Going Through Directional Changes: Evolving Human Movement Classifiers Using an Event Based Encoding

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

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

Jeremy Cosgrove
Department of Neurology
Leeds General Infirmary
Leeds, UK

Stuart Jamieson
Department of Neurology
Leeds General Infirmary
Leeds, UK

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

ABSTRACT

Directional changes (DC) is an event based encoding for time series data that has become popular in financial analysis, particularly within the evolutionary algorithm community. In this paper, we apply DC to a medical analytics problem, using it to identify and summarise the periods of opposing directional trends present within a set of accelerometry time series recordings. The summarised time series data are then used to train classifiers that can discriminate between different kinds of movement. As a case study, we consider the problem of discriminating the movements of Parkinson's disease patients when they are experiencing a common effect of medication called levodopa-induced dyskinesia. Our results suggest that a DC encoding is competitive against the window-based segmentation and frequency domain encodings that are often used when solving this kind of problem, but offers added benefits in the form of faster training and increased interpretability.

CCS CONCEPTS

•Computing methodologies → Genetic programming; •Applied computing → *Health informatics;*

KEYWORDS

Genetic programming, directional changes, time series analysis, movement analysis, Parkinson's disease, dyskinesia

ACM Reference format:
Michael A. Lones, Jane E. Alty, Jeremy Cosgrove, Stuart Jamieson, and Stephen L. Smith. 2017. Going Through Directional Changes: Evolving Human Movement Classifiers Using an Event Based Encoding. In *Proceedings of GECCO '17 Companion, Berlin, Germany, July 15-19, 2017,* 7 pages.
DOI: http://dx.doi.org/10.1145/3067695.3082490

*Corresponding author: email m.lones@hw.ac.uk

GECCO '17 Companion, Berlin, Germany

978-1-4503-4939-0/17/07…$15.00
DOI: http://dx.doi.org/10.1145/3067695.3082490

1 INTRODUCTION

In order for time series data to be processed by a machine learning algorithm, it is often necessary to segment the data. A common approach to doing this is to use a sliding window [5], which involves splitting the time series into a sequence of (generally overlapping) n-tuples. Each of these n-tuples, or windows, is then used as input to the machine learning algorithm and elicits a response. These responses are then combined, typically through summing or averaging, into a single output for the whole time series. Whilst often effective, this approach can be sensitive to parameters such as the length of the window and the degree of overlap.

In the computational finance community, there has been a growing interest in event-based encodings of time series, notably the directional changes (DC) encoding developed by Olsen et al. [2, 15]. This approach has become particularly popular in evolutionary algorithm-based approaches to financial time series analysis [8, 9]. Rather than segmenting a time series into segments of equal length, these approaches segment a time series into a sequence of consecutive events. In the case of DC, these events are upturn events and downturn events, and the associated overshoot periods that occur between events. Both types of event occur where there is a significant and sustained change in the magnitude of a value being tracked.

Computational finance is not the only domain that might benefit from event based encodings. In this paper, we consider medical time series analysis, focussing on the analysis of abnormal human movements. As for a financial time series, important patterns in movement data are often indicated by changes in direction, such as a change in the sign of a velocity or a switch from acceleration to deceleration. These changes are the equivalent of upturn and downturn events in financial time series data, and can be captured using a DC encoding.

As a case study, we use a dataset collected to study levodopa-induced dyskinesia (LID) in Parkinson's disease patients [10]. LID is a common side-effect of dopamine replacement therapy, and causes uncontrollable muscle spasms that can significantly impair a patient's quality of life [14]. It can be managed through changes to dosage, but only if there is a reliable means for recognising and recording when a patient is experiencing dyskinesia. In previous work, Cartesian genetic programming was used to train classifiers that can identify periods of dyskinesia within accelerometry time series data, employing a window-based segmentation of the data

Table 1: Number of examples of each dyskinesia grade

UDysRS	Study 1	Study 2
0	2933	1747
1	1227	971
2	1688	562
3	681	183
4	64	361

[10]. In this work, we consider the benefits of using a segmentation based on DC.

