Exploring Diagnostic Models of Parkinson's Disease with Multi-Objective Regression

Marta Vallejo
Department of Computer Science
Heriot-Watt University, UK

Jeremy Cosgrove
Department of Neurology
Leeds General Infirmary, UK

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

Stuart Jamieson
Department of Neurology
Leeds General Infirmary, UK

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

David W. Corne
Department of Computer Science
Heriot-Watt University, UK

Michael A. Lones
Department of Computer Science
Heriot-Watt University, UK
Email: M.Lones@hw.ac.uk

***Abstract*—Parkinson's disease is a progressive neurodegenerative disorder. The biggest risk factor for developing Parkinson's disease is age and so prevalence is increasing in countries where the average age of the population is rising. Cognitive problems are common in Parkinson's disease and identifying those with the condition who are most at risk of developing such issues is an important area of research. In this work, we explore the potential for using objective, automated methods based around a simple figure copying exercise administered on a graphics tablet to people with Parkinson's disease. In particular, we use a multi-objective evolutionary algorithm to explore a space of regression models, where each model represents a combination of features extracted from a patient's digitised drawing. The objectives are to accurately predict clinical measures of the patient's motor and cognitive deficit. Our results show that both of these can be predicted, to a degree, and that certain sub-sets of features are particularly relevant in each case.**

I. Introduction

In this work we investigate and describe the use of multi-objective evolutionary techniques to explore predictive models that can differentiate individuals with Parkinson's disease (PD) from healthy controls. Each subject was asked to trace a pre-defined shape using a digitising tablet. Various features were then extracted from the digitised drawings, and a multi-objective evolutionary algorithm (MOEA) was used to explore a space of linear and non-linear regression models in order to identify combinations of features that are predictive of clinical measures. We focus on two target measures: the Movement Disorders Society Unified Parkinson's Disease Rating Scale (MDS-UPDRS) Part 3 [1], which is a measure of motor features in PD, and the Montreal Cognitive Assessment (MoCA) [2], a validated measure of global cognition in PD. PD is often associated with motor dysfunction; however, cognitive aspects are becoming increasingly recognised [3], [4]. After ten years of PD diagnosis, half will have developed dementia [3], and so identifying biomarkers that might predict those most at risk of cognitive decline is an important research area. In addition, a simple mechanism of diagnosing PD would itself be desirable as some estimates place the misdiagnosis rate at 25% [5].

This paper is organised as follows: Section II gives a brief introduction to PD and its clinical assessment. Section III reviews related work. Section IV summarises the materials and methods used in this work. Section V presents results. Section VI concludes.

II. Parkinson's Disease

PD is a progressive neurodegenerative condition pathologically characterised by death of dopaminergic neurons in a part of the basal ganglia called the substantia nigra pars compacta. The cause of PD is unknown. Loss of dopaminergic neurons is primarily responsible for the motor features of PD, which are tremor, slowness (bradykinesia) and stiffness (rigidity). The other pathological hallmark of PD is the deposition of abnormal protein aggregates called Lewy bodies in the brain. Lewy bodies are thought to progressively infiltrate upwards through the brainstem towards the cortex, and the resultant damage to multiple neurotransmitter systems (e.g. acetylcholinergic and noradrenergic) is thought to drive the development of cognitive impairment [3].

Accurate diagnosis of PD can be challenging, particularly amongst non-specialists where high misdiagnosis rates have been reported [5], [6]. In non-research settings the diagnosis of PD is usually a clinical one, based on a history and examination for the pertinent motor features. The MDS-UPDRS Part 3 incorporates a full motor examination and is primarily used in research settings to allow the severity of motor symptoms to be objectively recorded. Each motor task is scored using a five-point scale but this process is known to be subjective, with considerable variance amongst clinicians [7]. A quick and accurate way of diagnosing PD would be most welcome.

As well as Parkinson's disease dementia (PDD), where subjects are, by definition, unable to manage with activities of daily living (ADLs), there is also a state where abnormalities on cognitive tests can be identified in PD but a person is able to manage independently with ALDs; this has become known as PD-mild cognitive impairment (PD-MCI). In our study we were able to classify subjects with PD into those with normal cognition (PD-NC), those with PD-MCI and those with PDD using recognised diagnostic criteria [8], [9].

Biomarkers to identify those with PD or PD-MCI most at risk of developing PDD are urgently needed. One avenue of investigation is to look for associations between motor function and cognition, and our work was motivated by this.

III. Related Work

Figure copying tests, such as clock drawing [10] and the Rey-Osterrieth complex figure [11], have long been used in the assessment of cognitive deficit in neurodegenerative conditions. In most cases, a subject's drawing is assessed by a clinician using a standardised procedure that considers, for example, the placement and sizing of different figure elements. Automating this process is challenging, given that a subject's drawing may be quite different from the original template [12]. Nevertheless, there has been some success. Notably, Souillard-Mander et al. [13] used machine learning techniques to analyse the elements of digitised clock drawings, achieving reasonable discrimination between PD subjects and age-matched controls.

