Automated, High Accuracy Classification of Parkinsonian Disorders: A Pattern Recognition Approach

Andre F. Marquand[1*], Maurizio Filippone[2], John Ashburner[3], Mark Girolami[4], Janaina Mourao-Miranda[1,4], Gareth J. Barker[1], Steven C. R. Williams[1], P. Nigel Leigh[5], Camilla R. V. Blain[1]

1 Department of Neuroimaging, Centre for Neuroimaging Sciences, Institute of Psychiatry, King's College London, London, United Kingdom, 2 School of Computing Science, University of Glasgow, Glasgow, United Kingdom, 3 Wellcome Trust Centre for Neuroimaging, University College London, London, United Kingdom, 4 Centre for Computational Statistics and Machine Learning, University College London, London, United Kingdom, 5 Brighton and Sussex Medical School, Trafford Centre for Biomedical Research, University of Sussex, Falmer, East Sussex, United Kingdom

Abstract

Progressive supranuclear palsy (PSP), multiple system atrophy (MSA) and idiopathic Parkinson's disease (IPD) can be clinically indistinguishable, especially in the early stages, despite distinct patterns of molecular pathology. Structural neuroimaging holds promise for providing objective biomarkers for discriminating these diseases at the single subject level but all studies to date have reported incomplete separation of disease groups. In this study, we employed multi-class pattern recognition to assess the value of anatomical patterns derived from a widely available structural neuroimaging sequence for automated classification of these disorders. To achieve this, 17 patients with PSP, 14 with IPD and 19 with MSA were scanned using structural MRI along with 19 healthy controls (HCs). An advanced probabilistic pattern recognition approach was employed to evaluate the diagnostic value of several pre-defined anatomical patterns for discriminating the disorders, including: (i) a subcortical motor network; (ii) each of its component regions and (iii) the whole brain. All disease groups could be discriminated simultaneously with high accuracy using the subcortical motor network. The region providing the most accurate predictions overall was the midbrain/brainstem, which discriminated all disease groups from one another and from HCs. The subcortical network also produced more accurate predictions than the whole brain and all of its constituent regions. PSP was accurately predicted from the midbrain/brainstem, cerebellum and all basal ganglia compartments; MSA from the midbrain/brainstem and cerebellum and IPD from the midbrain/brainstem only. This study demonstrates that automated analysis of structural MRI can accurately predict diagnosis in individual patients with Parkinsonian disorders, and identifies distinct patterns of regional atrophy particularly useful for this process.

Introduction

The akinetic-rigid syndromes of progressive supranuclear palsy (PSP), multiple system atrophy (MSA) and idiopathic Parkinson's disease (IPD), can be clinically indistinguishable in the early stages [1] despite having distinct characteristic patterns of molecular pathology [2–4]. Finding sensitive and specific objective biomarkers for predicting disease state in these disorders is an important aim for several reasons: first, the disorders have different prognoses, where MSA and PSP are characterised by relentless disease progression and carry a life expectancy of only a few years after diagnosis, IPD does not convey a substantial reduction in life expectancy. Second,

the disorders have differential responses to treatment; IPD responds moderately well to dopaminergic therapy and deep-brain stimulation [5], but PSP and MSA are both associated with a poor response [6]. Third, objective biomarkers predictive of early disease state may be useful to reduce the misdiagnosis rate in clinical trials of potential disease-modifying compounds. However, for any objective measure to facilitate clinical decision making in the long term, it must accurately and simultaneously discriminate between all the disorders.

Magnetic resonance imaging (MRI) holds the potential to provide objective diagnostic markers for the disorders. However, no published studies have demonstrated an automated approach to predict diagnosis in individual subjects

with accuracy that could be considered clinically useful. Existing studies have employed either manual measurements derived from radiological examination of MRI scans (rMRI) [7–10] or automated approaches based on voxel-based morphometry (VBM) [11–14]. Both approaches have disadvantages: rMRI markers are operator-dependent and time-consuming to construct and are not sufficiently specific for discriminating between MSA and PSP despite good sensitivity for discriminating both from IPD [15]. Whilst VBM has been successful in identifying neuroanatomical changes associated with these disorders at the group level, it has limited ability to predict disease state at the level of individual subjects.

Pattern recognition (PR) is an analytic approach increasingly being applied in clinical neuroimaging studies [16,17]. In contrast to rMRI and VBM, PR aims to predict disease state at the single-subject level based on distributed patterns of anatomical abnormality. PR has been highly successful for discriminating other neurological disorders [16–19], but only two studies have applied PR to Parkinsonian disorders and were unable to accurately discriminate all diagnostic groups [20,21].

