NeuroImage: Clinical 2 (2013) 883–893

Contents lists available at ScienceDirect

NeuroImage: Clinical

journal homepage: www.elsevier.com/locate/ynicl

Multiclass classification of FDG PET scans for the distinction between Parkinson's disease and atypical parkinsonian syndromes☆☆

Gaëtan Garraux [a,c,*,1], Christophe Phillips [a,b,1], Jessica Schrouff [a], Alexandre Kreisler [f,g,h,i], Christian Lemaire [a], Christian Degueldre [a], Christian Delcour [e], Roland Hustinx [d], André Luxen [a], Alain Destée [f,g,h,i], Eric Salmon [a,c]

[a] Cyclotron Research Centre, Sart Tilman B30, University of Liège, 4000 Liège, Belgium
[b] Department of Electrical Engineering and Computer Science, Sart Tilman B28, University of Liège, 4000 Liège, Belgium
[c] Department of Neurology, University Hospital Centre, Sart Tilman B35, 4000 Liège, Belgium
[d] Department of Nuclear Medicine, University Hospital Centre, Sart Tilman B35, 4000 Liège, Belgium
[e] Department of Neurology, Centre Hospitalier Peltzer-La Tourelle, 4800 Verviers, Belgium
[f] Université Lille Nord de France, 59000 Lille, France
[g] UDSL (Université Droit et Santé de Lille), Lille, France
[h] CHU de Lille, France
[i] Movement Disorders Unit and EA 2683 MENRT, France

ARTICLE INFO

Article history:
Received 5 August 2012
Received in revised form 6 June 2013
Accepted 7 June 2013
Available online 14 June 2013

Keywords:
Computer-aided diagnosis
Data mining
Pattern recognition
Boostrap resampling
Bagging
Error-Correcting Output Code
Multiclass classification
Relevance vector machine
FDG PET
Parkinson's disease
Multiple system atrophy
Progressive supranuclear palsy
Corticobasal syndrome

ABSTRACT

Most available pattern recognition methods in neuroimaging address binary classification problems. Here, we used relevance vector machine (RVM) in combination with booststrap resampling ('bagging') for non-hierarchical multiclass classification. The method was tested on 120 cerebral ^{18}fluorodeoxyglucose (FDG) positron emission tomography (PET) scans performed in patients who exhibited parkinsonian clinical features for 3.5 years on average but that were outside the prevailing perception for Parkinson's disease (PD). A radiological diagnosis of PD was suggested for 30 patients at the time of PET imaging. However, at follow-up several years after PET imaging, 42 of them finally received a clinical diagnosis of PD. The remaining 78 APS patients were diagnosed with multiple system atrophy (MSA, N = 31), progressive supranuclear palsy (PSP, N = 26) and corticobasal syndrome (CBS, N = 21), respectively. With respect to this standard of truth, classification sensitivity, specificity, positive and negative predictive values for PD were 93% 83% 75% and 96%, respectively using binary RVM (PD vs. APS) and 90%, 87%, 79% and 94%, respectively, using multiclass RVM (PD vs. MSA vs. PSP vs. CBS). Multiclass RVM achieved 45%, 55% and 62% classification accuracy for, MSA, PSP and CBS, respectively. Finally, a majority confidence ratio was computed for each scan on the basis of class pairs that were the most frequently assigned by RVM. Altogether, the results suggest that automatic multiclass RVM classification of FDG PET scans achieves adequate performance for the early differentiation between PD and APS on the basis of cerebral FDG uptake patterns when the clinical diagnosis is felt uncertain. This approach cannot be recommended yet as an aid for distinction between the three APS classes under consideration.

