Abstract

Around four million people are affected by Parkinson's disease (PD) around the world and this number is expected to increase over the following years. PD is a neurodegenerative disorder that causes a large number of symptoms that seriously affect the lives of patients and their families. The reliability of current diagnosing and monitoring methods depend on the examiner, so better tools to objectively evaluate the presence and severity of the disease and its symptoms are needed.

This project is part of a large ongoing research. This research focused on applying traditional data mining and machine learning tools using R's data mining library to help diagnosing, monitoring and understanding the features of PD symptoms from a dataset. A dataset is a collection of information collected in an experiment where more than 100 people, PD patients and healthy control subjects, were involved. This participants had to draw a picture and data was collected from this task. In this report, the raw dataset will be processed to find velocities of sub all subject's drawing, then information will be extracted from the data using data mining and machine learning. The results from different trained classifiers will be explained. The aim of this project is to achieve more than 90% accuracy and comparing performances from various classifiers and.

1. Objective

The objective of this project is to develop ways of expressing and classifying patterns in drawings that help doctors to diagnosing, monitoring and understanding Parkinson's disease and its symptoms. This goal is approached following these steps:

- Find the velocities of the patients' drawings.
- Applying machine learning techniques to the variables to extract features of PD's symptoms.
- Comparing results of various classifier.
- Comparing performances of target images

2. Literature Review

2.1 Introduction

"Parkinson's disease (PD) is a progressive neurological disorder characterised by a large number of motor and non-motor features that can impact on function to a variable degree." (Jankovic, 2008, p. 368)

It is estimated that there are four million people with PD around the world and this number is expected to triple in the next 50 years (Lones, et al., 2013). The incidence in men is 46% higher than in women's. PD affects 84 people in every 100,000 over age 50 in the UK (Horsfall, et al., 2013).
PD causes different symptoms:

- Rest tremor
- Slowness of movement (bradykinesia)
- Rigidity
- Poverty of movement (hypokinesia)
- "Absence or reduced functionality of movements" (akinesia) (NationalCollaboratingCentreFor, 2006).
- Postural instability

Symptoms are not present in all patients and their intensity vary over time, they can disappear and (re-) appear. After a long treatment of the disease, drugs used may cause other symptoms. Dyskinesia is an example of them, which caused parts of the body to be involuntary moved. All these make the symptomatic history for each patient unique (Ahlrichs and Lawo, 2013). It is also known that PD affects more to one side of the body (left or right side) than the other at the onset of the disease. Symptoms are affected by this asymmetry as well (Verreyt, et al., 2011), (NationalCollaboratingCentreFor, 2006).

Currently there are several scales to evaluate the severity of symptoms. The Unified Parkinson's Disease Rating Scale (UPDRS) is the most used one (Jankovic, 2008), (NationalCollaboratingCentreFor, 2006).The reliability of this scale depends on the examiner, it is a subjective way of evaluating the severity of the disease. When used by trained specialists it is reliable, but not when specialists have not enough experience. Tools to make the evaluation objective or to help specialists to measure the severity, the progression and to better understand the disease are needed (Saunders-Pullman, et al., 2008).

*work in progress

2.2 Previous work

Various methods to automatically diagnose PD have already been developed using different Artificial Intelligence techniques. Ahlrichs, C. and Lawo, M.
(2013) analyse some of them. Neural Networks (NNs), Hidden Markov Models (HMMs) and Support Vector Machines (SVMs) have already been

used. Many of them have focused on detecting just one symptom and these authors consider that, as PD is really heterogeneous in terms of symptomatology, they should be focused on multiple motor symptoms. Each research analysed by them was focused on a different symptom. These researches also used different sensors to collect data.

One important observation Ahlrichs, C. and Lawo, M. (2013) make at their conclusion is that datasets used on reviewed works are very different in terms of quantity and quality. Usually authors using small or synthetic datasets obtain higher accuracies than those using larger datasets from real people. They think that even high accuracies have been achieved, there are still improvements to be made.

