variant of CGP that uses a non-positional low level encoding to improve evolvability [4]. Prior to evaluation, IRCGP solutions are mapped into CGP solutions. Full details of IRCGP can be found in [11].

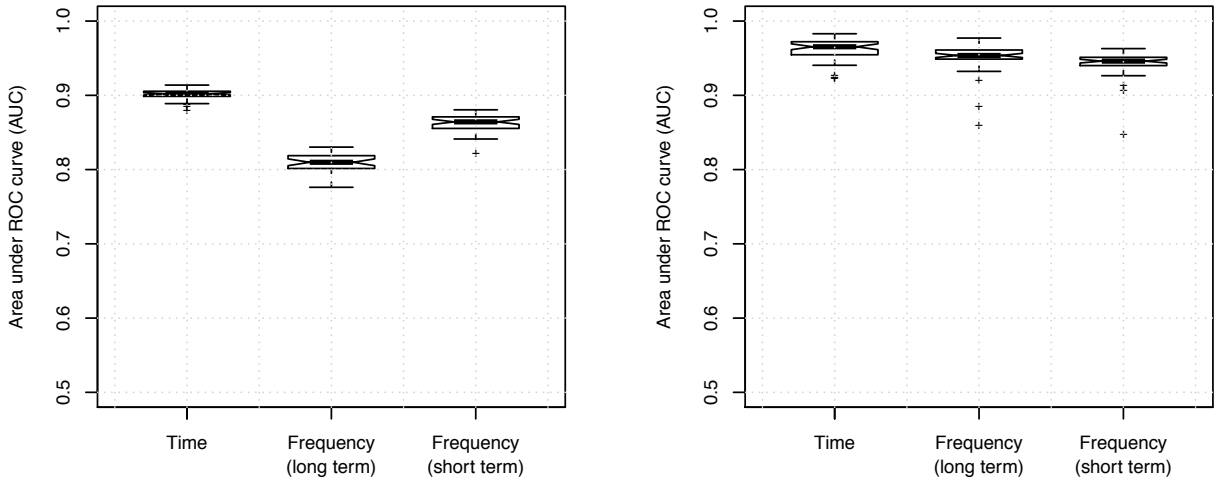
Evolutionary runs use a population size of 200 and a generation limit of 100. Point mutation is applied using a Gaussian distribution centred around the current value, with rates of 6% for functions and 3% for functionality profile elements (see [11]). Uniform crossover is applied with crossover points occurring with a probability of 15%.

3.3 Classifier Models

Three kinds of classifier were used: sliding window time series classifiers, long term spectral classifiers, and window-based short term spectral classifiers. In each case, classification is based on the evaluation of an evolved mathematical expression: in all cases these use up to 32 inputs, and contain up to 36 functions nodes ($\{+, -, \times, \div, \text{mean}, \text{min}, \text{max}, \text{mod}\}$) laid out on a 6×6 CGP grid. This uniformity eases comparisons, since the evolutionary algorithm is searching a space of equal size and dimensionality for each classifier model. The evolved expressions are used as follows:

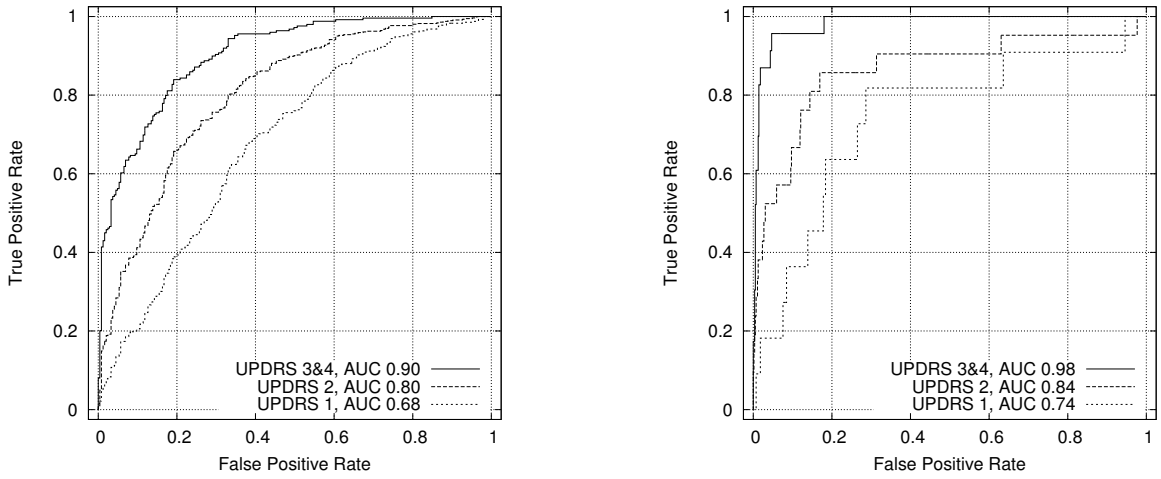
Time domain classifiers For each data point, a time series is created by calculating the magnitude of acceleration at each time index. The inputs to the expression represent the acceleration values in a contiguous time series window of length 32 ($0.32s$). This window is slid along the entire time series, generating an output for each of the $L - 31$ overlapping windows, where L is the length of the time series. The classification for the time series is then the mean of these values.