A more effective approach may be to use figures that are amenable to computational analysis. An example of this is our previous work with the pentagon spiral figure, a relatively simple drawing template that is nevertheless designed to amplify the clinical symptoms of PD (if present) by requiring the subject to repeatedly accelerate and decelerate. We have previously had some success with applying evolutionary algorithms to these recordings, by using them to train classifiers that respond to over-represented patterns of acceleration [14]. However, a limitation of this approach is interpretability, since it is challenging to understand how these evolved classifiers reach a decision. This is important, since a classifier can only be useful as a decision support tool if a clinician has confidence in the basis of its decision.

The idea of automated, objective diagnosis for PD has also been explored in the context of voice signals. For instance, Tsanas et al. [15] recorded the speech of a number of PD subjects and controls. They, and others, have since used various machine learning algorithms to build predictive models that discriminate between the vocal patterns of these two groups. However, this system is yet to be used in practice, perhaps due to the relatively small number of recordings in the corpus and the frequent use of black box models.

In this work, we are not aiming to generate a specific predictive model. Rather, we are using MOEAs to explore a space of models, with the intention that these models will be used to inform diagnosis and prognosis of PD more generally. As such, we focus on relatively simple regression models which are interpretable and also widely understood by clinicians. The results reported in this paper build upon an earlier feasibility study [16].

IV. Materials & Methods

A. Pentagon Spiral Figure

Following [14], we used the pentagon spiral figure (see Fig. 1a) as a means of capturing information about a subject's physiological state. This figure-copying task involves a subject repeatedly copying a pentagon shape whilst moving steadily outwards from a central point. Repeated movements are known to cause fatigue in PD patients, and continual acceleration and deceleration emphasises the presence of bradykinesia, the cardinal motor feature of PD. As well as providing a measure of motor dysfunction, the pentagon spiral figure also tests a subject's visuospatial abilities, which is a cognitive domain that can be affected in PD-MCI and PDD.

B. Assessment Scores

We use two clinical assessment scores in this work: the MDS-UPDRS Part 3 (referred to as UPDRS from now on) [1] and the MoCA [2], [17]. Both are composite scores that result from a series of tasks and observations administered by a clinician, each of which is scored separately. UPDRS scores have a composite value between between 0 and 132, with higher scores indicating greater motor dysfunction. Composite MoCA scores are ranged between 0 and 30, with lower scores indicating worsening cognitive dysfunction. A MoCA score below 26 is abnormal, although an assessment of ADLs is required to further classify PD subjects with cognitive impairment into PD-MCI and PDD.

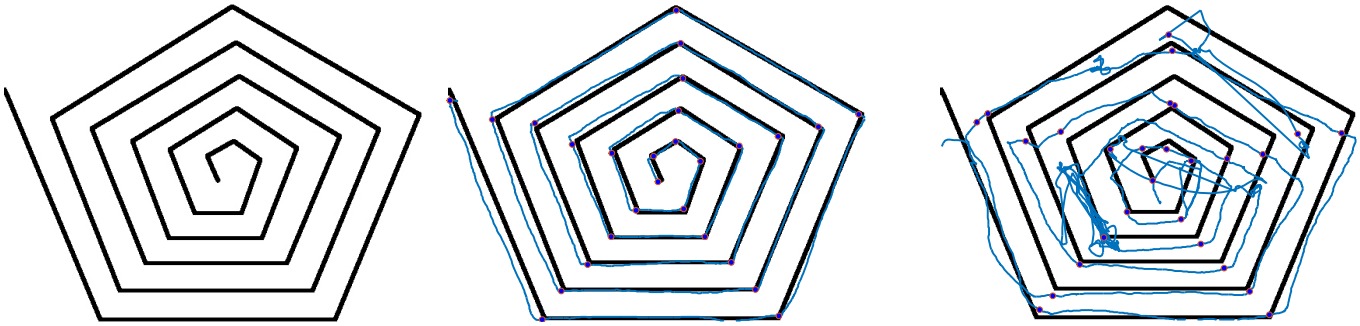
C. Data Collection

After obtaining full ethical approval, 58 PD subjects and 29 age-matched healthy controls were recruited at the Leeds General Infirmary. Each subject first underwent a standard clinical assessment, during which their UPDRS and MoCA scores were calculated. Of the PD subjects, 22 were classified as PD-NC, 26 as PD-MCI and 10 as PDD. Of the controls, 19 had normal cognition, and 10 had impaired cognition (MoCA< 26).

The subjects were then asked to trace over a template of the pentagon spiral figure on an A4-sized Wacom digitising tablet using an inking stylus. As they carried out the drawing, their movements were digitised and stored to a disk file as a time series of $< x, y >$ coordinates, alongside associated pen pressure and orientation information (not currently used in this study). The sampling frequency of the tablet was set at 200Hz. Each subject was asked to repeat the tracing process three times using their dominant and non-dominant hands in turn.