The primary objective of this work was to assess the capability of anatomical patterns (networks) of brain regions for automated discrimination of Parkinsonian disorders, aiming to discriminate between all disorders simultaneously and identify which networks would provide the best discrimination of each disorder. To achieve this, networks of subcortical regions were defined prior to the automated analyses, based on the known distribution of tau (PSP) or α-synuclein (MSA/IPD) pathology [2–4]. An advanced multi-class PR approach was then employed to assess the diagnostic capability of the full network, each component region and the whole brain. A secondary aim was to determine whether MSA subtypes MSA-P and MSA-C (predominantly Parkinsonian or cerebellar symptoms) could be discriminated and further, whether regarding them as single or distinct entities yields more accurate discrimination, since they have different burdens of brainstem and basal ganglia pathology [22].

We hypothesized that discrimination of PSP and MSA would be achieved with high accuracy while discrimination of IPD would be more challenging since most MRI studies report only subtle abnormalities in early- or mid-stage IPD [13,23]. Additionally, we hypothesized that: (i) the midbrain and cerebellum would be predictive of PSP, because atrophy of the midbrain and superior cerebellar peduncles (SCP) are rMRI markers for PSP [7,9,10]; (ii) the cerebellum would be predictive of MSA because middle cerebellar peduncle (MCP) width is an rMRI marker of MSA [8–10] and (iii) the midbrain/ brainstem would be the most predictive region for IPD based on its distribution of pathology [4] and a recent report of ponto-medullary degeneration in early IPD [23]. Finally, we sought to test whether the network or any of its components outperformed a whole-brain approach, which is important because cortical atrophy has been reported in all disorders [11,13,24].

Methods

Case selection

Seventeen patients with PSP, 19 with MSA and 14 with IPD participated (all diagnosed according to established criteria [25–27]) and were recruited according to procedures described elsewhere [28,29]. Five PSP patients met diagnostic criteria for definite, 11 for probable (clinically definite [1]) and one for possible PSP. All PSP patients could be considered to have the classical PSP-Richardson phenotype [30]. Twelve MSA patients were categorized as having MSA-P (one patient could be considered to have possible-, nine to have probable and two to have definite MSA according to recent updates to the diagnostic criteria [31]). Seven MSA patients were categorized as having MSA-C (six probable and one definite MSA). All IPD patients fulfilled criteria for clinically definite IPD [25]. All 13 IPD patients taking dopaminergic medication reported a good or excellent response and the six PSP and 13 MSA patients taking dopaminergic medication all described their response as poor. Nineteen healthy controls (HCs; spouses and friends of patients) with no known neurological disorder also participated. Disease severity was recorded using the Unified Parkinson's Disease Rating Scale (UPDRS), plus Hoehn and Yahr (HY) [32] and Schwab and England Activities of Daily Living (ADL) scales [33]. Cerebellar ataxia was assessed using the Parkinson's plus scale [34] and postural instability using the Postural Instability and Gait Disorder (PIGD) scale [35] (Table 1). All participants provided informed written consent and the study was approved by the Research Ethics Committees of King's Healthcare NHS Trusts and the Institute of Psychiatry.

Neuroimaging data acquisition/preprocessing

For each subject, a whole-brain T1-weighted 3-dimensional inversion recovery prepared spoiled gradient echo (SPGR) structural image was acquired using a 1.5T General Electric, Signa LX NV/i scanner (General Electric, WI, USA) with parameters: repetition time = 18ms, echo time = 5.1ms, inversion time = 450 ms, acquisition matrix = 256×152 over a 240×240 field of view, reconstructed as a 256x256 matrix, yielding in-plane voxel size of 0.94×0.94mm and 124 1.5 mm thick slices. In addition, a 2D T2-weighted structural image (used to screen participants for incidental structural lesions) and a diffusion-tensor imaging (DTI) sequence were acquired as described elsewhere [28]. Since SPGR images are more widely available and faster to acquire than DTI, we focus on these for the present work. The data from a subset of the subjects used in the present work were used in a companion paper where we validated the analytic methodology [36] and the DTI images from a different subset have been reported separately [28].

The SPGR images were used to derive a set of "scalar momentum" features [37] to describe anatomical variability amongst subjects (see materials S1 for details). The components of these images corresponding to grey- and white-matter were masked anatomically to constrain them to either: (i) the whole brain, (ii) a subcortical motor network comprising bilateral cerebellum, brainstem (including midbrain and decussations of SCP but excluding the MCP), caudate,

Table 1. Demographic and clinical information.