1. Introduction

Computer-aided diagnosis (CAD) integrates data processing, mathematics and statistics into computerized techniques to maximize the information that may be extracted from medical imaging datasets. One of the goals of CAD is to assist the clinicians in the differential diagnosis between several conditions with overlapping clinical features. This problem is commonly encountered in patients with a presumed progressive adult-onset chronic neurodegenerative disorder, in which the clinical phenotype only fully expressed several years after the onset of brain damage. Most CAD in this context addressed a binary classification problem i.e., involving the distinction between two diagnostic classes. One of the challenges of CAD is multiclass classification (Kloppel et al., 2012), which better reflects a situation encountered in routine clinical practice. As compared with binary classification, multiclass classification is a more complex problem and their performances are difficult to compare directly. Here, we present simple binary and new multiclass classification methods and test their performance for the distinction between different forms of degenerative parkinsonism.

☆☆ *Financial support*: This research was supported by FRS-FNRS and a grant from the Rahier Foundation, University of Liège, Belgium.

* Corresponding author at: MoVeRe Group, Cyclotron Research Centre, University of Liège, Sart Tilman B30, 4000 Liège, Belgium. Tel.: +32 4 366 23 16; fax: +32 4 366 29 46.
E-mail address: ggarraux@ulg.ac.be (G. Garraux).

[1] Both authors equally contributed to this work.

2213-1582 © 2013 The Authors. Published by Elsevier Inc. Open access under CC BY-NC-ND license.
http://dx.doi.org/10.1016/j.nicl.2013.06.004

Parkinsonism is clinically defined by the association of motor slowness, with muscle rigidity and/or tremor and/or a postural instability (Gibb, 1988). The most common cause of degenerative parkinsonism in adults is Parkinson's disease (PD). Much of the difficulty in the early diagnosis of PD is differentiating it from other forms of degenerative parkinsonism. A common source of misdiagnosis of PD is atypical parkinsonian syndromes (APS) that have a much poorer long-term prognosis such as multiple system atrophy (MSA), progressive supranuclear palsy (PSP) and corticobasal syndrome (CBS). In a clinico-pathological study conducted in a specialist movement disorder service, more than 60% of cases with a final clinical diagnosis of a parkinsonian syndrome other than PD had their diagnosis changed during the course of their illness. Of these, 60% were changed from an initial clinical diagnosis of PD (Hughes et al., 2002; Rajput et al., 1991).

Resting-state cerebral ^{18}fluorodeoxyglucose (FDG) uptake patterns measured using positron emission tomography (PET) has been recommended by the European Association of Nuclear Medicine Neuroimaging Committee for the differentiation between degenerative parkinsonisms (Varrone et al., 2009) under the assumption that FDG PET can capture specific functional and anatomical consequences of neuropathologic abnormalities specific of each condition. This is supported by the demonstration of group differences in regional FDG uptake between PD, MSA, PSP and CBS (Antonini et al., 1998; Eckert et al., 2005; Eidelberg et al., 1993; Feng et al., 2008; Ghaemi et al., 2002; Juh et al., 2004; Klein et al., 2005; Laureys et al., 1999; Otsuka et al., 1997; Teune et al., 2010). One of the most consistent abnormalities at visual inspection and semi-quantitative analyses is a relative decrease in striatal and frontal lobe tracer uptakes in APS as compared with PD or normal control populations. While this has been very informative at the group level, its diagnostic yield has been lower than expected in the early stages of these disorders because of overlapping individual *regional* FDG uptake across groups, which were often composed of small series of either well established cases studied with PET after a relatively long disease duration or early cases but without information on clinical follow-up to ascertain the initial clinical diagnosis (Garraux et al., 2000; Ghaemi et al., 2002; Juh et al., 2004; Laureys et al., 1999; Otsuka et al., 1997).