The paper is organised as follows. Section 2 describes the dataset used in this work, and previous findings. Sections 3 and 4 describe the DC encoding that is applied to this data, and the evolutionary settings, respectively. Section 5 presents results, showing how the predictive ability of classifiers is affected by the use of a DC encoding. Section 6 concludes.

2 MOVEMENT DATASET

We use the dataset described in [10], which was collected from accelerometry devices worn by a cohort of 7 Parkinson's disease patients as they went about unscripted movement for a period of 6 hours. The resulting recordings were segmented into regions of interest by three trained clinicians, who identified periods of movement when the patients were displaying dyskinesia, labelling these with the corresponding clinical grade on the unified dyskinesia rating scale (UDysRS) [7], along with indications of the sensor positioning and the subject's current activity. The clinicians also labelled periods where the patients were displaying no dyskinesia. Samples of movement data with UDysRS ratings from 0 (no dyskinesia) to 4 (severe dyskinesia) were then extracted and each saved in the form of an acceleration time series. For the purpose of training and evaluating classifiers, the dataset was uniformly split three ways into training, validation and test sets. The training set was used for fitness evaluation, the validation set for early stopping, and the test set was used to estimate generality.

In this work, we make use of an additional, more recent, dataset which was collected using a similar experimental protocol from a larger group of 17 patients using a newer set of accelerometry modules. We refer to this as Study 2, and use it to provide an unbiased estimate of classifier accuracy for selected classifiers. Table 1 summarises the datasets from both studies.

In previous work with the first dataset [10], implicit context representation Cartesian genetic programming (IRCGP) [3] was used to train classifiers which could discriminate clinically significant grades of dyskinesia (3 & 4 on the UDysRS scale) from periods of motion when no dyskinesia was present. Cartesian genetic programming (CGP) is a form of genetic programming in which functions are laid out on a grid and connected together as a graph [12]. In addition to allowing implicit reuse of evolved sub-expressions, and displaying useful forms of neutral evolution, it has the practical benefit of limiting the size of the evolved expression and thereby preventing bloat. Consequently it tends to generate expressions which are more interpretable than those produced by conventional GP. This is particularly useful in a medical context, where it is often

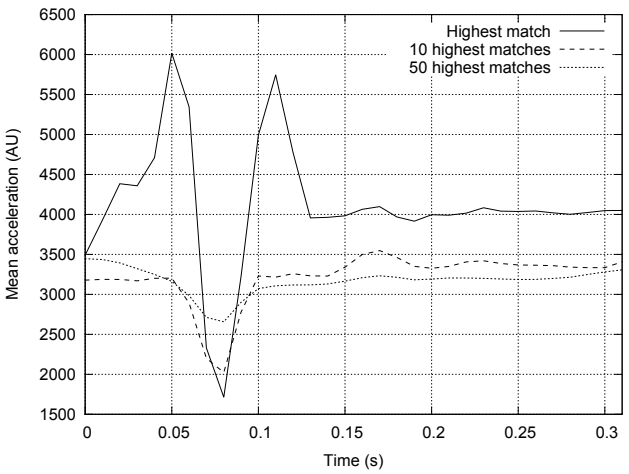


Figure 1: Example of an acceleration pattern recognised by a classifier trained using a fixed-size sliding window. Reproduced from [10].

necessary to understand the basis of a classification in order to for clinicians to have confidence in its prediction. IRCGP is a variant of CGP that is designed to improve the behaviour of crossover [3]; for details, see [11].

The results of this earlier study showed that classifiers could be evolved with a mean AUC (area under ROC curve) of about 0.9 when segmenting the time series data using overlapping sliding windows. This was significantly higher than the AUCs that could be achieved when training classifiers using frequency domain features, a notable result since most studies in this area focus on spectral analysis [13]. Analysis of the evolved classifiers suggested that their classification was based upon the shape of movements, rather than their magnitude or frequency. Fig. 1 shows an example of a pattern of acceleration that was found to be over-represented [10].

3 DIRECTIONAL CHANGES

DC segments a time series into a sequence of upturn and downturn events (see Algorithm 1, adapted from [2]). Both of these are sustained changes in the direction of a variable, measured in terms of a percentage change in magnitude using a threshold parameter, Δx_{dc}. For example, if $\Delta x_{dc} = 0.05$, an upturn event is recognised as having occurred when the variable increases in magnitude by at least 5% from its lowest point in the previous cycle.