Miguel (2014) have developed very interesting model to diagnose PD. He combined traditional data mining with Bio-inspired computing technics. Bio-inspired technics are usually use to achieve the best performance from model. His evolved classifiers achieved more than 90% accuracy in almost all classifier. His datasets are a bit outdate so he have to process them first before he use them result in a limitation of his model which has to be change if the images are not the same (It is Pentagon in this case).

*work in progress

3. Methodology

3.1 Dataset

The raw dataset has been obtained by Leeds General Infirmary. It belongs to 41 control subjects and 48 patients. This dataset is considerably larger than most mentioned by Ahlrichs, C. and Lawo, M. (2013). Participants were asked to draw on the digital device follow the target images.

Each subject had to do the task three times with the dominant hand and another three times with the non-dominant hand. Data was recorded using Wacom digitising tablets. For each repetition a text file with collected data is generated.
For every trial (it was recorded with a 197 Hz frequency) it was measured:

- Position of the pen
- Orientation of the pen
- Pressure exerted at that moment

Each file contains as many lines (records) and each record represented

- Timestamp
- X coordinate
- Y coordinate
- Pen orientation altitude
- Pen orientation azimuth
- Pressure

The raw dataset will be processed record by record to find velocity of the drawing of all subjects then the processed dataset will be used to teach classifiers.

3.1.1 The Benson Figure.

In this project, we will be using recordings of patients performing a standard cognitive test, the copy and recall of the Benson complex figure (see Figures: below). This test was developed by UCSF (Possin et al, 2011). It has recently become a standard part of Alzheimer's assessment in the USA, but is equally appropriate for measuring the cognitive skills of patients with PD and other NDDs. Although primarily designed to measure cognitive skills, such as visuospatial awareness and memory, the Benson figure is also well suited to measuring motor skills, since it involves the patient carrying out a variety of different movements. Ordinarily this motor information would be lost; however, by using a high-resolution digitising tablet with an inking stylus, we are able to record detailed information about how the patient moves (e.g., tremor, hesitations, slowing of movement, abnormal acceleration features). Since the patient's movements are recorded as a time indexed series, it is also possible to infer information about task planning, such as shape ordering and pauses. Because the tablet can detect the position of the pen whilst lifted from the paper, potentially valuable information about what the patient is doing between drawing strokes is also recorded. In effect, digitised recordings can be expected to contain motor and cognitive signals that are very relevant to the diagnosis of PD and for its discrimination from other NDDs. Importantly, the use of a standard test and an inking stylus means that the traditional clinical environment can be maintained.

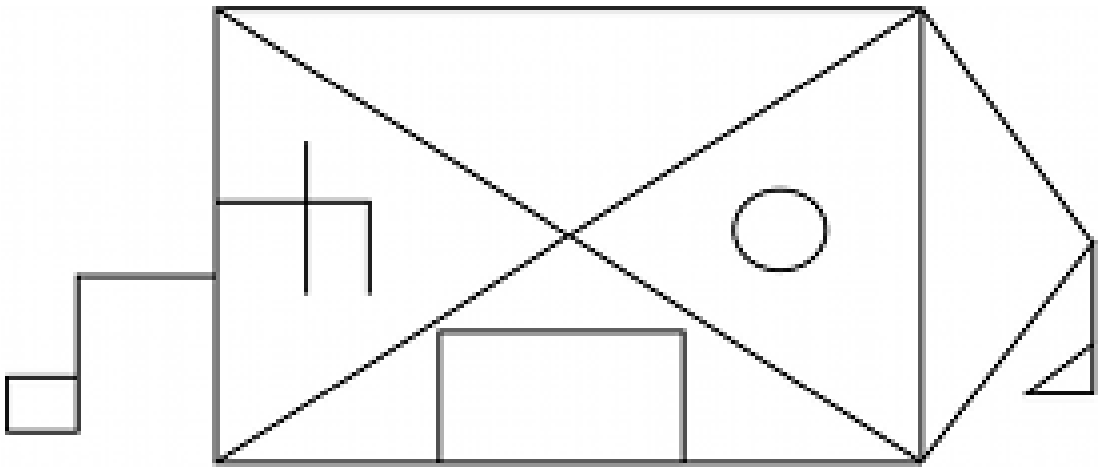

Figure: Original Benson image.