Long term spectral classifiers The inputs to the expression represent spectral densities at 32 equally spaced points in the frequency range 0–50Hz. These frequencies are taken from the Fourier transform of the acceleration time series. This is calculated using the method described in [8], which involves segmenting the time series into non-overlapping windows (of length 64 in this case), applying Fourier transforms to each window, and then averaging these to produce a robust estimate of spectral density across the whole time series.



(a) Dyskinesia (b) Tremor

Figure 1: Discriminative ability of classifiers evolved to recognise dyskinesia and tremor when evaluated on the independent test sets. Notched box plots show distributions of AUC over 50 runs. Non-overlapping notches indicate a strong likelihood of statistically significant differences between means.



(a) Dyskinesia (b) Tremor

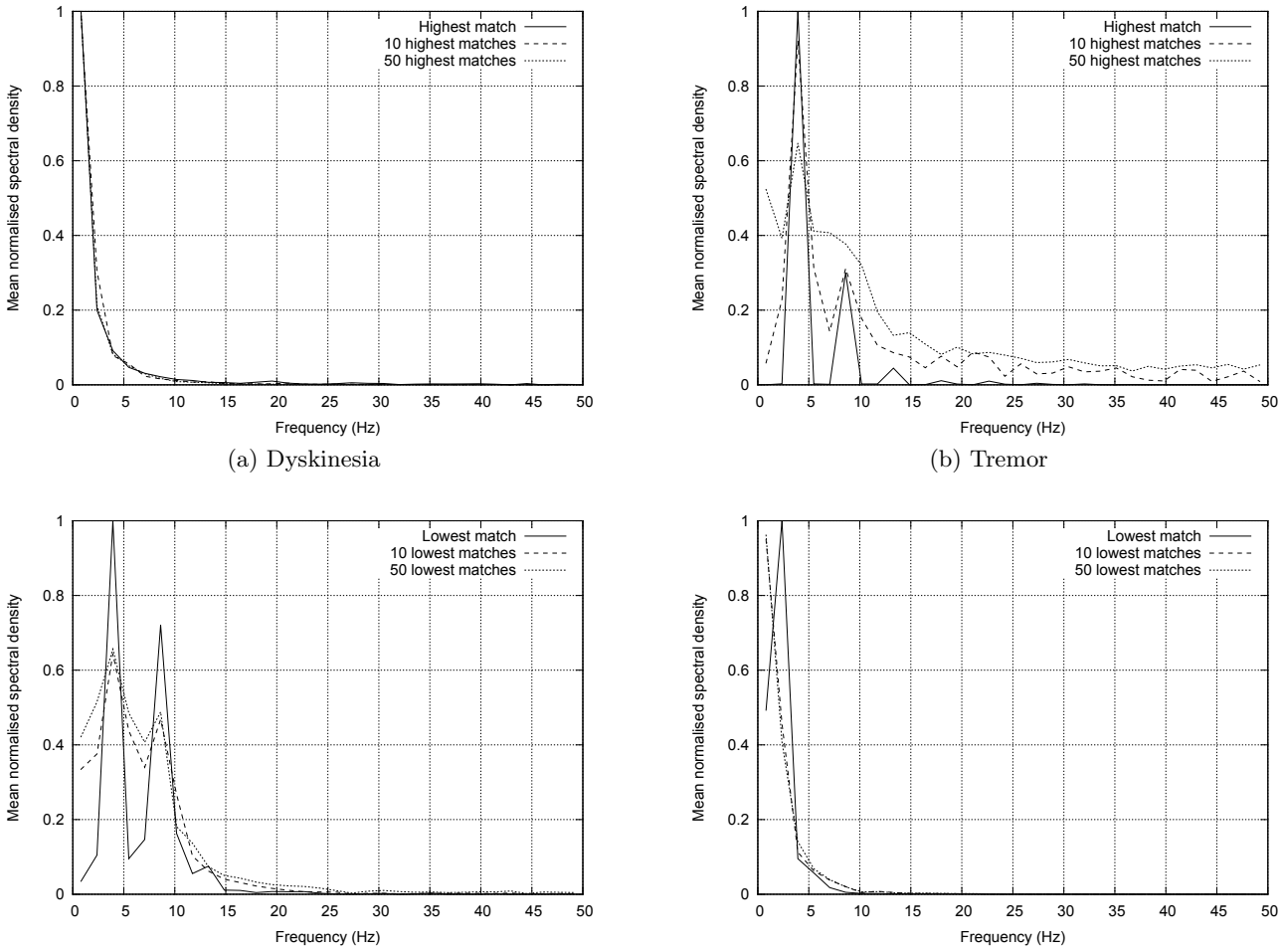
Figure 2: Representative ROC curves for classifiers evolved to recognise UPDRS level 3 and 4 dyskinesia and tremor when evaluated on the independent test sets for all UPDRS levels.