DC is essentially a method for detecting turning points within a time series, and has commonalities with other methods that achieve this. However, the method does have several advantages over other methods when considered in the context of human movement analysis. Perhaps most notable in the ease of comprehension, since, for example, a 5% change in acceleration has an obvious physiological interpretation.

Also useful is the ability to change the scale of analysis by varying the threshold parameter. Fig. 2 shows an acceleration time series being encoded as a sequence of upturn and downturn events, illustrating the effect of changing Δx_{dc} from 20% to 10%. Multi-scale analysis of this kind using DC has been instrumental for the

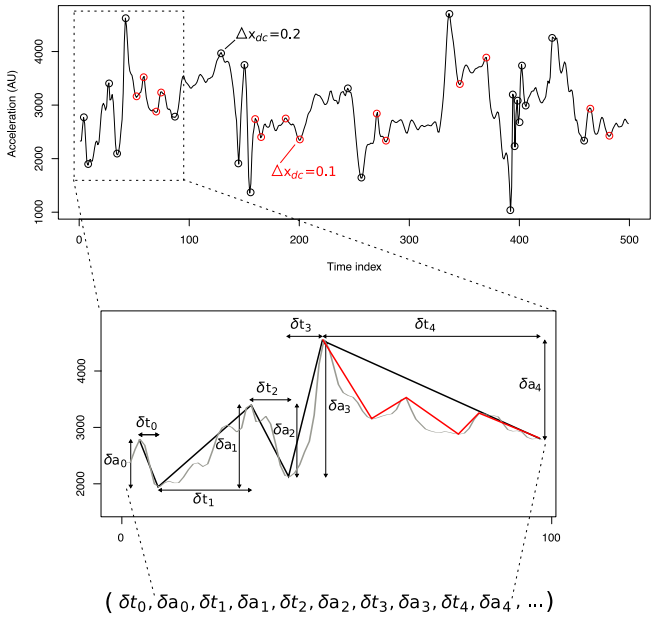


Figure 2: Encoding an acceleration time series as a sequence of directional change events, each summarised as a duration δt and a change in acceleration δa. The effect of varying the sensitivity parameter Δx_{dc} is also shown.

discovery of scale-free laws in financial time series data [6]. Since different neural programmes are thought to govern motion over different timescales, it could also help to understand how different disease pathways contribute to movement disorders [4].

Finally, the method always detects the exact point of inflection (i.e. the maximum or minimum in a cycle), rather than an approximation. This could be particularly beneficial for the analysis of movement data, where changes in behaviour can be abrupt and short-lived, and where peak acceleration is thought to be an important discriminator of motor ability.

4 CLASSIFIER EVOLUTION

For comparative purposes, we use the same approach reported in earlier work [10], where IRCGP was used to optimise mathematical expressions on a 6×6 CGP grid. A standard set of arithmetic functions is used $\{+, -, \times, \div, |x|, \text{mean}, \text{min}, \text{max}\}$. As in earlier work, the evolved expression is applied using a sliding window, and the overall output for a sequence is the mean of the outputs from the individual windows. However, unlike earlier work, the time series is first re-coded as an event sequence (as shown in Fig. 1), comprising ordered $(\delta t_e, \delta a_e)$ pairs for each event e. Each window contains a fixed number of events. A window is always aligned to start at a δt, and is slid along one event at a time.

Each run uses a population size of 200 and a generation limit of 50. 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%.