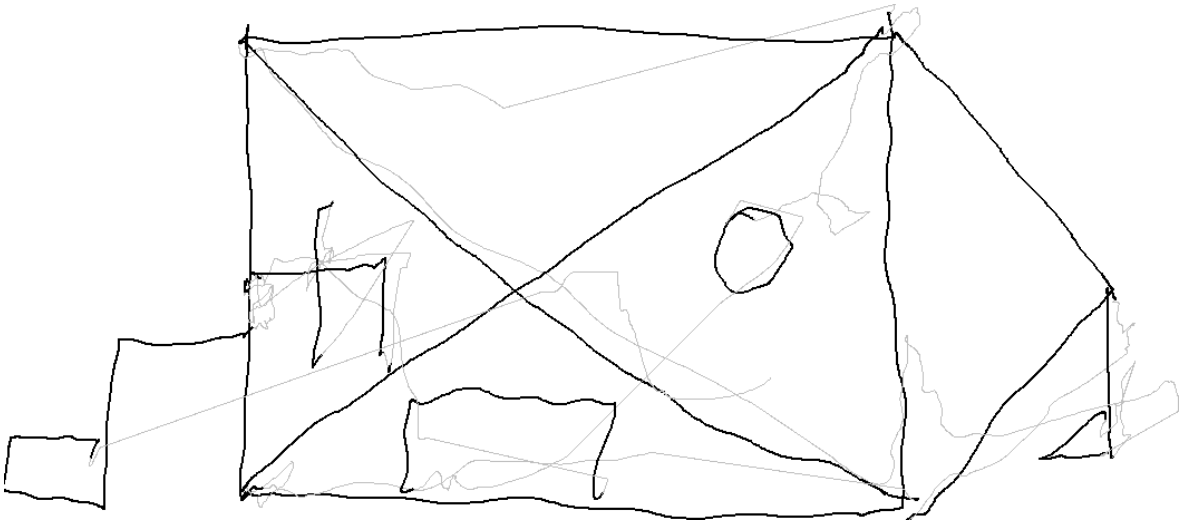

Figure: Benson image drawn by patient.

3.2 Data Mining

"Data mining is defined as the process of discovering patterns in data. The process must be automatic or (more usually) semiautomatic. The patterns discovered must be meaningful in that they lead to some advantage" (Witten and Frank, 2005, pp. 5)

Machine learning consists in developing systems that are able to make predictions after being exposed to a learning process. These systems can

learn in different ways, the most commonly used are (Witten and Frank, 2005), (Russel and Norving, 2014):

- Supervised learning
  A set of input-output pairs are used by the system to learn.

- Unsupervised learning

The system learns just observing the inputs.

- Reinforcement learning

Rewards and punishments are used to learn.

As it is given the output of the data that is going to be used (I know if each piece of data belongs to a patient or a control subject), during the project supervised learning methods are used.

3.2.2 Classifiers

There are different ways of approaching classification problems or, in other words, classifiers can grouped in different categories. Here some of the most relevant ones are listed (Witten and Frank, 2005):

- Decision trees
  They follow the 'divide-and-conquer' approach. Nodes test attributes against constants or against other attributes, and leaf nodes assign a class prediction to instances reaching them. When testing the tree to certain instance, values of the variables from the instance are tested against nodes until a leaf is reached, the class of that leaf is predicted for the instance. Trees are easy to visualise and to interpret.
- Classification rules
  Rules are formed by some preconditions that are tested and a conclusion that is assigns a class if the preconditions are met. Rules can easily be extracted from trees (preconditions would be the nodes and the conclusion a leaf). Rules can also have attached a probability and, when different rules with different conclusions are applied at the same time for the same instance, a final probability of belonging to certain class is provided.
- Instance-based learning
  This method memorises the training set and, when a new instance is tested, the class assigned is the same as the one assigned to the instance that is the most similar from all those that were memorised. There are different ways of calculating which one is the most similar. It is a simply way of doing classification.