Short term spectral classifiers This is similar to the long term spectral classifier. However, the expression is applied independently to the Fourier transforms produced from each time series window, rather than the averaged Fourier transform. The classification for an acceleration time series is then the mean of these values.

In essence, these three classifier models operate on different views of the movement data, allowing them to recognise different kinds of pattern. The sliding window time domain classifier is designed to recognise conserved patterns of acceleration that occur during short periods of movement, the short term spectral classifier is similar, but operates in the frequency domain, and the long term spectral classifier is able to recognise stationary processes that occur over longer time periods.

3.4 Classifier Evaluation

The evolutionary algorithm is used to find diagnostic classifiers that have high predictive accuracy. This is done using an objective function that measures the area under the ROC curve (AUC) when separating classes, i.e. a classifier that maximises accuracy across all trade-offs between specificity and sensitivity. AUC is equivalent to the probability that a randomly chosen subject will be assigned to the correct class [10]. An AUC of 1 means that a classifier achieves 100% specificity and 100% sensitivity. An AUC of 0.5 indicates performance no better than random. Generally speaking, AUC>0.9 is considered excellent, and AUC>0.8 is considered good. An AUC less than 0.5 indicates the same predictive power as one with $1 - \text{AUC}$, but with a reversed ordering of the classes within its output range. To simplify



(a) Dyskinesia (b) Tremor

(c) Not dyskinesia (d) Not tremor

Figure 3: Frequency responses of spectral classifiers evolved to predict dyskinesia and tremor, showing (a-b) periodograms matched as positive, (c-d) periodograms matched as negative.

the presentation of results, all AUCs are normalised to the interval [0.5,1.0].

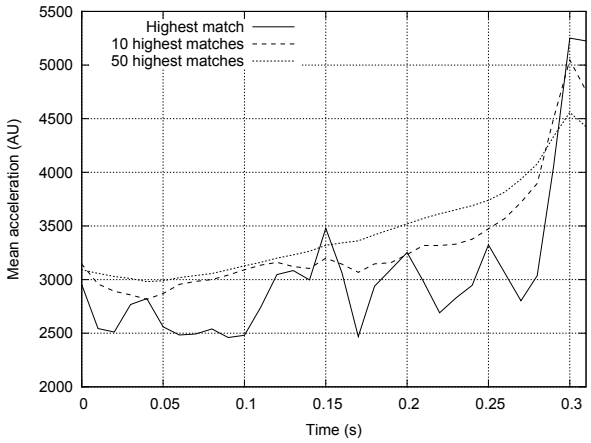
Data classes are divided equally into three sets: a training set, a validation set, and a test set. The training set is used by the objective function. The validation set is used for early stopping, and to select the final generation solution with the highest generality. The test set is then used to give an unbiased measure of predictive accuracy for the selected solution.