Algorithm 1 Encode time series *in* as *out* using directional changes

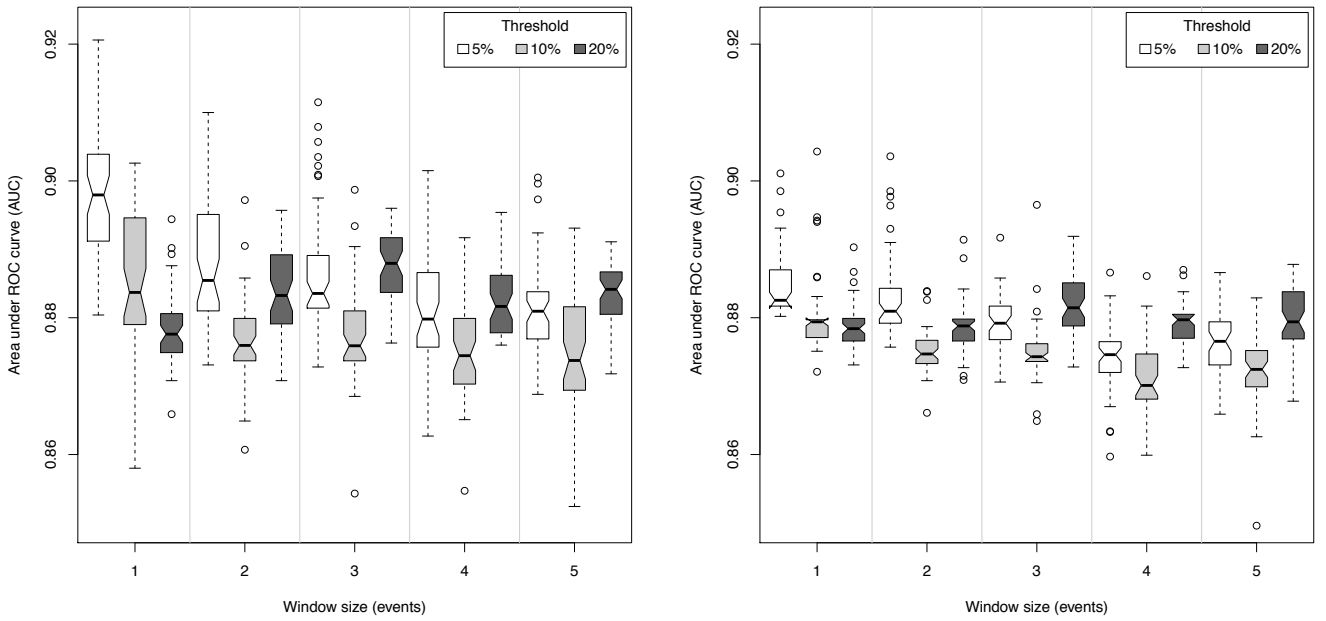
```
out ← ()                                          ▷ out is a sequence of numbers
event ← Upturn
high ← low ← in[0],  t_h ← t_l ← 0
for t ∈ 0 . . . |in| do                           ▷ Iterate through in
    if event = Upturn then
        if in[t] ⩽ high × (1 − Δx_dc) then        ▷ Upturn over
            event ← Downturn
            out ← out⌢(t_h − t_l, high − low)     ▷ Append (δt, δa)
            low ← in[t],  t_l ← t
        else
            if high < in[t] then                  ▷ Update high point
                high ← in[t],  t_high ← t
            end if
        end if
    else
        if in[t] ⩽ low × (1 + Δx_dc) then         ▷ Downturn over
            event ← Upturn
            out ← out⌢(t_l − t_h, low − high)     ▷ Append (δt, δa)
            high ← in[t],  t_h ← t
        else
            if low > in[t] then                   ▷ Update low point
                low ← in[t],  t_l ← t
            end if
        end if
    end if
end for
```

5 RESULTS

Fig. 3a shows the distributions of AUC across 50 runs using various combinations of window size and DC threshold. There are several notable observations. First, all combinations achieve a good level of discrimination, with AUCs ranging from 0.85 to 0.92. This is comparable to previous results on this data set, showing that the reduced information present in the DC encoding does not impair classification accuracy, possibly because it helps to emphasise patterns of interest. Second, the highest mean AUCs (~ 0.90) were achieved for a window size of one with a DC threshold of 5%. This is interesting, since it indicates that discrimination can be achieved by analysing each period of increasing or decreasing acceleration in isolation. Third, for a DC threshold of 20%, the results are significantly better when using a window size of three. This suggests that different discriminative patterns may be present at different temporal scales.

5.1 Expressions

To gain more insight into these results, we analysed individual classifiers. Equation 1 shows the evolved expression used by the classifier which performed best overall on the test set. This uses a window size of one, a DC threshold of 5%, and achieved an AUC of 0.92 on the test set.