- Naïve Bayes
  It is a probabilistic classifier that applies the Bayes' theorem, it assigns a probability. This theorem relates the probability of A knowing B with the probability of B knowing A. Naïve Bayes classifier assumes that both distributions are independent, this is to say, the presence of one feature does not imply the presence of any other one. It is fast and accurate, what makes it a really used technique.

3.2.3 Prediction

*Work in progress

3.2.4 R's data mining package

Rattle package will be used to perform all experiment in this research project. Rattle is R's package aimed to provide multiple sets of functions and classifiers for data mining task.

Rattle is in daily use by Australia's largest team of data miners and by a variety of government and commercial enterprises, worldwide. A number of international consultants also use Rattle in their daily business. Users include the Australian Taxation Office, Australian Department of Immigration, Ulster Bank, Toyota Australia, US Geological Survey, Carat Media Network, Institute of Infection and Immunity of the University Hospital of Wales, US National Institutes of Health, AIMIA Loyalty Marketing, Added Value, and many more. It is or has been used for teaching by the McMaster University, Australian National University, University of Canberra, University of Technology Sydney, Yale University, University of Southern Queensland, Revolution Analytics, Habin Institute of Technology Graduate School Shenzhen, and Many more.

2.2.4.1 List of Classifier that will be used inside this project.

*Work in progress

3.3 Evaluation

When evaluating, classifiers there are different factors that have to be considered. Obtaining high accuracies does not mean that the classifier is working as expected or desired. For example, for a testing dataset where 90% of instances belong to certain class A, always predicting A would give us an accuracy of 90%, but this classifier would probably not work well when exposed to other data, not the one used to test it. Statistical methods

are used to measure confidence bounds in case the test set is not big enough to be statistically significant (Witten and Frank, 2005).

3.3.1 Cross-validation

When applying a k-fold cross-validation, the dataset is split into k different groups of equal size, this division of the data is done randomly. The classifier being used is trained and tested also k times. Each time, the classifier is trained using k-1 of those groups and tested on the remaining one, using each time a different one. The accuracy of the classifier is the average of all k executions (Kohavi, 1995).

This technique is really convenient when the dataset is not big enough. It reduces the possibility of being unlucky and getting a training, or a testing, set not being representative. The standard method is using 10-fold cross-validation as 10 seems to be the best number to get a good estimation error (Witten and Frank, 2005).

3.3.2 Confusion matrix

This type of matrix is used in the machine learning field. Columns represent predicted classes and rows the real classes. When working with two classes, the matrix represents: false positives, false negatives, true positives and true negatives (Witten and Frank, 2005).

		Predicted class		
		PD	not PD	
Actual class	PD	true positive	false negative	sensitivity
	not PD	false positive	true negative	specificity

Figure: Confusion Matrix

Sensitivity and specificity are two relevant values:

- Sensitivity
  In this case, ability to detect people with PD properly.

$$sensitivity = \frac{true\ positives}{true\ positives + false\ negatives}$$

- Specificity
  In this case, ability to identify healthy people.

$$specificity = \frac{true\ negatives}{true\ negatives + false\ positives}$$

4. Requirement Analysis

4.1 Aim and Objective

Due to aging populations around the world, the incidence of neurodegenerative diseases (NDD), such as Parkinson's (PD) and Alzheimer's (AD), is increasing rapidly. Despite this rapidly increasing, the characteristics of these diseases remain incompletely understood. As a consequence, they are difficult to diagnose, with current assessment procedures suffering from poor accuracy, objectivity and repeatability. Diagnosis in primary care is especially poor, with GPs often failing to identify potential conditions and the patient receiving no diagnosis (Schrag, 2001). Even when the patient is referred to secondary care, the diagnosis reached by expert consultants is frequently incorrect (Bajaj et al, 2013), leading to patients receiving inappropriate treatment.