D. Feature Extraction

Table I summarises the list of features that were retrieved from each drawing. These features were all selected in consultation with PD experts, taking into account the PD literature and insights from other standard clinical tests like finger



(a) The drawing template

(b) Example of a control subject's spiral pentagon drawing overlaid on the template

(c) In this case, the concentric lines of the pentagon and the corners are challenging to distinguish

Fig. 1: The pentagon spiral figure. On the left is the template presented to a subject. The other two images show digitised recordings by study subjects overlaid on the template: the first from a control, the second a difficult case from a PD subject.

tapping, studies on grasp control and the gold standard clinical definition of bradykinesia.

Most of the features, with the exception of 1 and 2, were extracted automatically from the tablet recordings with the use of a set of custom-written Matlab scripts. The data extraction started with the automatic alignment of each drawing with respect to the template (Fig. 1a), focusing especially on finding the corners with the use of the minimum eigenvalue method [18]. In some cases the proper assignment of corners was problematic even if the task was done manually (see Fig. 1c), and this is likely to introduce some error during feature extraction.

TABLE I: Features extracted from the digitised drawings.

ID	Name	Feature Description
1	dominant	Dominant or non-dominant hand used
2	attempts	Repeat number: between 1 and 3 for each hand
3	totalTime	Total time: time taken to trace the entire figure
4	areaError	Area error: Area between the template pentagon and the pentagon drawn by the patient
5	distance	Total distance travelled by the pen
6	leaveSurface	Times the pen leaves the tablet surface: patients are instructed to not remove the pen from the tablet
7	timeContact	Total time the pen was in contact with the tablet
8	zeroVel	Duration of zero velocity during the whole task
9	zeroAcc	Duration of zero acceleration during the whole task
10	peakVel	Maximum peak velocity
11	avgVel	Average velocity
12	distPeakV	Distance to maximum peak velocity
13	timePeakV	Time to maximum peak velocity
14	timePeakV	Maximum peak acceleration
15	peakDesAcc	Minimum peak deceleration
16	avgAcc	Average acceleration
17	avgDec	Average deceleration
18	timePeakA	Time to maximum peak acceleration
19	timePeakD	Time to minimum peak deceleration
20	distPeakA	Distance to maximum peak acceleration
21	distPeakD	Distance to minimum peak deceleration
22	timeAAbs	Time in acceleration. Absolute
23	timeARel	Time in acceleration. Relative (% of movement time)
24	timeDAbs	Time in deceleration. Absolute
25	timeDRel	Time in deceleration. Relative (% of movement time)
26	totalNumPeaks	Total number of peaks in acceleration-deceleration
27	peakP	Number of peaks until maximum peak
28	peakN	Number of peaks until minimum peak
29	covV	Coefficient of variance (COV) in velocity
30	covA	COV in acceleration
31	covD	COV in deceleration

E. Polynomial Regression

Regression modelling is one of the most widely used tools for classification of samples and feature selection. The method allows the identification of statistical associations and causal relationships and may provide greater understanding of the underlying data generating process. In multivariable regression, a single dependent variable that is the outcome is inferred with the use of multiple independent variables or predictors. The formulation of a basic linear regression model can be described as follows:

$$y = \alpha + \beta x + \epsilon \tag{1}$$

where y is the dependent variable, x the independent one, α is the intercept, β is the predictor and ϵ is the error. This model can be extended in order to include more than one predictor. Then it is called multivariable regression modelling:

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n + \epsilon \tag{2}$$

where in this case β_0 is the intercept, $\beta_1 - \beta_n$ is a vector of n predictors, $x_1 - x_n$ a vector of independent variables and ϵ the error.

The model can be additionally enriched to include non-linear relationships in order to improve the capability of fitting more complex functions. In such a scenario, the final polynomial expression could be formulated as:

$$y = \beta_0 + \beta_1 x_1^{\theta_1} + \cdots + \beta_n x_n^{\theta_n} + \epsilon \tag{3}$$

where $\theta_1 - \theta_n$ depict the exponents to which the predictors β_i and independent variables x_i are raised.

A general characteristic of regression modelling is that the predictors derived using the final selected structure of the regression model parameters are applied globally over the problem under consideration assuming spatial non-stationarity. This could be the major limitation of the technique if local differences among areas of the study are significant.

This work uses a conventional ordinary least squared (OLS) regression analysis. This is a typical metric applied to linear regression problems to assess the suitability of a model. The calculation can be mathematically defined as follows:

$$a = \arg\min_{a} \sum_{j=1}^{n} \Big(y_j - F(x_j) \Big)^2 \tag{4}$$

where y_j is the final score or dependent variable for the i-th training example, x_j is the predictor for the same training example over the model and n is the total number of samples.

F. Pareto Archived Evolution Strategy

Multi-objective optimisation is an active research area, with many options available for algorithm design. The available techniques mainly differ in the details of the selection operator, which determines why one solution is chosen rather than another to be a progenitor of new candidate solutions. While this step is a relatively straightforward decision in single-objective optimisation, the area of multi-objective optimisation is essentially defined in terms of precisely how this process is handled.

Reflecting the variance and complexity of real-world applications, no single MOEA design approach has emerged as dominant in the field. However there is a subset of commonly used MOEAs that tend to perform well across the board. We make the pragmatic decision to choose the Pareto Archived Evolution Strategy (PAES) [19] from this commonly used subset, in part because the application in this paper is relatively compute-intensive, and PAES tends to perform well in terms of the speed/performance trade-off, in part due to its maintenance of a small population of solutions.