(a) 6×6 CGP grid **(b) 3×3 CGP grid**

Figure 3: Ability of evolved classifiers to discriminate samples of Grades 3 and 4 dyskinesia from movements where no dyskinesia is present. For each combination of window size and DC threshold, notched box plots show distributions of AUC on the test set over 50 independent runs. Non-overlapping notches indicate a strong likelihood of statistically significant differences between means. Results are shown for two different CGP grid sizes.

$$
\begin{aligned}
exp &= \min(\min(s_1, \text{mean}(s_1 + 1, (s_3)^2)), \tfrac{s_1}{s_2} - \tfrac{s_3}{s_5}) \\
s_1 &= s_4 + 0.005 \\
s_2 &= \tfrac{s_5}{s_3} \\
s_3 &= \delta t_0 + 0.88 \\
s_4 &= \text{mean}(s_2, s6) \\
s_5 &= |\delta a_0| \\
s_6 &= \text{mean}(s_5 - \delta a_0, -\delta t_0 - 0.88)
\end{aligned} \tag{1}
$$

$s1 \ldots s5$ are sub-expressions in the solution graph that are used more than once. The overall expression is not easy to interpret. However, it is notable that event duration (s_3), the magnitude of the acceleration change (s_5), and the gradient (s_2) are all used in reaching a decision. Also notable is sub-expression s_6, where the first argument expands to $|\delta a_0| - \delta a_0$, leading to a different response depending on whether the event is an upturn or a downturn. This suggests that periods of increasing and decreasing acceleration should be treated differently in reaching a classification. This is in keeping with our understanding of human movements, where acceleration and deceleration are thought to be governed by different neural pathways [1].

Given the difficulty of analysing large evolved expressions such as this, we repeated the experiments using a CGP grid of 3×3, limiting expressions to a maximum of nine function instances. Fig. 3b shows the results, showing that the relationships between the distributions remain very similar to the original experiments, but with an overall drop in AUC of about ~ 0.02, which is small but significant. For a window size of one and a DC threshold of 5%, the solution shown in Equation 2 was found several times. This has a test set AUC of 0.88.

$$exp = |\delta a_0| - \delta t_0 \tag{2}$$

This expression is very simple in comparison to Equation 1. When applied to all the events in a particular movement sequence, the output of the classifier will be the difference between the mean magnitude of acceleration changes and the mean duration of each event. Given that acceleration values are in general numerically higher than durations, the magnitude term will be the dominant of the two, indicating that large changes in acceleration are particularly significant indicators of dyskinesia in this data set. However, when the duration term is numerically large—which will be the case when directional changes are infrequent—the classifier's output will be significantly reduced, indicating non-dyskinetic behaviour. This may help to filter out voluntary movement, such as walking, where the changes in acceleration are large but infrequent. It is particularly interesting that this expression evolved at the 5% DC threshold level, suggesting that the magnitude and duration of acceleration changes at a small scale are more useful for discriminating dyskinesia than those that occur at larger scales of movement, i.e. those that a clinician would tend not to notice during a clinical assessment of dyskinesia may be the most significant.

When a DC threshold of 20% is used, the best classifiers are found for larger window sizes. This suggests that classifiers may be

responding to different patterns at larger scales of movement. Equation 3 shows the expression used by the classifier which performed best overall (0.90) on the test set for a DC of 20%, and Equations 4 and 5 show two shorter expressions from the 3×3 runs, both achieving AUCs of 0.89 on the test set.

$$\begin{aligned} exp &= \text{mean}(\max(-\delta a_0, s_1 + -\delta t_1) - (\delta a_0 \delta a_1 + s_1 - \delta a_0), \\ &\quad 0.89(0.65 + \min(s_1 - \delta a_1, \delta t_0 \delta t_1))(s_1 - \delta a_0)) \\ s_1 &= \max(\delta a_0, \delta a_1) \end{aligned} \tag{3}$$

$$exp = \max(|0.007\delta a_0 - \delta a_1|, 0.06\delta a_1(\delta a_1 - \delta t_0)) \tag{4}$$

$$exp = \max(0.63|\delta a_0|, \delta a_0 - \min(\delta a_1, \delta a_0 \delta t_1) - \delta t_0) \tag{5}$$