4. RESULTS

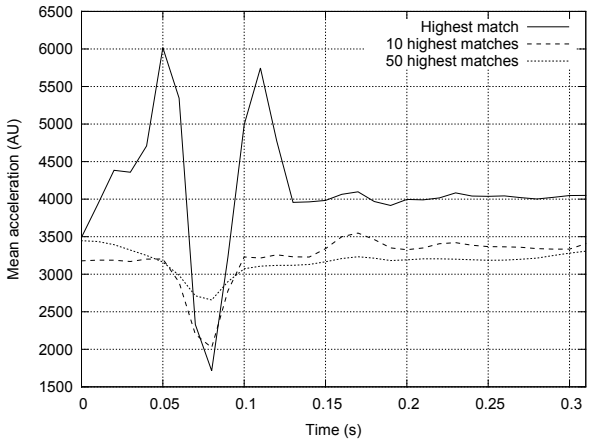
We evolved two sets of classifiers: those that can recognise dyskinesia, and those that can recognise tremor. In both cases, the objective was to discriminate grade 3 (moderate) and grade 4 (severe), treated as a single class, from grade 0 (normal). Data for grades 1 (slight) and 2 (mild) were not used during training in order to focus classification on clinically significant levels of abnormal movement, and to prevent training being affected by misclassified examples (which is more likely to happen at lower grades). Note that the grade 0 set for dyskinesia contains examples of tremor in addition to normal movement, and vice versa. Hence, the evolved classifiers are expected to discriminate their movement of interest from both normal movement and other abnormal movements.

To compensate for the stochasticity of evolutionary algorithms, we carried out 50 independent runs for each combination of symptom class and classifier model. The resulting distributions of predictive accuracy are shown in Fig. 1. This shows that we obtained AUCs of around 0.9 for the best dyskinesia classifiers, and 0.98 for the best tremor classifiers. Both of these represent levels of discrimination that are likely to be useful for monitoring purposes. Fig. 2 shows ROC curves for representative examples of evolved dyskinesia and tremor classifiers. In both cases, the recognition rates are roughly proportional to clinical grade, suggesting that classifier outputs may also be useful for providing objective measurements of symptom severity.

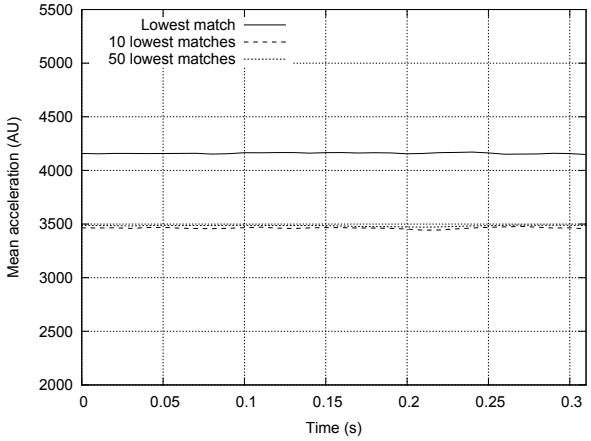
It is notable that the time series classifiers performed, on average, significantly better than the spectral classifiers. This is most pronounced for dyskinesia, but is also true for tremor. This is an interesting result because a lot of existing classification work on both these movement types has