The disadvantage of a small population is the lack of diversity, which may lead to premature convergence to a local optimum. To avoid this, the algorithm was run multiple times using the same problem configuration, with the archive of best solutions shared and modified among these executions.

G. Model Training

A regression learning procedure aims at searching for the best set of parameters that fits a determined real function. In cases where a polynomial structure is required to capture the essence of a non-linear system, the selection of the degrees of each term that form the polynomial normally relies on a personal decision. Since this is a crucial feature which influences the effectiveness of the function approximation, a more systematic approach should be considered.

In this regard, the traditional Anderson's procedure [20] proposes the discovery of the optimum degree in a polynomial regression by testing in sequence the entire set of coefficients from its maximum boundary to 0. However, if the system under consideration is known to be non-linear and highly complex, or characterised by a significant number of possible features, an exhaustive search of all plausible combinations quickly becomes infeasible.

Algorithm 1 PAES algorithm

```
Require: Max_iter
  Generate int_sol and set it as Current_sol
  Evaluate fitness values of the Current_sol
  Add Current_sol to archive
  for i = 1 To Max_Iter do
    Randomly select one factor to mutate (power, feature);
    Generate new_sol by mutating Current_sol;
    Evaluate fitness values of the New_sol
    if New_sol dominates Current_sol then
      Set New_sol as Current_sol
      Update archive
    else
      if Current_sol dominates New_sol then
        Discard New_sol
      else    ▷ Current_sol and New_sol do not dominate each other
        Update archive using New_sol
        if New_sol dominates any member of the archive then
          remove them
          add New_sol to archive
        else
          add New_sol to archive
          randomly select a Current_sol among New_sol and Current_sol
        end if
      end if
    end if
  end for
  return Non-dominated solutions
```

An alternative way of discovering the best structure is with the use of metaheuristics like evolutionary algorithms. The use of hybrid methods that merge both technologies were successfully reported in different works [21], [22].

Following this approach, the present work is structured as a two-stage method that searches for a model structure that best fits two scores, UPDRS and MoCA. These scores are conflictive in nature due their applicability to different symptomatology areas of PD, cognitive and motor deficit.

By wrapping both a weighted polynomial regression analysis and the conventional OLS that pursues the minimisation of a given cost function which computes the difference between the scores and the result of the polynomial solution, the EA approach can provide a successful model to select the most relevant set of features for both aspects of the disease. In this regard, the EA is intended to estimate not only the most significant set of parameters, but also the best structure within the polynomial expression.

In this case this minimisation is multi-objective $(\mathcal{F}_1(x), \mathcal{F}_2(x)) = \mathcal{F}(x)$. The set of optimal solutions, also called the Pareto optimal front, is the set of non-dominant solutions where no further Pareto improvements can be made. The objective is to find a Pareto approximation that is located as closer as possible to the true Pareto front.

The starting point of the process will be the execution of the evolutionary algorithm PAES by the random generation of a single individual solution. This solution, also called a chromosome, represents a model composed by a subset of features each linked with a corresponding degree with a polynomial structure. A polynomial can be defined as an expression consisting of variables and coefficients. Consequently the formal chromosome representation can be characterised by two different vectors represented as follows:

$$f = (f_1, f_2 \ldots, f_8), f_i \in \{1 - 31\} \tag{5}$$

$$a = (a_1, a_2 \ldots, a_8), a_i \in \{1 - 6\} \tag{6}$$

where f is a vector of features and a denotes the vector of exponents/powers of the predictors. If a polynomial evolves to an expression that contains like terms, which means the same features with the same exponent, these terms are not combined.

Every time the algorithm finds a good non-dominated solution, this solution is stored in a repository, called an archive, that for our purpose is defined with unlimited size. The archive represents the current approximation of the Pareto front. For each execution of the algorithm, the candidate solution is evolved for 400 generations using selection and mutation operators.

Mutation is defined twofold. Firstly it can be aimed at producing a disturbance in the exponent value of a single term. This type of mutation has the capability of adding and removing terms in the equation. Secondly the mutation can be intended to modify the feature selected within this formula. Features can be repeated within an equation.

The algorithm is executed for each exponent a number of times until achieving five consecutive runs without finding a new non-dominated solution. These executions cover all the range of exponents investigated in this work. Finally the algorithm returns the solutions allocated in the archive with the best prediction using the root-mean-square metric.

H. Model Validation

In order to assess the adequacy of a given model and to avoid overfitting this work uses a division of the data into three subsets: a training and validation set that are used to fit each individual optimisation execution, and a hold-out test that is used to measure the generalization error of the final model [23]. Concretely, from the 50 subjects in the dataset, 35 of them are used to train the regression model, and 15 are assigned to the hold-out set used for testing.