Although the window size is 3, it is evident from these examples that the fittest classifiers among the 20% DC threshold group tend to use only the first two events in each window, suggesting that the third event is only useful in terms of evolvability. Equ. 5 is the easiest to interpret. The max and min functions both act as switches. Unless the first period of motion is much longer than the second, the max function returns the second term, which then evaluates to one of two expressions depending upon whether the window comprises an upturn followed by a downturn, or a downturn followed by an upturn. For the former case, it approximately returns the sum of magnitudes of the two changes in acceleration. For a downturn followed by an upturn, it approximately returns the magnitude of the first acceleration change multiplied by the duration of the second period of change. Notably, this is usually much larger than the value returned for the cycle from upturn to downturn, and so dominates the overall mean when applied to all the windows in a movement sequence. This seems to suggest that the relationship between a period of reducing motion and its following period of increasing motion is useful for discriminating dyskinesia.

The use of max and min functions for switching appears frequently within the fittest evolved solutions. This suggests that, in future work, it could be useful to introduce a more explicit mechanism for switching between behaviours, for instance the use of conditional execution.

5.2 ROC Curves

To gain a better understanding of the generality and relative utility of the selected classifiers, we re-evaluated them on the second data set. Fig. 4 shows the resulting ROC curves, showing how well they recognise the different grades of dyskinesia, and also how appropriate they are for different body positions and activities. For comparative purposes, we show a classifier from the previous study [10] alongside the overall best DC classifier (Equ. 1), the shortest classifier (Equ. 2), and the best classifier for a 20% threshold (Equ. 3).

These ROC curves give a more nuanced view of the DC classifiers' performance. Although they do generalise well to the second data set, the AUCs they achieve are slightly below those of the best non-DC classifier. However, the difference is small, and the considerable reduction in computational effort required to train a classifier using a DC sequence rather than a full time series may make them preferable in some circumstances; for example, when online training is used. For instance, the number of windows of data that need to be evaluated in the training data set is reduced by a factor of about 30 when using DC.

The 5% threshold classifiers do well at discriminating grade 3 and 4 dyskinesia, but are relatively poor at identifying instances of grade 1 and 2 dyskinesia. The 20% threshold classifier, on the other hand, has a slightly lower accuracy on grades 3 and 4, but discriminates grades 1 and 2 relatively well. They also differ in their trade-offs between sensitivity and specificity. This can be seen both in the lower grades and in the effect of sensor position, where the 5% classifiers favour specificity over sensitivity and the 20% classifier favours sensitivity over specificity. These observations, again, suggest that they are responding to different patterns in the data.

Fig. 4 also compares the ability of the classifiers to discriminate dyskinesia based on the subject's activity. All the classifiers perform well when the subject is sitting down, achieving high AUCs for all grades. However, discriminating dyskinesia whilst walking is known to be a difficult problem, and this is reflected in the plots. The DC classifiers are affected more than the non-DC classifier, suggesting that the shape of an acceleration change event is particularly important in this context. Nevertheless, it is notable that the 5% classifier does significantly better than the 20% classifier, indicating the relative benefit of considering fine movements when recognising dyskinesia in the presence of large-period voluntary movements, such as walking.

6 CONCLUSIONS

In this paper, we have described the use of a DC encoding when evolving classifiers for discriminating abnormal movements in Parkinson's disease. By segmenting an accelerometry time series into a sequence of events, each summarised by its duration and acceleration change, DC significantly reduces the size of movement data. This means that movement sequences can be processed with considerably less computational effort during fitness evaluation, which is important in applications such as online training and embedded learning, both of which are desirable in telemonitoring systems. It also reduces the information content. However, our results suggest that this loss of information has only a small effect upon classification accuracy. Furthermore, the reduction in extraneous information promotes the evolution of interpretable expressions. This is useful in the application domain described in this paper, where the movement disorder under study (dyskinesia) is incompletely understood, and where information about how to clinically discriminate it is potentially useful.

