Classification of Resting-State fMRI for Olfactory Dysfunction in Parkinson's Disease using Evolutionary Algorithms

Dr. Amir Dehsarvi [1], Professor Stephen L. Smith [2]

[1] Department of Electronic Engineering, University of York, York, UK, YO10 5DD, ad953@york.ac.uk
[2] Department of Electronic Engineering, University of York, York, UK, YO10 5DD, stephen.smith@york.ac.uk

ABSTRACT

Accurate early diagnosis and monitoring of neurodegenerative conditions is essential for effective disease management and treatment. This research develops automatic methods for detecting brain imaging preclinical biomarkers for olfactory dysfunction in early stage Parkinson's disease (PD) by considering the novel application of evolutionary algorithms. Classification will be applied to PD patients with severe hyposmia, patients with no/mild hyposmia, and healthy controls. An additional novel element is the use of evolutionary algorithms to map and predict the functional connectivity using rs-fMRI. Cartesian Genetic Programming (CGP) will be used to classify dynamic causal modelling (DCM) data as well as timeseries data. The findings will be validated using two other commonly used classification methods (ANN and SVM) and by employing *k*-fold cross-validation. Developing methods for identifying early stage PD patients with hyposmia is relevant since current methods of diagnosing early stage PD have low reliability and accuracy. Furthermore, exploring the performance of CGP relative to other methods is crucial given the additional benefits it provides regarding easy classifier decoding. Hence, this research underscores the potential relevance of DCM analyses for classification and CGP as a novel classification tool for brain imaging data with medical implications for disease diagnosis, particularly in early stages.

CCS CONCEPTS

• **Computing methodologies** → **Machine Learning** → Learning Paradigms → Supervised Learning → Supervised Learning by Classification

KEYWORDS

Evolutionary Algorithms, Cartesian Genetic Programming, Classification, Parkinson's Disease, Olfactory Dysfunction, Resting-State fMRI, Dynamic Causal Modeling

1 INTRODUCTION

There is poor differential diagnosis of neurodegenerative diseases with high rates of misdiagnosis and low test-retest reliability [1]; indeed, PD has rates of misdiagnosis of 15-26%. Hence, this research examines a key question: Can early stage PD be diagnosed using EAs on rs-fMRI data? Early diagnosis of PD is fundamental in providing patients with palliative care during the early phases before motor symptoms are present, enabling effective disease management and maintaining patient quality of life. Moreover, once a neuroprotective drug to treat PD is developed, the early diagnosis of PD would have even greater clinical implications [2].

This research explores an automatic and non-invasive method of confirming the diagnosis of PD, specifically, examining the classification of PD patients with severe hyposmia (olfactory dysfunction), patients with no/mild hyposmia, and healthy controls. This research will involve an analysis of rs-fMRI data taken from the OpenfMRI (accession number: ds000245). This study will apply EAs (CGP and RCGP) for the classification of rs-fMRI in PD using DCM and timeseries analyses. The timeseries values and DCM values from the rs-fMRI data will be subjected to supervised classification and the findings will be validated with two other commonly used classification methods (ANN and SVM) as well as employing *k*-fold CV.

A key aim of this research is to identify the applicability of CGP and RCGP classification for both timeseries and DCM analyses regarding the analysis of PD data. CGP and RCGP have not previously been used in the classification of brain imaging data. This study examines an additional novel question: is DCM analysis useful for classification for PD data? Previous research has not explored the applicability of DCM values in classification and, to date, little research has applied DCM to PD data [3]. Hence, by doing both, this research develops automatic procedures for identifying PD brain imaging preclinical biomarkers for olfactory dysfunction, which can be used for aiding/confirming early PD diagnosis.

Based on previous work from the authors revealing classification accuracies of up to 92% for identifying a biomarker for early stage PD, it is expected that evolutionary algorithms can be applied to olfactory dysfunction for PD patients with similarly high accuracies.

2 EXPERIMENTAL AND COMPUTATIONAL DETAILS

2.1 Participants

Forty-five participants were recruited with an age range of 52–78 years (20 male): 15 PD patients with severe hyposmia, 15 patients with no/mild hyposmia, and 15 healthy controls were examined. Brain atrophic changes and functional connectivity alterations in all participants were evaluated using Odour Stick Identification Test for the Japanese and Addenbroke's Cognitive Examination Revised for their odour-identification performance and general cognitive function, respectively. The study was approved by the institutional review board of participating sites. Written informed consent was obtained from all participants.