V. Results

One wide-spread method used in the multi-objective community to evaluate the performance of the evolutionary algorithm is the hypervolume indicator [24]. This indicator allows the transformation of the problem into a single objective by mapping a set of points into a scalar. This scalar is the result of calculating the volume covered by the approximation of the Pareto front $P \subset \mathbb{R}^d$ and a reference set $R \subset \mathbb{R}^2$. In this work, the dimension of the multi-objective problem is $d = 2$ and the reference set is formed by two points called utopia and anti-utopia $r_u, r_a \subset \mathbb{R}^2, R = \{r_u, r_a\}$. The calculation of the hypervolume is formalised as follows:

$$I(H)_A := \lambda(H(P, R)) \tag{7}$$

where H represents the set of objective vectors that are enclosed in this front and λ depicts the Lebesgue measure.

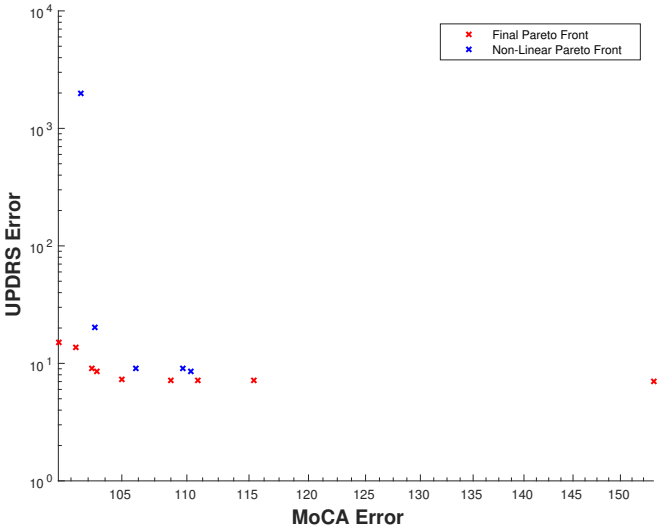


Fig. 2: The set of the non-dominated solutions. Red points show the solutions of the Pareto front, and blue points show the best non-linear regression models. For visualisation purposes, UPDRS errors are shown on a logarithmic axis.

The model training process was repeated for maximum regression exponents from 1 (linear models) to 5. Fig. 2 shows both the overall Pareto front and the front comprised of non-linear models (i.e. those containing exponents greater than 1), showing that the Pareto front is comprised exclusively of linear models. This is further highlighted in Fig. 3, depicting the hypervolumes achieved when using models of different complexity, and showing that runs where only linear models were evolved achieved the highest accuracy. This is perhaps surprising, given the limited expressiveness of linear regression models. However, more complex models (i.e. polynomial models in this case) are more likely to overfit data, and this effect is likely to be more pronounced when using a small data set. The instability of higher order solutions is reflected in the non-linear Pareto front, which is relatively sparse, presumably

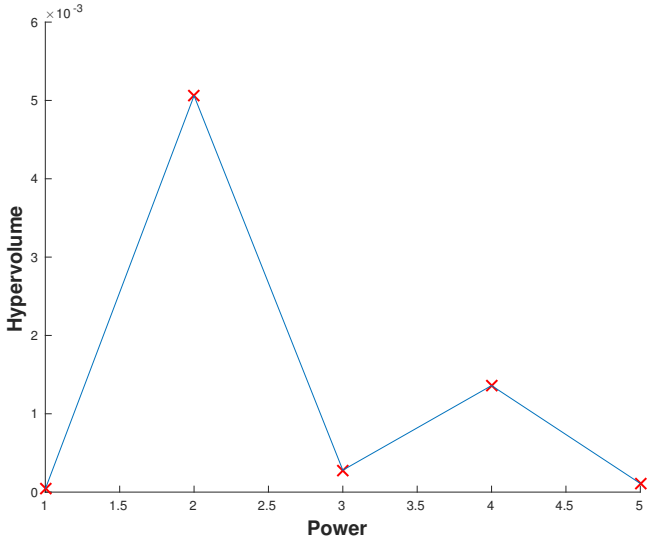


Fig. 3: The relationship between the performance of the MOEA, calculated by means of the hypervolume, and the maximum exponent allowed in the construction of the polynomial regression models.

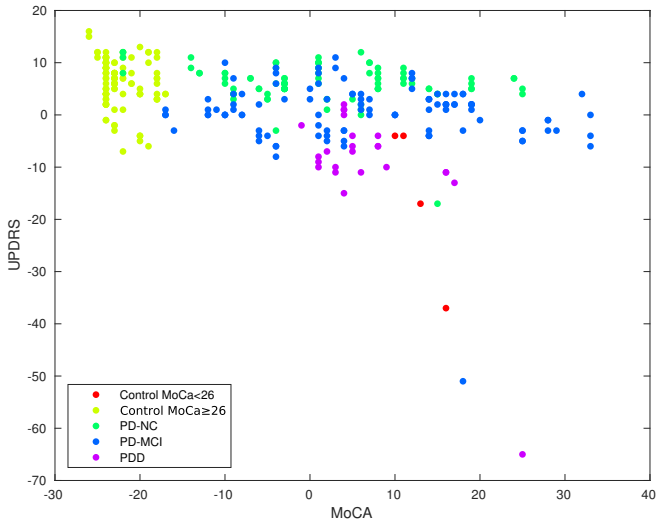


Fig. 4: Visual representation of the residuals resulted from taking the best model found for each of the scores, UPDRS and MoCA, in the archive of best solutions. The scatter plot illustrates regression errors from each subject where colours represent different cognitive conditions that are used to classify the disease: patients without cognitive impairment (PD-NC), mild cognitive impairment (PD-MCI) and patients with dementia (PDD).

because many solutions decreased in fitness when re-evaluated on the hold-out set. The linear front, by comparison, maintains a reasonable spread of solutions.