Instead, that a patient visits their GP or specialist, carries out a short and simple task that involves copying and then recalling a complex images on a tablet, and then receives an accurate diagnosis and prognosis of their condition. By looking for ways of identifying neurological signals within recordings of complex figure drawings made by PD patients, this project will provide a significant step in making this a reality. By correlating these different neurological signals with data collected in an on-going clinical study at a major neurological research centre, this project will also help researchers to understand the basis features of PD and related diseases.

This project focuses on PD. PD is the second most common neurodegenerative disorder, affecting around 1% of the population over the age of 60 (Lau and Breteler, 2006). Over the last decade, it has been increasingly recognised that PD is a neuropsychiatric disorder rather than a

pure motor disorder (Berg et al, 2013). Cognitive impairments have been detected at the onset of subtle motor symptoms (Foltynie et al, 2004) making them an important marker for early diagnosis and treatment. Approximately 10% of PD patients per year develop full-blown dementia (Pagonabarraga et al, 2012). However, at present cognitive aspects of the disease are poorly understood, and an accurate prognosis is unlikely. An important aim of this project is to use data mining techniques that can identify and discriminate different PD groups based on the occurrence of cognitive disorders. Another problem, which this project also addresses, is the difficulty of performing differential diagnosis between NDDs. These often have overlapping symptoms, meaning that patients can be incorrectly diagnosed and go on to receive inappropriate treatments. PD can be readily confused with AD, progressive supranuclear palsy (PSP), multi-system atrophy and corticobasal degeneration in its early stages (NICE, 2006). Whilst there is no curative treatment for PD, this is an area of considerable research effort (Kansara et al, 2013). If a neuroprotective drug were developed, early diagnosis of PD will become even more important. By contributing towards better understanding of the disease and better methods for monitoring disease progress, the proposed research will help address some of the obstacles faced by neuroprotective drug researchers (Stocchi and Olanow, 2013).

To achieve the aim, the following specific objectives will be met:

- **Computational exploration** of the digital recordings of patients drawing the Benson Complex Figure Copy test (Possin et al, 2011), a neurological assessment task, in order to identify the patterns of cognitive and motor signals underlying different neurological states and neurodegenerative conditions.
- **Construction** of accurate, objective diagnostic and prognostic classifiers for Parkinson's disease.
- **Evaluation** of this approach on patients' data within a clinical environment.

4.2 Motivation

Neurodegenerative Diseases. Over the next 10-20 years, NDDs are predicted to become a major social and economic problem in countries (such as the UK) with aging populations. All NDDs lead to degeneration of neural tissue, though the biological pathways through which this occurs varies considerably between diseases. For most NDDs, these pathways

remains poorly understood. An important contemporary issue, with significant implications for drug development, is whether current disease ontologies reflect the biological truth (Berg et al, 2013). Nevertheless, regardless of their underlying biology, most NDDs lead to widespread damage, affecting diverse regions of the brain. As a consequence of this, there is considerable overlap at the symptomatic level. This, in turn, can make it challenging to perform a differential diagnosis. In early stage disease, when the benefits to the patient of a correct diagnosis are greatest, the patient's symptoms may make it impossible to identify the disease using current clinical assessment techniques. Since different NDDs have different physiological causes, they require different medication. Hence, an incorrect diagnosis can have considerable bearing on a patient's welfare. Early diagnosis also has significant economic benefits, and will become increasingly important with the predicted development of curative therapies (Kansara et al, 2013).