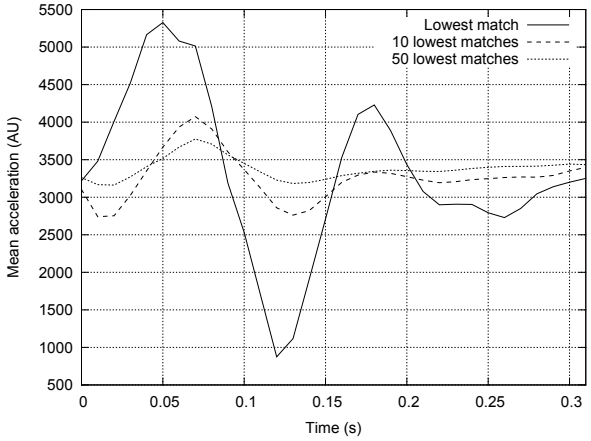
(a) Dyskinesia



(b) Dyskinesia



(c) Not dyskinesia



(d) Not dyskinesia

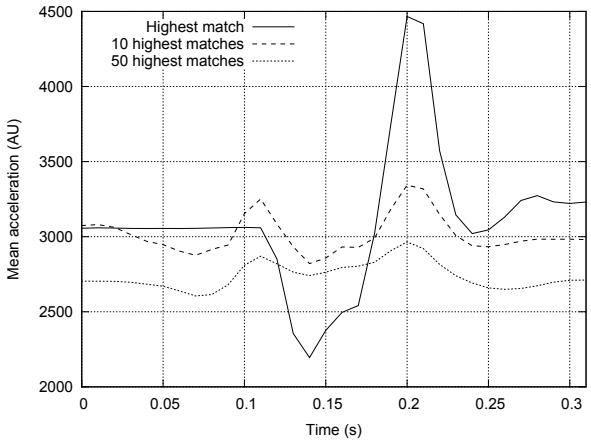
Figure 4: Acceleration patterns matched by time-domain classifiers evolved to predict dyskinesia, showing (a-b) time series windows matched as positive, (c-d) time series windows matched as negative.

focused on the frequency domain [1, 9, 14]. For tremor, a frequency analysis seems natural, since it is defined as a regular rhythmic oscillation of a body part, and the frequency ranges for different kinds of tremor are relatively well characterised. However, it is also possible to recognise oscillatory patterns in the time domain. In addition, time domain analysis makes it easier to characterise the shape of a waveform. This may underlie the performance of the time series classifiers, since the shape of the waveform may allow better separation of tremor and dyskinesia from other processes that occur within the same frequency range. For example, it has previously been observed that the frequencies associated with dyskinesia overlap with those of voluntary movements [9], especially walking.

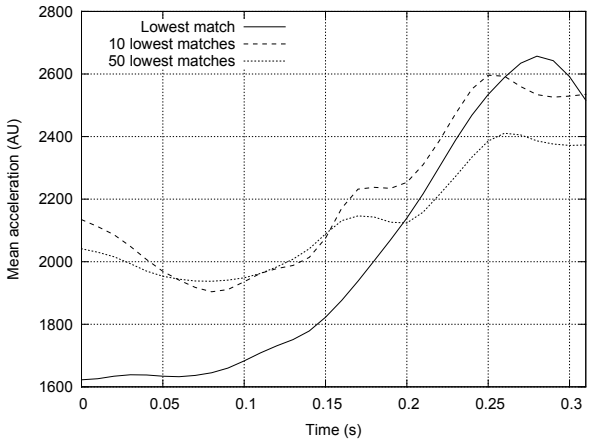
Also notable are differences between the discriminative ability of the two spectral classifier models, particularly the observation that short term classifiers recognise dyskinesia better than long term classifiers. This suggests that dyskinesia does not occur with a constant frequency during dyskinetic periods, which reflects previous observations that periods of dyskinetic movement are intrinsically complex [5]. This is in contrast to the slightly better performance that long term classifiers show for tremor recognition, which is consistent with current understanding of the stationarity of Parkinson's tremor.

4.1 Discriminative Patterns

A benefit of using raw temporal and spectral data as a basis for classification, rather than more complex features derived from these, is the potential for identifying discriminative patterns of motor movements. In particular, it is possible to look at the data windows that lead to a strong positive or negative response from a classifier in order to identify significant over-represented patterns of movement [12]. This has more value in the time domain than the frequency domain. However, for completeness, Fig. 3 shows the periodograms which lead to strong positive and negative responses from evolved tremor and dyskinesia classifiers. It can be seen that tremor classifiers favour movements with periodograms that are dominated by spikes of around 4Hz and 8Hz. This corresponds well with the known bands of rest and action tremors. Dyskinesia classifiers, by comparison, favour periodograms with the power spectrum concentrated in the low frequencies. Again, this corresponds well with ex-



(a) Tremor



(b) Not tremor

Figure 5: Acceleration patterns matched by time-domain classifiers evolved to predict tremor, showing (a) time series windows matched as positive, (b) time series windows matched as negative.

isting understanding of the frequency range of LID. It is also evident that the tremor and dyskinesia classifiers are symmetric, with the dyskinesia classifier most strongly rejecting periodograms that indicate tremor, and vice versa. This is likely to be a desirable behavior in practice, since these two movement classes are opposing indicators of levodopa dosage.