Although the regression errors (see Fig. 2) are similar for UPDRS and MoCA, both < 10 on average, when taking into

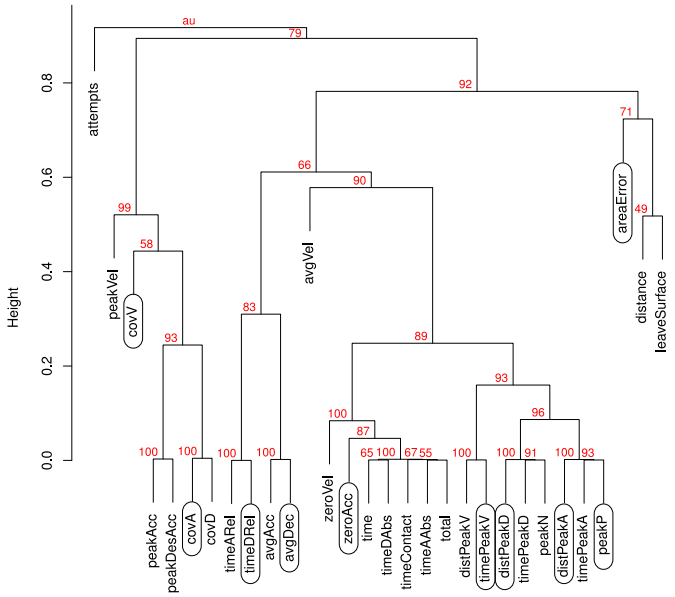


Fig. 5: Clustering dendrogram (generated using the R package pvclust) showing how the different extracted features correlate. Lower branches indicate higher correlations. Approximately unbiased p-values (AU) are shown in red text; branches with values greater than 95% are strongly supported by the data. Feature used in the evolved regression models shown in Table II are circled.

account the MoCA range the errors on this cognitive score are comparatively larger. This may suggest that motor deficit is easier to predict than cognitive deficit. It may also be the case that the drawings contains less cognitive information than motor information, or that MoCA is a less reliable regression target (since clinical scores are subjective and liable to different error rates). Nevertheless, it is apparent from Fig. 4 (which shows the per-subject error rates) that error rates are not uniform across the cohorts. For example, the UPDRS error rate for the PDD group is relatively low. For cognitively-normal controls, UPDRS scores are consistently under-predicted. This may be a limitation of using linear models, since the UPDRS scores for this cohort are much higher than the other cohorts and may be outside the range of the predictor.

Table II and III describe the solutions in the Pareto front. As noted, these are all linear solutions (hence powers of 1 for each feature). It is notable that the solutions include only a subset of the available features, and each has only a small number of terms. Again, this may be required to avoid over-learning in a small data set. However, it is quite advantageous from an interpretability perspective.

The features that are used across all solutions (including both the UPDRS and MoCA ends of the Pareto front) are the area error (f_4), the average rate of deceleration (f_{17}, avgDec) and the percentage of time spent in deceleration (f_{25}, timeDRel). The use of area error is unsurprising, since the accuracy of the drawing would expect to be lessened

	Power	MoCA	UPDRS	Features	Powers
1	1	102.763	9.044	4,9,25,4,21	1,1,1,1,1
2	1	110.893	7.202	17,4	1,1
3	1	100.339	15.005	17,4,27,19	1,1,1,1
4	1	153.575	7.055	17,4,29	1,1,1
5	1	105.010	7.247	25,25,25	1,1,1
6	1	103.129	8.520	25,25	1,1
7	1	115.359	7.106	25	1
8	1	101.589	13.735	20,25	1,1
9	1	108.737	7.216	30,25	1,1

TABLE II: Content of the archive with the non-dominated polynomial regression models gathered from all the sub-archives generated by the different exponents. These solutions constitute the Pareto front approximation. Columns of the table show maximum power/exponent of the polynomial, residual errors for both scores (UPDRS and MoCA), features selected and powers of the corresponding terms.

