# A Multi-Objective Approach to Predicting Motor and Cognitive Deficit in Parkinson's Disease Patients

Marta Vallejo
Dept. of Computer Science
Heriot-Watt University
Edinburgh, UK

Jeremy Cosgrove
Department of Neurology
Leeds General Infirmary
Leeds, UK

Jane E. Alty
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

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

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

## ABSTRACT

Parkinson's disease (PD) is a chronic neurodegenerative condition. Traditionally categorised as a movement disorder, nowadays it is recognised that PD can also lead to significant cognitive dysfunction including, in many cases, full-blown dementia. Due to the wide range of symptoms, including significant overlap with other neurodegenerative conditions, both diagnosis and prognosis remain challenging. In this paper, we describe our use of a multi-objective evolutionary algorithm to explore trade-offs between polynomial regression models that predict different clinical measures, with the aim of identifying features that are most indicative of motor and cognitive PD variants. Our initial results are promising, showing that polynomial regression models are able to predict clinical measures with good accuracy, and that suitable predictive features can be identified.

## 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

## Keywords

Multi-objective evolutionary algorithms, Predictive modelling, Parkinson's disease, Polynomial regression

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



*GECCO'16 Companion, July 20-24, 2016, Denver, CO, USA*



DOI: http://dx.doi.org/10.1145/2908961.2931731

## 1. INTRODUCTION

Parkinson's disease (PD) is a chronic progressive neurodegenerative disease with a high incidence. Though it mostly affects the elderly, it also occurs in younger people. The disease leads to neural cell death, and affects various regions of the brain. Notably, loss of dopamine-producing neurons in the *substantia nigra* region of the brain results in movement disorders such as tremor, slowing of movement, and unstable gait. However, it is increasingly recognised that cognitive dysfunction is also prevalent, and this is an important focus of contemporary research [4, 19]. Where cognitive dysfunction precedes the movement disorder, PD is often known as Lewy Body Dementia (LBD). When it appears later, it is often referred to as Parkinson's with Dementia (PDD). Cases with milder symptoms are known as PD with Mild Cognitive Impairment (PD-MCI). However, these labels do not necessarily reflect differences in the disease process, and are not necessarily exhaustive; it is likely that the true disease ontology is more complex, with PD comprising a group of diseases with overlapping cognitive and motor symptoms [1].

Previous work with evolutionary algorithms has focused on their use as predictive modelling techniques to diagnose whether a patient has PD. This approach has been successful, with individual evolved classifiers obtaining test set accuracies in excess of 90% in one study when predicting the presence of PD from time series recordings of a subject's movements [17]. By analysing evolved classifiers, it is also possible to obtain insight into the motor symptoms that are most indicative of PD, thereby informing clinical diagnosis more generally [16].

In this work, by comparison, we are looking at how multi-objective evolutionary algorithms (MOEAs) can help us to better understand PD ontologies. Rather than focussing on a single predictive model, MOEAs allow us to explore a space of predictive models that make different trade-offs. In this paper, we focus on the trade-off between models that predict motor and cognitive elements of PD, and how this might guide clinical practice when diagnosing and prognos-

ing different PD variants. In particular, we describe how a MOEA is used to optimise a Pareto front of polynomial regression models that aim to predict motor and cognitive scores using the best subset of features extracted from clinical assessment figures drawn by patients with different variants of PD.

## 2. RELATED WORK

Over the past decade, there has been a steady increase in the study of cognitive decline in PD [4]. A number of these studies have considered the utility of biomarkers in predicting the onset of both dementia and milder symptoms, identifying a range of different motor and non-motor symptoms, brain imaging features, and biochemical markers that have predictive power [11, 18]. A smaller number of studies have considered the predictive strength of multiple markers [7]. However, to our knowledge there have not been any previous applications of data mining and predictive modelling techniques to cognitive facets of PD.

By comparison, there have been numerous studies of the use of predictive modelling in the diagnosis of Parkinson's disease more generally. Many of these focus on a set of speech recordings available from the UCI machine learning repository [15]. Other studies, particularly work using evolutionary algorithms, have involved collection of clinical data in collaboration with neurologists [17]. Within this context, a significant issue with many predictive modelling techniques, and particularly those involving computational intelligence methods, is their interpretability. Some work has been done to address this [16].

In this work we evolve polynomial regression models to predict clinical measures. This is both to improve interpretability and to make our results more accessible to clinicians, who are unlikely to be familiar with more complex models such as the GP expression trees we used in previous work. There has been past work on hybridising polynomial regression and evolutionary algorithms, including [3] where the EA is used to evolve the model's coefficients, and [14] where the authors also use a multi-objective approach.

We use data from a clinical study in which subjects trace an image of a pentagon spiral (see Section 3.1). There has been some previous work in applying evolutionary computation techniques to pentagon spiral images [21]. Notable differences in the current work are a much larger data set, the extraction of clinically relevant features, the use of relatively simple, interpretable predictive models, a focus on regression rather than classification, and the use of MOEAs.