REFERENCES

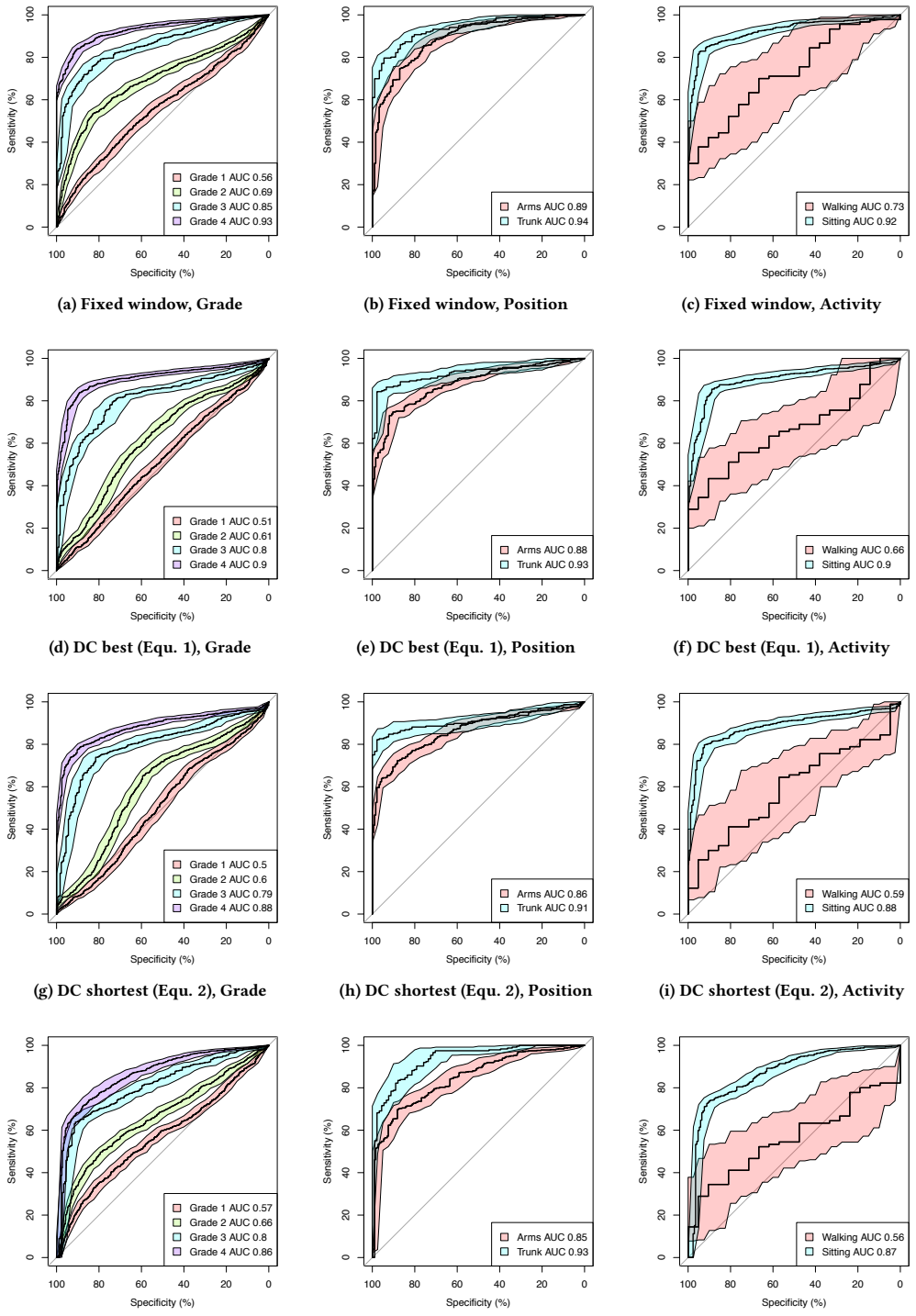
[1] Bhim M Adhikari, Kristen M Quinn, and Mukesh Dhamala. 2013. Is the brain's inertia for motor movements different for acceleration and deceleration? *PloS one* 8, 10 (2013), e78055.

[2] Monira Aloud, Edward Tsang, Richard B. Olsen, and Alexandre Dupuis. 2011. A Directional-Change Events Approach for Studying Financial Time Series. *Economics* 2011-28 (2011).