	Weights UPDRS	Weights MoCA
1	24.219,-0.778,0.098,-0.415,0.195,-0.166	21.329,4.722,3.079,1.505,1.386,1.480
2	24.219,-0.765,-0.225,0.181	21.329,2.915,-3.742,0.937,1.386
3	24.219,-0.790,-0.852,0.181,0.195,-0.162	21.329,3.033,4.864,0.937,1.386,1.516
4	24.219,-0.805,-0.870,0.181,0.195,-0.162	21.329,3.161,5.356,0.937,1.386,1.516
5	24.219,-0.127,-0.137,-0.169	21.329,1.760,1.397,3.494
6	24.219,-0.162,-0.169	21.329,1.397,3.494
7	24.219,-0.143	21.329,1.397
8	24.219,-0.132,-0.169	21.329,1.397,3.487
9	24.219,-0.137,-0.172	21.329,1.286,3.345

TABLE III: The weights of each polynomial depicted in II. During evaluation, each polynomial is fitted separately to each regression target, hence two sets of weights are shown.

with either motor or cognitive impairment. The importance of deceleration is more interesting, and may relate to a recent observation that deceleration and acceleration are not coordinated by the same brain region [25]. Hence, it is possible that PD particularly affects the brain region responsible for control of deceleration, and this could be a potential biomarker. The importance of observing deceleration during diagnosis also reflects our previous work on analysing finger tapping in PD subjects, where we identified the closing deceleration of a tap cycle to be particularly discriminative between PD patients and controls [26].

The predictors for MoCA additionally make use of the following features: the distance or time until peak deceleration or acceleration (f_{19} timePeakD, f_{20} distPeakA, and f_{21} distPeakD), the number of peaks until peak acceleration (f_{27}, peakP), and (to a lesser extent) the duration of time spent at zero acceleration (f_9, zeroAcc). This suggests that the presence of multiple bursts of acceleration may particularly be an indicator of cognitive dysfunction. This may further relate to the problem of under-scaling of movements seen in many PD subjects, where a subject under-estimates the amount of acceleration required to reach a target (in this case a corner) and consequently must perform further accelerations to compensate. It is quite plausible that this behaviour is over-represented in those with cognitive variants of the disease, i.e. PD-MCI and PDD.

Fig. 5 depicts more general relationships observed between the features, derived using bootstrap hierarchical clustering of their correlation matrix. The features selected for inclusion in the evolved regression models (circled in the figure) appear to be spread across the feature space. However, there are some measures that do not appear to be useful, including gross metrics such as the total time and distance travelled, and the overall peak and average velocities. This may point to the importance of considering more subtle features when carrying out diagnosis, and likewise the importance of using computational techniques to measure these subtle features.

VI. Conclusions

This study aims to use multi-objective evolutionary techniques to explore the space of diagnostic models associated with PD. In this paper, we applied this approach to studying predictive regression models that capture a patient's degree of motor and cognitive dysfunction. Our results suggest that clinical measures of motor and cognitive decline can be predicted, although to varying degree within different patient subpopulations. Analysis of the resulting Pareto front of solutions is informative, suggesting that patterns of deceleration are particularly diagnostically significant. In addition, the presence of multiple bursts of acceleration may be especially significant for predicting cognitive dysfunction. From a modelling perspective, linear models appear to be more predictive, though this may be an indication of their relative stability on small data sets in comparison to polynomial models.

In future work, we aim to look at a wider range of diagnostic objectives, for example the ability to predict PD in its early stages. It would also be interesting to look at different kinds of predictive model, for example regression trees. More work is also required to fully understand the medical significance of our findings.

ACKNOWLEDGEMENTS

This research was supported by the UK Engineering and Physical Sciences Research Council through the grant "A Multiobjective Evolutionary Approach to Understanding Parkinson's Disease" (Ref: EP/M013677/1). Data generated during this research is available at the following DOI: 10.17861/958af07e-d336-4202-854e-12188211873a.

REFERENCES

[1] 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*, vol. 23, no. 15, pp. 2129–2170, 2008.

[2] Z. S. Nasreddine, N. A. Phillips, V. Bédirian, S. Charbonneau, V. Whitehead, I. Collin, J. L. Cummings, and H. Chertkow, "The montreal cognitive assessment, moca: a brief screening tool for mild cognitive impairment," *Journal of the American Geriatrics Society*, vol. 53, no. 4, pp. 695–699, 2005.

[3] J. Cosgrove, J. E. Alty, and S. Jamieson, "Cognitive impairment in Parkinson's disease," *Postgraduate medical journal*, vol. 91, no. 1074, pp. 212–220, 2015.

[4] J. Pagonabarraga and J. Kulisevsky, "Cognitive impairment and dementia in Parkinson's disease," *Neurobiology of disease*, vol. 46, no. 3, pp. 590–596, 2012.

[5] C. B. Levine, K. R. Fahrbach, A. D. Siderowf, R. P. Estok, V. M. Ludensky, and S. D. Ross, "Diagnosis and treatment of parkinson's disease: A systematic review of the literature: Summary," 2003.

[6] N. P. Bajaj, V. Gontu, J. Birchall, J. Patterson, D. G. Grosset, and A. J. Lees, "Accuracy of clinical diagnosis in tremulous parkinsonian patients: a blinded video study," *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 81, no. 11, pp. 1223–1228, 2010.

[7] 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*, vol. 26, no. 10, pp. 1859–1863, 2011.