The use of MOEAs to explore trade-offs between objectives is a well-trodden path. However, there has been relatively little work on extracting more general insights by studying the evolved populations of MOEAs. A prominent example is the work of Deb et al. [6], who describe how analysis of MOEA populations can be used to understand broader design principles regarding decision variables and objectives.

## 3. MATERIALS AND METHODS

### 3.1 Clinical Data

Data collection took place at the Leeds Teaching Hospitals NHS Trust[1]. Fifty-eight patients and twenty-nine age-matched controls were recruited and underwent standard clinical assessments of their motor and cognitive abilities. The main exclusion criteria for controls were drug-induced parkinsonism, multisystem atrophy syndromes, Alzheimer's disease, vascular dementia and combined degenerative and vascular dementias, and significant impairment of upper limb function or visual acuity.

In this paper, we use composite scores from the MoCA (Montreal Cognitive Assessment), a cognitive screening test that has previously been used to assess cognitive impairment in PD [5], and from the motor section of the MDS-UPDRS (Movement Disorder Society sponsored revision of the Unified Parkinson's Disease Rating Scale) assessment. MoCA scores are between 0 and 30, with a value below 26 indicating impaired cognitive function. UPDRS motor scores are between 0 and 132, based on values between 0 (normal) and 4 (severe) over 33 measures. Amongst the PD patients, 22 were assessed as having PD with normal cognition, 26 had PD-MCI, and 10 had PDD. Of the 29 controls, 19 had normal cognition, and 10 had impaired cognition.

Following these standard assessments, the subjects were asked to carry out a series of figure drawing tasks, which are designed to identify particular motor and cognitive impairments. In this paper, we analyse data from a single drawing task which required the subject to trace, using an inking pen, a pentagon spiral figure (see Figure 1) that was overlaid on a Wacom digitising tablet. The subjects' movements were collected as a time series of pen locations within the tablet's frame of reference, sampled 200 times per second, allowing velocity and acceleration time series to be readily generated. Information about pen pressure and inclination was also recorded at each time step.

Pentagon spiral tracing is a relatively simple drawing task, yet it is able to highlight both motor and cognitive dysfunction. In particular, it requires the subject to repeatedly speed up and slow down, then change direction. This can highlight both rigidity and bradykinesia, both cardinal motor dysfunctions in PD. The drawing of straight lines also highlights tremor, another common motor symptom. In the cognitive domain, the task highlights impairments within visuo-spatial reasoning and task planning, two of the major faculties that are affected by cognitive variants of PD.

### 3.2 Features

Table 1 lists the features that are available for each drawing. Features 6–17 were extracted programmatically from the tablet recordings. This first involved aligning the subject's drawing against the spiral pentagon template, which was done by identifying the corner points (using gradient information and the angle between line segments) and associating each of these with the nearest corner point in the template. Figure 1 shows an example of a drawing aligned to the template, with the corner points marked by small circles.

The extracted features capture a range of different motor and cognitive functions. Total time is an indication of a subject's drawing speed; this is generally slower in PD patients than age-matched controls, and is also likely to be affected by the patient's cognitive state. Area error is primarily an indication of the accuracy of a subject's visuo-spatial rea-

[1]Permission to use this data was granted by the South Central - Oxford C NHS REC (ref: 15/SC/0365). Other data generated during this research is available at the following DOI: 10.17861/958af07e-d336-4202-854e-12188211873a.

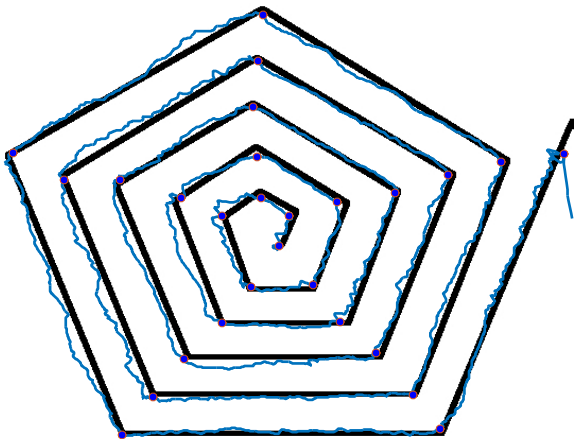

**Figure 1: Example of a patient's spiral pentagon trace overlaid on the template.**

soning, but may also be affected by motor dysfunction. Distance travelled is likely to increase if the subject is experiencing tremor. Whilst subjects are asked not to lift the pen, many do, and the number of times they lift the pen may be an indication of executive dysfunction. Zero velocity indicated hesitation, which may also be a sign of executive dysfunction (e.g. task planning). Unlike healthy controls, PD patients are known to exhibit multiple acceleration bursts when carrying out movements; they are consequently likely to have smaller periods of zero acceleration.

The features discussed in the previous paragraph all summarise an entire drawing. Initial analysis of the recordings, however, indicated that patients exhibit certain dysfunctions in only parts of the drawing, particularly the different radial triangular segments of the image. To capture this, we also extracted the total time (a mixed indicator of motor and cognitive function) separately for each of these segments. These are recorded as features 8–12.