[3] Xinye Cai, Stephen L Smith, and Andy M Tyrrell. 2006. Positional independence and recombination in Cartesian genetic programming. In *European Conference on Genetic Programming*. Springer, 351–360.

[4] AR Crossman. 2000. Functional anatomy of movement disorders. *Journal of anatomy* 196, 4 (2000), 519–525.

[5] Thomas G Dietterich. 2002. Machine learning for sequential data: A review. In *Structural, syntactic, and statistical pattern recognition*. Springer, 15–30.



(a) Fixed window, Grade (b) Fixed window, Position (c) Fixed window, Activity

(d) DC best (Equ. 1), Grade (e) DC best (Equ. 1), Position (f) DC best (Equ. 1), Activity

(g) DC shortest (Equ. 2), Grade (h) DC shortest (Equ. 2), Position (i) DC shortest (Equ. 2), Activity

(j) DC 20% threshold (Equ. 3), Grade (k) DC 20% threshold (Equ. 3), Position (l) DC 20% threshold (Equ. 3), Activity

Figure 4: ROC curves for selected classifiers when re-evaluated on the second data set. The position plots show ability to separate grades 3 and 4 dyskinesia (from no dyskinesia) when using data sensors mounted in particular body regions. The activity plots show the ability to recognise grades 1–4 dyskinesia whilst subjects perform different activities. Note that the wide confidence bands for walking reflect a relatively sparsity of data for this activity. Confidence bands are calculated by repeated re-sampling, as implemented by the pROC package in R.

[6] JB Glattfelder, A Dupuis, and RB Olsen. 2011. Patterns in high-frequency FX data: discovery of 12 empirical scaling laws. *Quantitative Finance* 11, 4 (2011), 599–614.
[7] Christopher G Goetz, John G Nutt, and Glenn T Stebbins. 2008. The unified dyskinesia rating scale: presentation and clinimetric profile. *Movement Disorders* 23, 16 (2008), 2398–2403.
[8] Jeremie Gypteau, Fernando EB Otero, and Michael Kampouridis. 2015. Generating directional change based trading strategies with genetic programming. In *European Conference on the Applications of Evolutionary Computation*. Springer, 267–278.
[9] Michael Kampouridis and Fernando EB Otero. 2017. Evolving trading strategies using directional changes. *Expert Systems with Applications* 73 (2017), 145–160.
[10] Michael A Lones, Jane E Alty, Phillipa Duggan-Carter, Andrew J Turner, DR Jamieson, and Stephen L Smith. 2014. Classification and characterisation of movement patterns during levodopa therapy for parkinson's disease. In *Proceedings of the Companion Publication of the 2014 Annual Conference on Genetic and Evolutionary Computation*. ACM, 1321–1328.
[11] Michael A Lones, Stephen L Smith, Jane E Alty, Stuart E Lacy, Katherine L Possin, DR Stuart Jamieson, and Andy M Tyrrell. 2014. Evolving classifiers to recognize the movement characteristics of Parkinson's disease patients. *IEEE Transactions on Evolutionary Computation* 18, 4 (2014), 559–576.
[12] Julian F Miller. 2011. *Cartesian genetic programming*. Springer.
[13] Carlos Pérez-López, Albert Samà, Daniel Rodríguez-Martín, Juan Manuel Moreno-Aróstegui, Joan Cabestany, Angels Bayes, Berta Mestre, Sheila Alcaine, Paola Quispe, Geroid Ó Laighin, and others. 2016. Dopaminergic-induced dyskinesia assessment based on a single belt-worn accelerometer. *Artificial Intelligence in Medicine* (2016).
[14] Bhomraj Thanvi, Nelson Lo, and Tom Robinson. 2007. Levodopa-induced dyskinesia in Parkinson's disease: clinical features, pathogenesis, prevention and treatment. *Postgraduate medical journal* 83, 980 (2007), 384–388.
[15] Edward P. K. Tsang, Ran Tao, Antoaneta Serguieva, and Shuai Ma. 2017. Profiling high-frequency equity price movements in directional changes. *Quantitative Finance* 17, 2 (2017), 217–225.