[8] M. Emre, D. Aarsland, R. Brown, D. J. Burn, C. Duyckaerts, Y. Mizuno, G. A. Broe, J. Cummings, D. W. Dickson, S. Gauthier *et al.*, "Clinical diagnostic criteria for dementia associated with parkinson's disease," *Movement disorders*, vol. 22, no. 12, pp. 1689–1707, 2007.

[9] I. Litvan, J. G. Goldman, A. I. Tröster, B. A. Schmand, D. Weintraub, R. C. Petersen, B. Mollenhauer, C. H. Adler, K. Marder, C. H. Williams-Gray *et al.*, "Diagnostic criteria for mild cognitive impairment in parkinson's disease: Movement disorder society task force guidelines," *Movement Disorders*, vol. 27, no. 3, pp. 349–356, 2012.

[10] E. Pinto and R. Peters, "Literature review of the clock drawing test as a tool for cognitive screening," *Dementia and geriatric cognitive disorders*, vol. 27, no. 3, pp. 201–213, 2009.

[11] M.-S. Shin, S.-Y. Park, S.-R. Park, S.-H. Seol, and J. S. Kwon, "Clinical and empirical applications of the rey–osterrieth complex figure test," *Nature protocols*, vol. 1, no. 2, pp. 892–899, 2006.

[12] R. Canham, S. Smith, and A. Tyrrell, "Location of structural sections from within a highly distorted complex line drawing," *IEE Proceedings-Vision, Image and Signal Processing*, vol. 152, no. 6, pp. 741–749, 2005.

[13] W. Souillard-Mandar, R. Davis, C. Rudin, R. Au, D. J. Libon, R. Swenson, C. C. Price, M. Lamar, and D. L. Penney, "Learning classification models of cognitive conditions from subtle behaviors in the digital clock drawing test," *Machine learning*, vol. 102, no. 3, pp. 393–441, 2016.

[14] S. L. Smith, P. Gaughan, D. M. Halliday, Q. Ju, N. M. Aly, and J. R. Playfer, "Diagnosis of Parkinson's disease using evolutionary algorithms," *Genetic Programming and Evolvable Machines*, vol. 8, no. 4, pp. 433–447, 2007.

[15] A. Tsanas, M. A. Little, P. E. McSharry, J. Spielman, and L. O. Ramig, "Novel speech signal processing algorithms for high-accuracy classification of parkinson's disease," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 5, pp. 1264–1271, 2012.

[16] M. Vallejo, J. Cosgrove, J. E. Alty, S. Jamieson, S. L. Smith, D. W. Corne, and M. A. Lones, "A multi-objective approach to predicting motor and cognitive deficit in parkinson's disease patients," in *Proceedings of the 2016 on Genetic and Evolutionary Computation Conference Companion*. ACM, 2016, pp. 1369–1376.

[17] J. Dalrymple-Alford, M. MacAskill, C. Nakas, L. Livingston, C. Graham, G. Crucian, T. Melzer, J. Kirwan, R. Keenan, S. Wells *et al.*, "The MoCA: well-suited screen for cognitive impairment in Parkinson disease," *Neurology*, vol. 75, no. 19, pp. 1717–1725, 2010.

[18] J. Shi and C. Tomasi, "Good features to track," in *Computer Vision and Pattern Recognition, 1994. Proceedings CVPR'94., 1994 IEEE Computer Society Conference on*. IEEE, 1994, pp. 593–600.

[19] J. Knowles and D. Corne, "The pareto archived evolution strategy: A new baseline algorithm for pareto multiobjective optimisation," in *Proceedings of the 1999 Congress on Evolutionary Computation (CEC)*, vol. 1. IEEE, 1999.

[20] T. W. Anderson, "The choice of the degree of a polynomial regression as a multiple decision problem," *The Annals of Mathematical Statistics*, pp. 255–265, 1962.

[21] B. Stojanovic, M. Milivojevic, M. Ivanovic, N. Milivojevic, and D. Divac, "Adaptive system for dam behavior modeling based on linear regression and genetic algorithms," *Advances in Engineering Software*, vol. 65, pp. 182–190, 2013.

[22] K. Phiwhorm and S. Arch-int, "Ldl-cholesterol levels measurement using hybrid genetic algorithm and multiple linear regression," in *International Conference on Information Science and Applications (ICISA)*. IEEE, 2013, pp. 1–4.

[23] T. Hastie, R. Tibshirani, J. Friedman, and J. Franklin, "The elements of statistical learning: data mining, inference and prediction," *The Mathematical Intelligencer*, vol. 27, no. 2, pp. 83–85, 2005.

[24] E. Zitzler and L. Thiele, "Multiobjective optimization using evolutionary algorithms- a comparative case study," in *Conference on Parallel problem solving from nature (PPSN V)*. Springer, 1998, pp. 292–301.

[25] B. M. Adhikari, K. M. Quinn, and M. Dhamala, "Is the brain's inertia for motor movements different for acceleration and deceleration?" *PloS one*, vol. 8, no. 10, p. e78055, 2013.

[26] M. A. Lones, J. E. Alty, S. E. Lacy, D. S. Jamieson, K. L. Possin, 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*. IEEE, 2013, pp. 76–82.