**Table 1: Features used in the developed approach.**

| ID | Feature |
|---|---|
| 1 | Dominant or non-dominant hand |
| 2 | Side of the illness |
| 3 | Number of the attempt the patient is performing the task |
| 4 | Patient/Control |
| 5 | Disease duration |
| 6 | Total time: time taken to complete one spiral pentagon task |
| 7 | Area error: Area created between the template pentagon and the pentagon drawn by the patient |
| 8-12 | Time spent in each triangle formed within the pentagon. Vector of five values |
| 13 | Total distance travelled by the pen |
| 14 | Times the pen leave the tablet surface: patients are instructed to not remove the pen from the tablet |
| 15 | Total time the pen was in contact with the tablet |
| 16 | Duration of zero velocity during whole task |
| 17 | Duration of zero acceleration during whole task |

## 3.3 Polynomial Regression

In regression modelling, the main goal is aimed at learning from a finite number of training data an unknown real-valued function in such a way that it has a high level of generalisation in its prediction capabilities when it is applied to a given test data [2, 8]. In such problems it is common the use of *dictionary methods* where basis functions are represented [9]. The most widely used approach to select the function expansions is the non-adaptive strategy in which the model always includes a predetermined set of basis functions that were not adapted from the training data. Contrary to this approach, an adaptive modelling method [12] fits the basis functions to the data by means of a search mechanism which implies that the degree is added as a new parameter to fit. However, in this polynomial regression problem the number of candidate functions presented in the dictionary will be exponentially increased [2, 8].

However, the power of the function cannot be extended infinitely. A common problem of considering high complex models is that they tend to overfit the data and the regression curve may oscillate wildly between data points. These effects normally induce large prediction errors [10]. In order to find a trade-off between not overfitting and capturing well enough the relationships within the data to achieve a good predictive performance, the most popular approach is to reduce the number of explanatory variables to a subset that potentially could provide good results. The search within this subset is performed by a combinatorial optimisation process where an exhaustive search is in fact impractical [20].

Simple heuristics were proposed to reduce computational time like forward selection [8, 20] or beam search, selected by Todorovski et al. [23] within his CIPER (Constrained Induction of Polynomial Equations for Regression) algorithm that learns polynomial algebraic equations from data. However, these approaches are not powerful enough to cope with problems of large dimensionality. Even facing a problem with limited complexity, the content of this subset is normally not known in advance and due to that a previous search sometimes is required to find the structure of the model before the main learning process starts, which may not be trivial.

A linear regression model can be defined by a linear expansion of functions as follows:

$$F(x) = \sum_{i=1}^{k} b_i f_i(x) \tag{1}$$

where $b = (b_1, b_2 \ldots, b_k)^T$ is a vector of predictors that are defined as parameters of the model, $k$ is the number of basis functions and parameters and finally $f_i(x)$, $i = 1, 2 \ldots, k$ are the basis functions of input $x$. The evolution towards more complex models can be defined by the nonlinear transformations of $x$ where the linearity in the constant parameters is maintained. As a result every polynomial is composed by a set of individual coefficients and degrees and standard linear regression methods can still be used to estimate the parameters efficiently. The general formulation for polynomial regression can be depicted as:

$$Z = \alpha + \sum_{i=1}^{m} \beta_i \cdot T_i \tag{2}$$

where $Z$ is the dependent term called the response vari-

able, $T = \prod_{j=1}^{n} x_j^{a_{i,j}}$ is a vector of the predictors, $\beta = (\beta_1, \beta_2 \ldots, \beta_N)$ is a vector of unknown coefficients where $\beta_i \neq 0$, $\alpha$ is a constant and $m$ is the degree of the polynomial. Assuming continuous derivatives of the regression function at points $T$, local polynomial fitting permits estimating the parameter vector $\beta$.

If the degree of a polynomial $dg$ is defined by the maximum degree of the terms $T$ in that polynomial $Z$, then $dg(Z) = \max_{i=1}^{r} dg(T_i)$ and the length of a polynomial $len$ is the sum of the degrees of all terms represented in that polynomial $len(Z) = \sum_{i=1}^{r} dg(T_i)$, then for a polynomial of degree $d$ and length $l$, the total number of possible basis functions within the dictionary is:

$$numFun = \prod_{i=1}^{l} (1 + \frac{d}{i}) \quad (3)$$

## 3.4 Pareto Archived Evolution Strategy

Pareto Archived Evolution Strategy (PAES) [13] is a simple and effective multi-objective optimisation algorithm that combines the use of local search techniques with a hill-climbing and random mutation strategy. In its original variant called (1+1)-evolution strategy, the algorithm uses a unique-parent and a single-offspring that are compared in each iteration. The calculation of the quality of new candidate solutions is supported by means of the information provided towards a set of diverse non-dominated solutions stored in the archive. The creation of the offspring is generated by the use of binary strings and a unique bitwise mutation operator which is compared with its single parent. This latter factor differentiates this approach from other MOEAs that maintain a population of solutions. The strategy uses an archive of previously visited non-dominated solutions to estimate the dominance rating of the new solution. A maximum size of the archive refers to the desired number of final solutions. Based on that structure, the algorithm is capable of distinguishing between good and bad quality solutions.