Fig. 4 shows a similar analysis from the time domain perspective, indicating over-represented patterns that occur within data windows classified strongly as dyskinesia or not dyskinesia. During the course of analysis, it was observed that classifiers evolved in different evolutionary runs often respond to different movement features in the data. For instance, the classifier whose matching behaviour is shown in Fig. 4a appears to respond strongly to the presence of a non-linear acceleration profile which starts flat and rises quickly, with a pronounced shoulder. Fig. 4b by comparison, suggests a classifier that responds to short periods of reduced acceleration during movements, with the highest matching patterns of acceleration resembling an M-shape. Some diversity is also seen in the negative responses from dyskinesia classifiers. However, Figs. 4c and 4d are representative of most. Fig. 4c suggests that constant acceleration is atypical of dyskinesia. Fig. 4d indicates the rejection of certain oscillatory patterns—most likely tremor, therefore suggesting a similar symmetric relationship to the ones seen in frequency domain classifiers. This is also supported by analysis of time domain classifiers evolved to recognise tremor (see Fig. 5), where negative responses somewhat resemble the dyskinetic acceleration profile shown in Fig. 4a.

5. CONCLUSIONS

In this paper, we have shown that genetic programming is able to design classifiers that can be used for monitoring a patient's response to levodopa therapy. We have also shown how the evolved classifiers can be analysed to identify the patterns of over-represented movement that are associated with tremor and dyskinesia.

Whilst the results seem promising, it should be born in mind that this is only an initial study involving a relatively small number of patients. Nevertheless, plans are currently underway to collect data from a larger group of patients, which will enable us to generate more robust classifiers, and also carry out a more rigorous study of the movement patterns underlying tremor, dyskinesia and other movement disorders associated with Parkinson's disease.

We can also expect to obtain better classifier performance by forming ensembles of evolved classifiers. Our previous work on Parkinson's diagnosis has shown the benefits of combining behaviourally diverse classifiers [11]. Given the diversity we have seen within the evolutionary runs in these experiments, this approach would seem particularly appropriate, especially if the movement patterns we see are complementary indicators of the underlying movement disorders.

Rather than generating a single classifier or ensemble for all patients, it may be more effective to train classifiers independently for each patient. Certainly our initial results (not reported here) seem to support this idea. Hence, in practice, we might expect the monitoring of levodopa treatment to be preceded by an initial evaluation that involves collecting movement data and training classifiers, rather than a one size fits all approach. However, such a personalised approach to medicine would be costly, especially given the growing incidence of Parkinson's disease. Nevertheless, the results presented in this paper indicate that sufficient generality can be obtained without considering patients individually.

6. ACKNOWLEDGMENTS

This work was part funded by Parkinson's UK and by the Wellcome Trust [ref: 097829] through the Centre for Chronic Diseases and Disorders (C2D2) at the University of York.

7. REFERENCES

[1] C. Ahlrichs and M. Lawo. Parkinson's disease motor symptoms in machine learning: A review. *Health Informatics*, 2, Nov. 2013.

[2] A. Bartolić, M. Šantić, and S. Ribarič. Automated tremor amplitude and frequency determination from power spectra. *Computer Methods and Programs in Biomedicine*, 94(1):77–87, Apr. 2009.

[3] P. Bonato, D. M. Sherrill, D. G. Standaert, S. S. Salles, and M. Akay. Data mining techniques to detect motor fluctuations in parkinson's disease. In *Engineering in Medicine and Biology Society, 2004. IEMBS'04. 26th Annual International Conference of the IEEE*, volume 2, pages 4766–4769. IEEE, 2004.

[4] X. Cai, S. L. Smith, and A. M. Tyrrell. Positional independence and recombination in cartesian genetic programming. In D. Hutchison et al., editors, *Genetic Programming, Proc. 9th European Conf., EuroGP 2006*, volume 3905 of *Lecture Notes in Computer Science*, chapter 32, pages 351–360. Springer Berlin Heidelberg, Berlin, Heidelberg, 2006.