Parkinson's disease. PD is a chronic progressive NDD 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. Where cognitive dysfunction precedes the movement disorder, PD is often known as Lewy Body Disorder (LBD). When it appears later, it is often referred to as Parkinson's with Dementia (PDD). However, given the wide range of symptoms and prognoses, it seems likely that PD is not a single disease (Berg et al, 2013). Better understanding of the true relationships between different forms of PD is likely to be an important outcome of the proposed work.

PD is currently assessed using MDS-UPDRS (Goetz, 2007), which involves a motor exam and an interview regarding the patient's cognitive state, with categories of motor and cognitive competency scored using standard scales and multiple-choice forms. Despite this, clinical assessment of PD and other neurodegenerative diseases are associated with high levels of misdiagnosis, a lack of objectivity, and poor repeatability (Bajaj et al, 2013). However, this is not necessarily due to limitations of the underlying assessment tools, but rather the way in which they are used. Poor training has been highlighted as a significant problem in this respect (Association of British Neurologists,

2011), but a more limiting issue is the extent to which humans are able to measure subtle elements of cognitive and motor ability. An example of this is the finger tapping task used in motor assessment, where the clinician attempts to measure abnormality of movement whilst a patient repeatedly taps together their thumb and forefinger. Recent results suggests that poor inter- and intra-rater performance on this task was not due to the marking scale, but rather the clinician's ability to measure motor performance by eye (Heldman et al, 2011). A potential solution, therefore, is to record the patient's behaviour and use computational techniques to measure motor skills objectively. This is something I did in a recent clinical study, showing that finger tapping alone can separate PD from age-matched controls with an accuracy of ~95%. This indicates that standard motor exams do have the potential to be accurate and objective, if supported by appropriate computational techniques.