PAES is an algorithm with a performance profile that is competitive with that of more recent algorithms, especially on larger-scale problems under the constraint of a reduced time budget. See Algorithm 1. In our approach, we run PAES multiple times from different initial points in the search space, maintaining a solution archive that is shared between subsequent executions of the algorithm. This permits a relatively broad exploration of the search space.

## 3.5 The Search Algorithm

A MOEA can be applied to the task of finding which predictor variables are informative in regard to different objectives. In this case, we consider the two objectives of predicting motor and cognitive deficit in Parkinson's Disease Patients, as measured by UDPRS and MoCA scores. The algorithm uses as input a dataset that contains information about independent and dependent variables which defined possible polynomial structure. The set of possible features have been extracted from clinical test data collected by Leeds Teaching Hospitals NHS Trust which are depicted in Table 1.

The type of MOEA algorithm applied in this paper is a modified version of $(1 + 1)$ PAES algorithm where a single chromosome is evolved. In each generation, a polynomial regression algorithm is used to calculate the fitness function

**Algorithm 1** PAES

**Require:** Max_iter
  Generate int_sol and set it as Current_sol
  Evaluate objective 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

by estimating the error of fitting this given polynomial to the UDPRS and MoCA scores assigned to the patients. Each polynomial equation included in the population of solutions is structured as Eq. 2.

It is not reasonable to consider the entire search space of all possible polynomials since the use of high-level complex functions normally entails the overfitting of data and could also have additional effects in which the model may even become numerically unstable. To control these effects a beforehand bound of the dictionary of basis functions should be performed. According to Equation 3, the number of possible subsets from a dictionary of size $m$ is $2^m$. Each of these subsets can be a candidate for our subset of selections. Since most of the times the complete revision of all of these possibilities are unhandleable due to its complexity, a maximum boundary for the length and the degree of the model has been selected, with values of $d = 8$ and $l = 8$. Other non-population based algorithms commonly start the search with the simplest possible equation, that is a constant term that is developed towards more complex structures by a process of adding new basis functions [23]. By means of an iterative process the application of the refinement operator can generate all the possible candidate polynomial equations.

The evolutionary algorithm PAES starts with the generation of a single individual solution. Every time the algorithm finds a good non-dominated solution, this solution is stored in the archive that for our purpose is defined with unlimited size. The archive constitutes the current approximation of the Pareto front. Each of these chromosomes

are constructed by the random generation of two different vectors represented as follows:

$$f = (f_1, f_2 \ldots, f_8), f_i \in \{1 - 17\} \tag{4}$$

$$a = (a_1, a_2 \ldots, a_8), a_i \in \{1 - 8\} \tag{5}$$

where $f$ is a vector of features and $a$ is the vector of exponents/powers of the predictors. Both elements depict the internal structure of a given polynomial.

In order to evaluate the suitability of a determined equation, the values of these generic constants have to be fitted against the data, using of a subset of the collected data selected for the training process. A different training subset is generated every time the polynomial is assessed.

The quality of the obtained equation is evaluated using the fitness function. This function returns a pair of values, each of them measuring the discrepancy between the assigned values of UDPRS and MoCA scores and the values predicted using the equation respectively. The estimation of the ability of the selected parameters to fit the data is performed using the Ordinary Least-Squares (OLS) method [8, 22] in which the function is aimed at minimizing the residual sum of squares. This is a typical metric applied to linear regression problems. The calculation can be mathematically defined as follows:

$$b = \arg\min_{b} \sum_{j=1}^{n} \Big(y_j - F(x_j)\Big)^2 \tag{6}$$

where $y_j$ is the response variable for the i-th training example, $x_j$ is the value predicted for the same training example using the equation, $n$ is the total number of data points and $F$ is defined in Equation 1.

Two mutation operators are used. One targets the power $a_i$ of one of the terms $f_i$ of the polynomial, which it randomly increases or decreases. To choose the same previous value is not permitted. This allows the modification of the level of complexity of the equation. The second perturbs the value of one the features of the equation $f_i$, selecting randomly another one. In each generation, an elitist mechanism tries to refine the polynomial equation to evolve to a more effective area of the search space. Special care is taken that the newly introduced term is different from the previous one and no redundancy is stored in the archive.

The stopping criterion is defined by a determined number of times that a given solution is evolved. This number could vary from 100 to 500. Since the archive is shared between runs, new non-dominated solutions are added to the structure every time the algorithm finds one in its different runs.

The outcome of the algorithm consists of the best non-dominated polynomial equations that have been found which are allocated in the archive for each of the different exponent values considered.

A solution can be depicted by two vectors: a vector $\beta$ of predictors:

$$\beta = (\beta_0, \beta_1 \ldots, \beta_8) \tag{7}$$

where $\beta_0$ is the independent term and a vector of residual errors which contains one value for each score considered.