[5] M. I. Chelaru, C. Duval, and M. Jog. Levodopa-induced dyskinesias detection based on the complexity of involuntary movements. *Journal of neuroscience methods*, 186(1):81–89, Jan. 2010.

[6] B. T. Cole, S. H. Roy, C. J. De Luca, and S. Nawab. Dynamic neural network detection of tremor and dyskinesia from wearable sensor data. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society.*, 2010:6062–6065, 2010.

[7] J. Gour, R. Edwards, S. Lemieux, M. Ghassemi, M. Jog, and C. Duval. Movement patterns of peak-dose levodopa-induced dyskinesias in patients with parkinson's disease. *Brain Research Bulletin*, 74(1-3):66–74, Sept. 2007.

[8] D. M. Halliday, J. R. Rosenberg, A. M. Amjad, P. Breeze, B. A. Conway, and S. F. Farmer. A framework for the analysis of mixed time series/point process data–theory and application to the study of physiological tremor, single motor unit discharges and electromyograms. *Progress in biophysics and molecular biology*, 64(2-3):237–278, 1995.

[9] N. L. W. Keijsers, M. W. I. M. Horstink, and S. C. A. M. Gielen. Online monitoring of dyskinesia in patients with parkinson's disease. *Engineering in Medicine and Biology Magazine, IEEE*, 22(3):96–103, May 2003.

[10] H. C. C. Kraemer, G. A. Morgan, N. L. Leech, J. A. Gliner, J. J. Vaske, and R. J. Harmon. Measures of clinical significance. *Journal of the American Academy of Child and Adolescent Psychiatry*, 42(12):1524–1529, Dec. 2003.

[11] M. Lones, S. Smith, J. Alty, S. Lacy, K. Possin, S. Jamieson, and A. Tyrrell. Evolving classifiers to recognise the movement characteristics of parkinson's disease patients. *IEEE Transactions on Evolutionary Computation*, 2013. Published online, in press.

[12] M. A. Lones, J. E. Alty, S. E. Lacy, D. R. S. Jamieson, K. L. Passin, N. Schuff, and S. L. Smith. Evolving classifiers to inform clinical assessment of parkinson's disease. In *Computational Intelligence in Healthcare and e-health (CICARE), 2013 IEEE Symposium on*, pages 76–82. IEEE, Apr. 2013.

[13] M. A. Lones and S. L. Smith. Objective assessment of Visuo-Spatial ability using implicit context representation cartesian genetic programming. In *Genetic and Evolutionary Computation: Medical Applications*, pages 174–189. John Wiley & Sons, Ltd, 2010.

[14] K. Niazmand, A. Kalaras, H. Dai, and T. C. Lueth. Comparison of methods for tremor frequency analysis for patients with parkinson's disease. In *Biomedical Engineering and Informatics (BMEI), 2011 4th International Conference on*, volume 2, pages 693–697. IEEE, Oct. 2011.

[15] G. Rigas, A. T. Tzallas, M. G. Tsipouras, P. Bougia, E. E. Tripoliti, D. Baga, D. I. Fotiadis, S. G. Tsouli, and S. Konitsiotis. Assessment of tremor activity in the parkinson's disease using a set of wearable sensors. *Information Technology in Biomedicine, IEEE Transactions on*, 16(3):478–487, May 2012.

[16] R. Saunders-Pullman, C. Derby, K. Stanley, A. Floyd, S. Bressman, R. B. Lipton, A. Deligtisch, L. Severt, Q. Yu, M. Kurtis, and S. L. Pullman. Validity of spiral analysis in early parkinson's disease. *Movement disorders*, 23(4):531–537, Mar. 2008.

[17] M. Tsipouras, A. Tzallas, E. Tripoliti, G. Rigas, P. Bougia, D. Fotiadis, S. Tsouli, and S. Konitsiotis. On assessing motor disorders in parkinson's disease. In J. Lin and K. Nikita, editors, *Wireless Mobile Communication and Healthcare*, volume 55 of *Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*, pages 35–38. Springer Berlin Heidelberg, 2011.