2.2 Classification

To have equal class representation, data from each class will be divided randomly into subsets of 70% (training), 15% (validation), and 15% (test). The geometry of the programs in the population (chromosomes) has fifty nodes with a function set of four mathematical operations (+, -, ×, ÷), multiple inputs (according to the datasets), and one output (class 1 for one participant group, class 0 for the other participant group). At each generation, the fittest chromosome is selected and the next generation is formed with its mutated versions (mutation rate = 0.1). Evolution stops when 15000 iterations are reached. To obtain statistical significance, the classification will be done in 10 runs for each combination of inputs and the accuracy was averaged over the runs. The results (the winning chromosome, the networks, and the accuracy values) will be stored for each run individually.

One of the core advantages of CGP is that it provides a *white box solution*, giving more information on the inputs used and enhanced knowledge concerning the final solution obtained in classification, which is not easily possible (if at all) with ANN and SVM classification techniques. The classification will be done using CGP, ANN, and SVM (the latter two being used to validate the classification results from CGP).

2.3.1 Classification of the DCM values.

Following preprocessing and processing the fMRI data, DCM analysis results will be subjected to classification implemented in CGP and RCGP. One group of participants will be assigned class 0 and a different group will be assigned class 1. To have equal class representation, data from each class will be divided randomly into subsets of 70% (training), 15% (validation), and 15% (test). The classification will be run in the CGP Library with 16 inputs (all the DCM values sorted by region and presented as only one vector per participant, see Figure 1), 50 nodes, 1 output (class 1 for one group, class 0 for the other one), and 15000 generations. The process will be repeated for 10 runs and the results (the winning chromosome, the networks, and the accuracy values) will be stored for each run individually. To facilitate comparison,

classification with data from the same participants will be run using ANN and SVM.
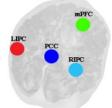

Figure 1: Default mode network and the four regions of interest used in the DCM analysis.

2.3.2 k-fold Cross-Validation (CV).

10-fold CV will be used to evaluate classification accuracy using an unbiased estimate of the generalisation accuracy [4]. CV is beneficial as it allows the generation of independent test sets with enhanced reliability. *10*-fold CV is repeated 10 independent times and the mean accuracy over all 10 trials is calculated.

3 RESULTS

Provisional results revealed that the CGP classifiers had an accuracy rate of 90% for classifying PD patients with severe hyposmia from healthy controls. The results from two other classification techniques (ANN and SVM) validated this finding as they were very similar.

4 CONCLUSIONS

To conclude, this research explores the classification of rs-fMRI data in early stage PD patients with severe hyposmia, patients with no/mild hyposmia, and healthy controls using novel data (DCM analyses) and a novel classification method (CGP). Classification accuracy for DCM analyses will be compared to that of timeseries analyses, and two other classification methods (ANN and SVM) will be used to validate the findings, as well as *k*-fold cross-validation. Hence, this research underscores the potential relevance of CGP as a classification tool for two types of brain imaging data (DCM and timeseries rs-fMRI analyses) and develops automatic procedures for identifying PD brain imaging preclinical biomarkers for olfactory dysfunction, which are fundamental in improving accuracy of early stage PD diagnosis methods.

SELECTED REFERENCES

<bib id="bib1" type="Periodical"><number>[1]</number>A. Schrag, Y. Ben-Shlomo, and N. Quinn, "How valid is the clinical diagnosis of Parkinson's disease in the community?," *J. Neurol. Neurosurg. Psychiatry*, vol. 73, no. 5, pp. 529–534, Nov. 2002.</bib>
<bib id="bib2" type="Periodical"><number>[2]</number>F. Stocchi and C. W. Olanow, "Obstacles to the development of a neuroprotective therapy for Parkinson's disease.," *Mov. Disord.*, vol. 28, no. 1, pp. 3–7, Jan. 2013.</bib>
<bib id="bib3" type="Periodical"><number>[3]</number>D. M. Herz, H. R. Siebner, O. J. Hulme, E. Florin, M. S. Christensen, and L. Timmermann, "Levodopa reinstates connectivity from prefrontal to premotor cortex during externally paced movement in Parkinson's disease," *Neuroimage*, vol. 90, pp. 15–23, 2014.</bib>
<bib id="bib4" type="Periodical"><number>[4]</number>R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *The International Joint Conference on Artificial Intelligence (IJCAI)*, 1995, vol. 14, no. 2, pp. 1137–1145.</bib>