# 4. RESULTS

Regression performance is estimated using a set of unseen examples using 10-fold cross-validation. Overall performance is measured by calculating the hypervolume [24] derived from the estimated Pareto front. This indicator retrieves information related to the convergence and the spread of the set of non-dominated solutions that belong to the known Pareto front. The mathematical formulation of hypervolume (Hv) can be depicted as follows:

$$Hv(P) = \lambda\Big( \cup_{p \in P}\, x | p \prec x \prec r \Big) \tag{8}$$

where $\lambda$ defines the Lebesgue measure, $P$ is the Pareto approximation to the real Pareto front and $p$ and $r$ represent the utopia and anti-utopia points used as a reference. In our analysis setup, lower values of hypervolume correspond to better outcomes which indicate that the approximate $P$ is closer to the true Pareto front.

To begin with, we considered the trade-off between model complexity and predictive ability, by running the algorithm with different limits placed on the maximum exponent until no new solutions have been found in ten consecutive runs for each different power considered. Figure 2 shows the hypervolumes for exponent limits between 1 (a linear model) and 8. This suggests that a linear regression model would not be sufficient to predict the clinical measures, and justifies our use of a polynomial model. However, it is not clear from this plot which is the optimal upper limit for model complexity.

Figure 3 depicts the Pareto front (in red) created by combining the archives from all the runs summarised in Figure 2. It can be seen that the polynomial regression models in both tails of the front are able to generate robust predictions for either motor and cognitive scores. This indicates that polynomial regression models are sufficiently expressive for this task, and that PAES is effective at searching for appropriate model instances. The shape of the front shows a clear trade-off, with models around the mid-point having fairly poor predictive ability for both regression targets. Table 2 lists the solutions in this Pareto front. Figure 3 also shows the Pareto front (in blue) when only linear models are used; this again highlights the relatively poor predictive ability of linear models on this task, though it is notable that the shape of the front is broadly similar to the higher-order polynomial case.

Analysis of the features used by the evolved models gives some insight into the utility of different features for predicting motor and cognitive deficit. In general, the most used features for both objectives are do to with the time spent drawing and the area error, i.e. speed and accuracy. The model with the smallest residual error for UPDRS scores is:

$$\begin{aligned} y = {} & 56.1 f_9^3 + 16.1 f_{12}^3 + 0.4 f_{10}^4 + 3.0 f_6^5 \\ & + 0.3 f_7^5 + 9.0 f_{11}^5 - 120.2 f_{12}^5 - 17.5 f_{13}^5 \end{aligned} \tag{9}$$

and the model for the smallest MoCA residual error is:

$$\begin{aligned} y = {} & -1.4 f_3^2 - 0.038 f_7^3 + 0.68 f_9^4 - 7.9 f_6^5 \\ & + 7.5 f_9^6 + 24.0 f_{12}^7 + 0.22 f_{13}^7 + 0.71 f_{15}^7 \end{aligned} \tag{10}$$

Further analysis is required to fully understand the medical significance of these results. However, the feature usage

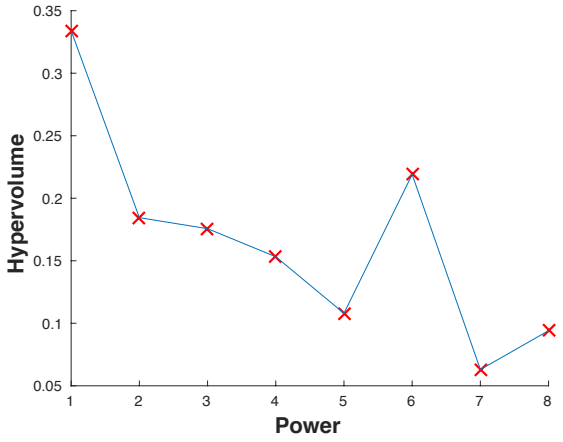


**Figure 2: Relationship between regression error and model complexity, showing the hypervolume for each of the exponents considered.**

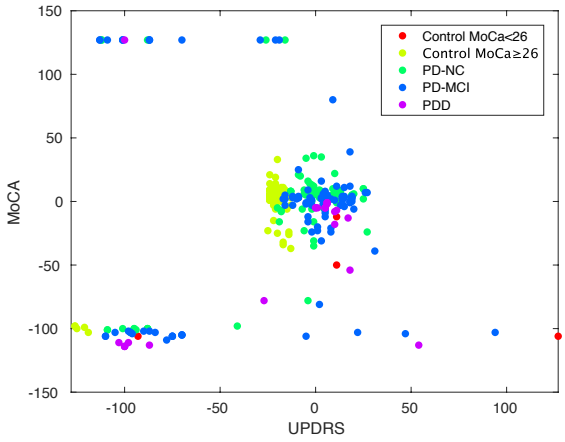


**Figure 4: Scatter plot showing regression errors for each objective for each subject, using the two polynomial models with the best overall fit to MoCA and UPDRS scores. Colours indicate disease classes.**

seems to confirm existing knowledge. For example, slowing of movement (indicated by time spent drawing, particularly over a series of repeats) is closely associated with the progress of the motor disorder, and impoverished visuospatial reasoning (indicated by area error) is associated with cognitive decline. In future work, we aim to extract a much larger group of features, particularly features within a drawing cycle, such as velocity and acceleration around corners. We will also consider a wider range of objectives, including other predictive variables (such as accuracy early in the disease, which is more useful for diagnosis), and non-predictive factors such as minimising the complexity of the model (e.g. number of features).