Here, we examined the value of CAD for the distinction between PD, MSA, PSP and CBS on the basis of cerebral FDG PET. The present study differs from previous ones by several methodological aspects with respect to both the population characteristics and analysis methods. First, to maximize the clinical significance of cerebral FDG PET for distinction between the diagnostic classes under consideration, we included scans performed in the first years after symptom onset (Table 1) at a time when clinical features were outside the prevailing perceptions for PD. Diagnostic classes were then defined later by the retrospective application of clinical diagnostic criteria for PD and APS at follow-up, on average ~8.0 and ~2.8 years after PET assessment (i.e., standard of truth). Second, a crucial difference with previous studies is the analysis of neuroimaging data using an automatic voxel-based multivariate supervised machine learning method, "Relevance Vector Machine" (RVM) (Tipping, 2001), that we have previously applied on a binary case for the distinction between patients with or without an altered state of consciousness on the basis of cerebral FDG uptake patterns (Phillips et al., 2011). We profoundly modified this method to be suitable for multiclass classification.

Classification was both performed in a binary sense, PD versus all the APS subcategories pooled into a single class, and in a multiclass sense, PD and the 3 APS categories considered separately. For multiclass classification, pairwise coupling is a popular approach that combines all comparisons for each pair of classes (Fürnkranz, 2002). Here, we used a one-versus-one approach involving six binary RVM classifiers from which a single prediction was obtained using an Error-Correcting Output Code (ECOC) approach (Dietterich and Bakiri, 1995) (see Section 2.4.3). For cross-validation and assessment of prediction accuracy, RVM was combined with bootstrap aggregation (also known as "bagging") (Breiman, 1996; Efron and Tibshirani, 1993) and the final RVM class assigned to each FDG PET scan was defined by the prediction that received the most votes (see Section 2.4.4).

The final class assigned to each FDG PET scan was then compared with the clinical diagnosis at follow-up to estimate prediction accuracy, sensitivity, specificity, positive (PPV) and negative (NPV) predictive values. The statistical significance of RVM classification accuracy was assessed using a permutation testing (see Section 2.4.4). We also compared binary RVM classification with the radiological diagnosis of the nuclear medicine specialist at the time of PET imaging (i.e., for the distinction between PD and APS).

Finally, from the vote counting in the bootstrap procedure, a "majority confidence ratio" was estimated for each scan on the basis of class pairs that were the most frequently assigned by RVM. This level of confidence was further linked to the PPV (see Section 2.4.5). We believe that this qualification of the classification outcome may provide clinically relevant information at the individual level for physicians who usually request FDG PET scans as an aid to solve a multiclass diagnosis problem.

2. Methods

2.1. Subjects

Patients were all referred for cerebral FDG PET at the Cyclotron Research Centre (CRC), University of Liège, or the University Hospital Center (CHU) of Liège by neurologists because clinical features were outside the prevailing perceptions for PD. In many cases, no other specific diagnosis was mentioned in the PET order form and no standardized clinical assessment was available in this retrospective study. The most frequent atypical features at referral were an equivocal clinical response to scheduled L-DOPA administration, prominent axial symptoms, greater than expected asymmetry of parkinsonian signs, early falls, or the co-occurrence of other features such as a pyramidal and/or cerebellar syndrome, limb dystonic posturing, oculomotor abnormalities, or severe dysautonomic dysfunction. All subjects included in this research protocol gave their written informed consent to participate in the study; the study protocol was approved by the Ethical Committee of the University of Liège.

Table 1
Demographic and clinical data.

	N	Gender (F/M)	Data at the time of PET assessment			Last available follow-up
			Mean age (years)	Mean disease duration (years)	Mean LEDD (mg)	Mean disease duration (years)
PD	42	17/25	56.9 ± 10.3	3.6 ± 3.1	442 ± 239	11.6 ± 5.1
MSA	31	18/13	66.0 ± 8.8	3.4 ± 2.9	559 ± 298	6.4 ± 3.9
PSP	26	9/17	69.4 ± 7.3	3.1 ± 2.4	281 ± 250	5.9 ± 4
CBS	21	15/6	67.8 ± 7	3.3 ± 2	164 ± 189	5.9 ± 2.9
All classes	120	59/61	63.9 ± 10.2	3.4 ± 2.7	386 ± 284	8.0 ± 5.0

LEDD = L-DOPA equivalent daily dose (Tomlinson et al., 2010).