*Work in progress

5. Professional, Legal, Ethical, and Social Issues

*Work in progress

6. Project Plan

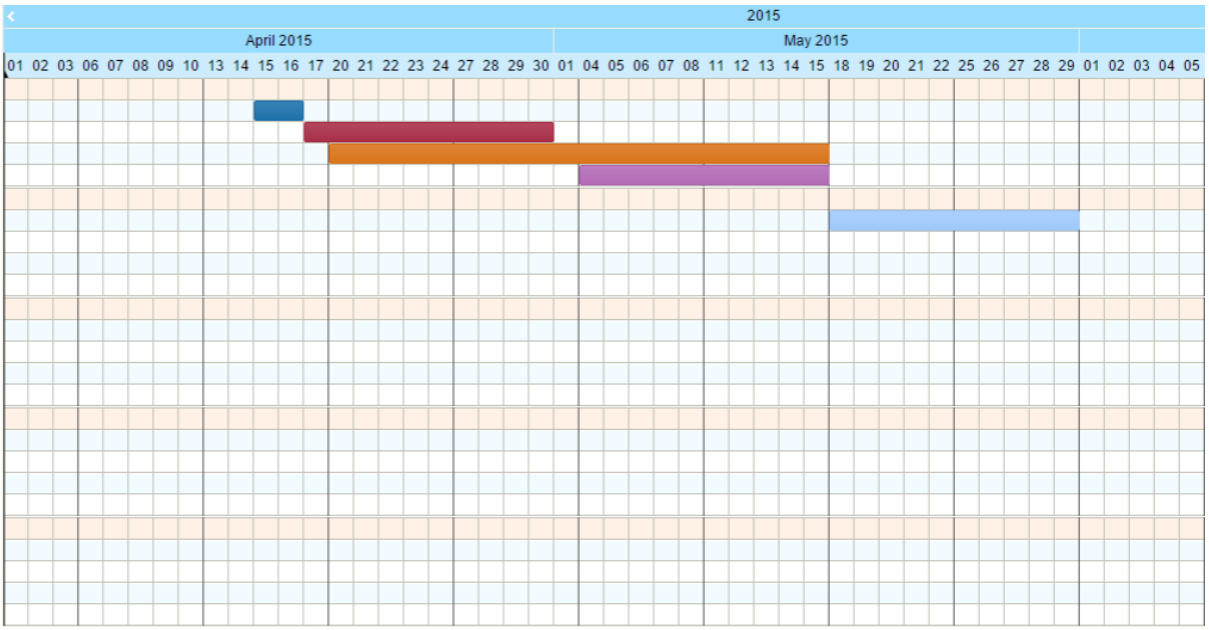

Figure: Project Time table

5.1 Preparing dataset (Blue)

There are multiple raw data files that need to be processed before they can use on the project. This task is a simple tasks but it's required some coding and times correspond to the size of the raw data sets which are large. The estimated time of this task is about 2 days.

5.2 Learning R's data mining tools (Red)

Processed datasets is required to run some test on R's data mining models. This task will be performed simultaneously with a coding task. The estimated time of this task is about 10 days.

5.3 Writing a programmes (Orange)

This task will begin on the second day of learning task. This task perform simultaneously with learning task and Evaluation task. The code and accuracy of results will be optimized correspond to the evaluation task. The estimated time of this task is about 20 days.

5.4 Evaluation (Purple)

The evaluation task is to compare an accuracy of each R's data mining models with each other also compare the accuracy with previous research projects. The estimated time of this task is about 10 days.

5.5 Writing a Report (Light Blue)

Collecting all results from experiments and write the report.

5.6 Risk Management

Several risks are clarify as follow.

- The time table above did not included Saturday and Sunday because sometimes, it's hard to work on weekend.
- The coding task is quite long (about 20 days) despite using the ready to use R's data mining models because there might be some problem discovered while coding and need to be fixed or required custom model to do certain tasks.

References

Association of British Neurologists, *Local adult neurology services for the next decade*, 2011.

Bajaj, N. et al, Accuracy of clinical diagnosis in tremulous parkinsonian patients: a blinded video study, *J Neurol Neurosurg Psychiatry* 81:1223-1228, 2010.

Berg, D. et al., Changing the research criteria for the diagnosis of PD, *Lancet Neurology* 12(5), 2013.

Deb, K. and Srinivasan, A., Innovization: Discovery of Innovative Design Principles Through Multiobjective Evolutionary Optimization, *Multiobjective Problem Solving from Nature*, pp243-262, 2008.

Foltynie, T. et al, The cognitive ability of an incident cohort of PD patients in the UK. *Brain* 27(3), 2004.

Goetz, C., MDS-sponsored revision of the Unified PD Rating Scale, *Movement Disorders* 22, 2007.

Heldman, D. et al, The Modified Bradykinesia Rating Scale for Parkinson's Disease: Reliability and Comparison with Kinematic Measures, *Movement Disorders* 26(10):1859–1863, 2011.

Kansara , S. et al., Early diagnosis and therapy of Parkinson's disease, *J. Neural Trans.* 120(1), 2013.

Lau, L. and Breteler, M., Epidemiology of Parkinson's disease, *Lancet Neurology* 5:525-535, 2006.

National Institute for Health and Clinical Excellence (NICE), Parkinson's disease: diagnosis and management in primary and secondary care, Royal College of Physicians, 2006.

Pagonabarraga, J. et al, Cognitive impairment and dementia in PD, *Neurobiology of Disease* 46, 2012.

Possin, KL. et al., Distinct neuroanatomical substrates and cognitive mechanisms of figure copy performance in AD and behavioral variant FTD, *Neuropsychologia* 49(1): 43-8, 2011.

Schrag, A., How valid is the clinical diagnosis of PD in the community?, *J Neur Neuro Psy* 73(5), 2001.

Stocchi, F. and Olanow, C., Obstacles to the Development of a Neuroprotective Therapy for Parkinson's Disease, *Movement Disorders* 28(1):3–7, 2013.