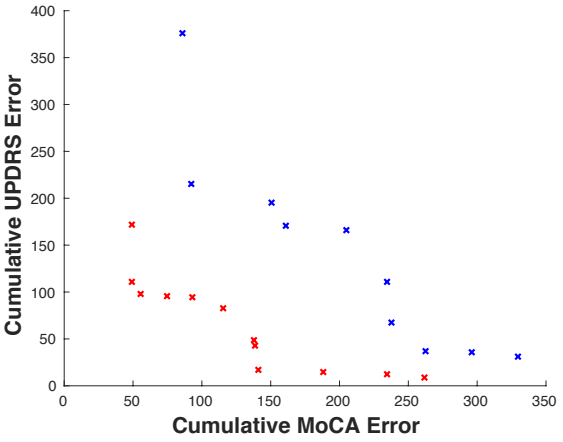


**Figure 3: Non-dominated sets of solutions. Red points show the Pareto front for all polynomial regression models, blue points show the Pareto front for linear models.**

Further analysis is required to fully understand the medical significance of these findings. However, some of these observations seem to confirm existing knowledge. For example, slowing of movement (indicated by time spent drawing, particularly over a series of repeats) is closely associated with the progress of the motor disorder, and impoverished visuospatial reasoning (indicated by area error) is associated with cognitive decline. Hence, the approach seems valid. In future work, we aim to extract a much larger group of features, particularly features within a drawing cycle, such as velocity and acceleration around corners. We will also consider a wider range of objectives, including other predictive variables (such as accuracy early in the disease, which is more useful for diagnosis), and non-predictive factors such as minimising the complexity of the model (e.g. number of features).

It is interesting to consider whether the evolved models have uniform error rates across the different patient and control groups. To test this, Figure 4 shows the regression errors for the motor and cognitive measures for each subject, using the models from both ends of the Pareto front. In general, it appears that the predictive behaviour of the models is not uniform. The motor ability of controls with normal cognition is consistently under-estimated, as indicated by the left-offset of the olive green group in the figure. The motor ability of Parkinson's patients with normal cognition is, by comparison, predicted quite accurately, with errors distributed fairly equally around the origin.

The cognitive ability of controls and patients with normal cognition is also predicted fairly well. However, there is significantly more variance in the predictions of the cognitive abilities of subjects with cognitive impairment. Notably, the cognitive scores of Parkinson's patients with dementia are generally under-estimated.

These differences may in part reflect proportionally smaller numbers of subjects with low MoCA scores (dementia) and high UPDRS scores (non-PD) in the data set, and a consequent over-fitting of the models to certain parts of the ranges. However, they may also indicate real differences between the groups. Another interesting facet is the clustering of subjects at the top and bottom of the MoCA error range, seen as horizontal lines in the scatter plot. For these subjects, there seems to be a fairly consistent under- or over-prediction. Again, it is unclear whether this is due to properties of the model, or whether it points to more sig-

nificant differences within the subject population. We plan to explore this further in future investigation.

## 5. CONCLUSIONS

Parkinson's is a complex disease that leads to different symptoms in different patients. Most patients experience impaired motor function. However, many patients also experience cognitive dysfunction, and a significant proportion of these patients go on to develop dementia. The ontology of Parkinson's disease is only partially understood, and it is unclear which symptoms are of diagnostic significance at different stages of the disease, and whether a particular patient will go on to develop cognitive symptoms. In this paper, we have taken an initial look at whether multi-objective evolutionary algorithms can be used to better understand trade-offs when diagnosing and prognosing Parkinson's. In particular, we focused on the trade-off between predicting motor and cognitive symptoms using regression models trained on objective features extracted from a clinical assessment task. The results are promising, showing that both motor and cognitive features can be predicted fairly well. We chose to use a relatively simple regression model in order to improve interpretability. Analysis of the trained models shows that different groups of features are particularly relevant to assessment of motor and cognitive signs, and a number of these match with existing clinical knowledge. In future work we will consider a much broader range of features, and also carry out a more in depth analysis.

## 6. ACKNOWLEDGEMENTS

This research was supported by the EPSRC through the project "A Multiobjective Evolutionary Approach to Understanding Parkinson's Disease" (Ref: EP/M013677/1).

## 7. REFERENCES

[1] D. Berg, A. E. Lang, R. B. Postuma, W. Maetzler, G. Deuschl, T. Gasser, A. Siderowf, A. H. Schapira, W. Oertel, J. A. Obeso, et al. Changing the research criteria for the diagnosis of Parkinson's disease: obstacles and opportunities. *The Lancet Neurology*, 12(5):514–524, 2013.

[2] V. Cherkassky and F. M. Mulier. *Learning from data: concepts, theory, and methods*. John Wiley & Sons, 2007.

[3] J. Clegg, J. F. Dawson, S. J. Porter, and M. H. Barley. The use of a genetic algorithm to optimize the functional form of a multi-dimensional polynomial fit to experimental data. In *Evolutionary Computation, 2005. The 2005 IEEE Congress on*, volume 1, pages 928–934. IEEE, 2005.

[4] J. Cosgrove, J. E. Alty, and S. Jamieson. Cognitive impairment in Parkinson's disease. *Postgraduate medical journal*, 91(1074):212–220, 2015.

[5] 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*, 75(19):1717–1725, 2010.

[6] K. Deb and A. Srinivasan. Innovization: Discovery of innovative design principles through multiobjective evolutionary optimization. *Multiobjective Problem Solving from Nature: From Concepts to Applications*, page 243, 2007.

[7] K. T. O. Dubbelink, A. Hillebrand, J. W. Twisk, J. B. Deijen, D. Stoffers, B. A. Schmand, C. J. Stam, and H. W. Berendse. Predicting dementia in Parkinson disease by combining neurophysiologic and cognitive markers. *Neurology*, 82(3):263–270, 2014.

[8] J. Friedman, T. Hastie, and R. Tibshirani. *The elements of statistical learning*. Springer, Berlin, 2001.

[9] J. H. Friedman. *An overview of predictive learning and function approximation*. Springer, 1994.

[10] I. Guyon, S. Gunn, M. Nikravesh, and L. A. Zadeh. *Feature extraction: foundations and applications*, volume 207. Springer, 2008.

[11] M. Hu, K. Szewczyk-Królikowski, P. Tomlinson, K. Nithi, M. Rolinski, C. Murray, K. Talbot, K. P. Ebmeier, C. E. Mackay, and Y. Ben-Shlomo. Predictors of cognitive impairment in an early stage Parkinson's disease cohort. *Movement Disorders*, 29(3):351–359, 2014.

[12] G. Jekabsons. Adaptive basis function construction: an approach for adaptive building of sparse polynomial regression models. In Y. Zhang, editor, *Machine Learning*. INTECH, 2010.

[13] 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)*, volume 1. IEEE, 1999.

[14] D. Laucelli and O. Giustolisi. Scour depth modelling by a multi-objective evolutionary paradigm. *Environmental Modelling & Software*, 26(4):498–509, 2011.

[15] M. A. Little, P. E. McSharry, E. J. Hunter, J. Spielman, and L. O. Ramig. Suitability of dysphonia measurements for telemonitoring of Parkinson's disease. *Biomedical Engineering, IEEE Transactions on*, 56(4):1015–1022, 2009.

[16] M. Lones, J. E. Alty, S. E. Lacy, D. Jamieson, K. L. Possin, N. Schuff, S. L. Smith, et al. 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, 2013.

[17] M. Lones, S. L. Smith, J. E. Alty, S. E. Lacy, K. L. Possin, D. Jamieson, A. M. Tyrrell, et al. Evolving classifiers to recognize the movement characteristics of Parkinson's disease patients. *Evolutionary Computation, IEEE Transactions on*, 18(4):559–576, 2014.

[18] B. Mollenhauer, L. Rochester, A. Chen-Plotkin, and D. Brooks. What can biomarkers tell us about cognition in Parkinson's disease? *Movement Disorders*, 29(5):622–633, 2014.

[19] J. Pagonabarraga and J. Kulisevsky. Cognitive impairment and dementia in Parkinson's disease. *Neurobiology of disease*, 46(3):590–596, 2012.

[20] J. Reunanen. Search strategies. In *Feature Extraction*, pages 119–136. Springer, 2006.

[21] 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*

| | Power | MoCA | UPDRS | Features | Power |
|---|---|---|---|---|---|
| **1** | 7 | 261.669 | 86.37 | 7, 16, 12, 11, 6, 13, 7, 7 | 7, 7, 5, 6, 6, 7, 7, 3 |
| **2** | 5 | 234.812 | 12.029 | 12, 9, 11, 6, 12, 7, 13, 10 | 3, 3, 5, 5, 5, 5, 5, 4 |
| **3** | 5 | 141.165 | 16.881 | 12, 9, 6, 6, 12, 11, 13, 10 | 3, 3, 5, 5, 5, 5, 5, 4 |
| **4** | 7 | 188.139 | 14.546 | 7, 16, 12, 11, 6, 13, 7, 7 | 7, 7, 5, 6, 3, 6, 7, 3 |
| **5** | 6 | 138.869 | 43.16 | 6, 5, 12, 12, 11, 7, 16, 7 | 5, 5, 6, 4, 6, 6, 6, 5 |
| **6** | 2 | 138.031 | 49.216 | 12, 15, 6, 8, 12, 17, 16, 16 | 1, 2, 2, 1, 1, 1, 2, 1 |
| **7** | 6 | 115.074 | 82.795 | 6, 5, 12, 12, 12, 7, 16, 7 | 5, 5, 6, 4, 6, 6, 6, 5 |
| **8** | 7 | 93.073 | 93.888 | 7, 16, 13, 11, 6, 13, 3, 7 | 7, 7, 5, 6, 6, 7, 7, 3 |
| **9** | 7 | 74.324 | 95.174 | 7, 16, 12, 11, 6, 13, 3, 7 | 7, 7, 5, 6, 6, 7, 7, 3 |
| **10** | 7 | 55.477 | 97.556 | 7, 8, 13, 11, 6, 13, 3, 7 | 7, 7, 5, 6, 1, 7, 7, 3 |
| **11** | 7 | 49.515 | 17.2109 | 12, 13, 9, 6, 15, 9, 3, 7 | 7, 7, 4, 5, 7, 6, 2, 3 |
| **12** | 7 | 49.529 | 110.828 | 12, 13, 9, 13, 15, 9, 3, 7 | 7, 7, 4, 5, 7, 6, 2, 3 |

| | Polynom1 | Polynom2 |
|---|---|---|
| **1** | 24.344, 1.477, 1.626, -0.082, -0.687, -0.659, -2.716, 1.477, -2.222 | 20.618, -3.689, -7.413, 1.836, 5.958, 1.493, 9.083, -3.689, 7.619 |
| **2** | 24.407, -1.629, -0.262, 0.098, -0.662, 1.511, 0.281, -0.109, -0.282 | 16.061, 56.12, 8.974, 2.985, 0.211, -120.216, 0.277, -17.487, 0.442 |
| **3** | 24.38, -1.935, -0.16, -0.308, -0.308, 1.81, -0.015, 0.044, -0.143 | 20.319, 32.059, 2.672, 0.75, 0.75, -30.153, -0.903, -1.092, 2.751 |
| **4** | 23.508, 1.338, 1.524, -14.899, -0.631, -0.426, -2.527, 1.338, -1.98 | 21.201, -1.779, -6.404, 1.808, 7.796, 1.657, 5.943, -1.779, 5.609 |
| **5** | 24.117, -0.398, 0.062, 1.243, -1.366, 0.187, 6.408, -0.456, -6.372 | 20.7, 1.422, 2.895, -99.253, 100.954, -0.023, -13.267, 1.016, 12.303 |
| **6** | 24.409, 0.407, -0.342, -0.535, -0.947, 0.407, 1.501, 0.25, -1.292 | 20.267, 0.346, 5.355, 1.731, 4.352, 0.346, -7.081, -2.272, 3.477 |
| **7** | 25.012, 7.157, -0.173, 0.319, -0.717, 0.319, 17.879, 0.038, -15.055 | 20.474, 1.509, 1.268, -53.395, 108.71, -53.395, -16.665, 1.776, 16.084 |
| **8** | 24.324, 2.607, 1.759, -2.796, -0.533, -0.599, 0.122, -1.212, -1.911 | 20.862, 9.17, -5.347, 31.799, 5.424, 0.697, -33.56, 12.354, -2.466 |
| **9** | 24.291, 1.197, 1.3, -0.058, -0.673, -1.725, -1.549, -1.277, -1.04 | 20.112, 2.672, -4.889, 1.584, 4.445, 0.802, 2.514, 11.667, -0.318 |
| **10** | 24.17, 2.746, -0.377, -12.657, 0.04, -0.567, 10.823, -1.589, -1.086 | 20.397, 1.213, 0.344, 17.215, 2.098, 1.181, -17.575, 11.943, -0.205 |
| **11** | 24.017, 0.216, 0.677, -7.926, -0.392, 0.713, 7.455, -1.353, -0.038 | 22.584, 40.264, -0.329, 4.642, 0.678, 0.64, -3.153, 11.959, 0.317 |
| **12** | 24.26, -1.319, 2.176, -10.821, -1.491, 0.469, 10.317, -1.336, 0.473 | 20.102, 0.934, -15.343, 16.881, 14.652, 0.607, -15.007, 11.753, -1.007 |

**Table 2: Non-dominated polynomial regression models, showing cumulative errors for each regression target, features used, with corresponding powers, and the coefficients that are fitted to the model for each of the regression targets.**


*Programming and Evolvable Machines*, 8(4):433–447, 2007.

[22] D. Stronger and P. Stone. Polynomial regression with automated degree: A function approximator for autonomous agents. *International Journal on Artificial Intelligence Tools*, 17(01):159–174, 2008.

[23] L. Todorovski, P. Ljubič, and S. Džeroski. Inducing polynomial equations for regression. In *Machine Learning: ECML 2004*. Springer, 2004.

[24] E. Zitzler and L. Thiele. Multiobjective optimization using evolutionary algorithms - a comparative case study. In *Parallel problem solving from nature PPSN V*, pages 292–301. Springer, 1998.