	HCs (n=19)	PSP (n=17)	IPD (n=14)	MSA (n=19) [MSA-P, n=12; MSA-C, n=7]
Age, mean ± SD	63.9 ± 7.8	68.6 ± 6.5	64.6 ± 6.9	64.0 ± 7.7 [64.0 ± 6.7; 60.6 ± 8.3]
Sex, M:F	10:9	7:10	7:7	10:9 [4:8 ; 6:1]
Disease duration, mean ± SD	–	5.3 ± 2.4	6.6 ± 2.0	4.9 ± 2.3 [4.4 ± 2.2; 5.5 ± 2.5]
HY, mean (range)	–	4.0 (3.0-4.0)	2.5 (2.0-3.0)	3.0 (2.5-5.0) [3.0 (2.5-5.0); 4.0 (3.0-4.0)]
ADL, median (range)	–	50% (20-80)	90% (80-100)	70% (40-80) [70% (40-80); 70% (60-80)]
UPDRS-III, mean ± SD	–	34.8 ± 7	21.7 ± 9.6	35.7 ± 13.8 [42.6 ± 12.3; 25.0 ± 8.0]
PIGD score, mean (range)	–	11.0 (7.0-18.0)	3.0 (1.0-6.0)	9.0 (5.0-14.0) [8.0 (5.0-14.0); 9.0 (8.0-11.0)]
Cerebellar, median (range)	–	2.0 (0.0-6.0)	0.0 (0.0-2.0)	8.5 (0.0-13.0) [4.0 (0.0-10.0); 10.0 (0.0-13.0)]

For patients taking levodopa, scores are given in the "on" state. Scales: HY: Hoehn and Yahr; ADL: Schwab and England Activities of Daily Living; UPDRS-III: Unified Parkinson's Disease Rating Scale-part 3, PIGD: Postural instability and gait disorder. Cerebellar scores are taken from the Parkinson's Plus Scale (maximum = 24).

putamen, pallidum and accumbens or (iii) each of these component regions, separately. Both components were concatenated and used as classifier inputs.

Pattern Recognition Analysis

Nearly all applications of PR to neuroimaging have employed pair-wise categorical classification, but here we employed a multi-class probabilistic approach. This is preferable for Parkinsonian disorders because: (i) it aims to separate all disease classes simultaneously, thus more closely resembling the clinical decision-making process and (ii) provides quantitative measures of diagnostic confidence. The PR approach employed here is described in detail in a companion methodological report [36] and is outlined in materials S1. Four contrasts were applied to discriminate different combinations of disease groups: classifier I aimed to separate disease groups, replicating the decision process employed clinically (i.e. PSP vs. IPD vs. MSA; chance level=33%); classifier II aimed to separate disease groups and HCs (PSP vs. IPD vs. HCs vs. MSA; chance=25%); classifiers III and IV were similar to classifiers I and II respectively, except the MSA class was separated into distinct MSA-P and MSA-C groups (classifier III: PSP vs. IPD vs. MSA-P vs. MSA-C; chance=25%. Classifier IV:PSP vs. IPD vs. HCs vs. MSA-P vs. MSA-C; chance=20%). All four classifiers were applied to the whole-brain and subcortical motor network and Classifier II was applied to assess the diagnostic value of regional features because it can be used to examine the relationship of each disease group to

HCs. The discriminative value of different brain regions was also assessed at a finer scale than was afforded by the anatomical network by examining the pattern of predictive voxel weights for classifier II. This represents a multi-class generalisation of an approach employed elsewhere for binary classification [38–42] (see materials S1 for details).

To estimate the generalisability of each model for new cases, it is crucial to evaluate it using data that has not been used in any way to build the model (e.g. to infer parameters). Leave-one-out cross-validation, which provides approximately unbiased estimates of the true generalizability, was used to achieve this (see materials S1 for details). Note that all data preprocessing steps were embedded within this cross-validation loop, including the creation of a study-specific template for volumetric normalisation. Thus, the training and test sets were entirely independent during all stages of model construction and assessment,

Classifier assessment

Each classifier's errors can be summarised using confusion matrices, which indicate the ease, or difficulty with which classes could be separated. In the binary case, these give rise to the sensitivity, specificity and positive/negative predictive value (PV). Here, straightforward multi-class generalisations were derived for the sensitivity and PV, which describe the performance for each class (Figure 1). The (balanced) accuracy and overall predictive value (OPV) were then computed by averaging these over all classes. Note that the class specificity as typically employed in the binary context does not straightforwardly generalise to multi-class cases, since more than one type of misclassification can occur. However, the PV indirectly measures specificity for each class. Significance of each metric was assessed using Monte Carlo testing (see materials S1).

Results

Demographic variables

Diagnostic groups did not differ significantly with respect to age ($F_{3,65}$=1.8, p=0.17), sex (X^2=0.62; p=0.89) or disease duration ($F_{2,47}$=1.9; p=0.15) (Table 1).

Classification performance: subcortical motor network

All subcortical network classifiers (Table 2) exceeded chance accuracy and OPV (p < 0.001, Monte Carlo test). Classifier I discriminated all classes with high sensitivity and PV, (Figure 2), making only four errors: one IPD case was predicted as PSP and three PSP cases were predicted as IPD. The MSA class was predicted perfectly (Figure 3).

Classifier II exceeded chance sensitivity and PV for all classes except IPD which was significant for sensitivity only at trend level (Figure 2). This was due to several IPD cases being mistaken for HCs; MSA and PSP remained well classified, one MSA case was mistakenly predicted as a control and two PSP cases were predicted as IPD (Figure 3). For MSA, the incorrectly labelled subject was an MSA-P patient (thus, MSA-
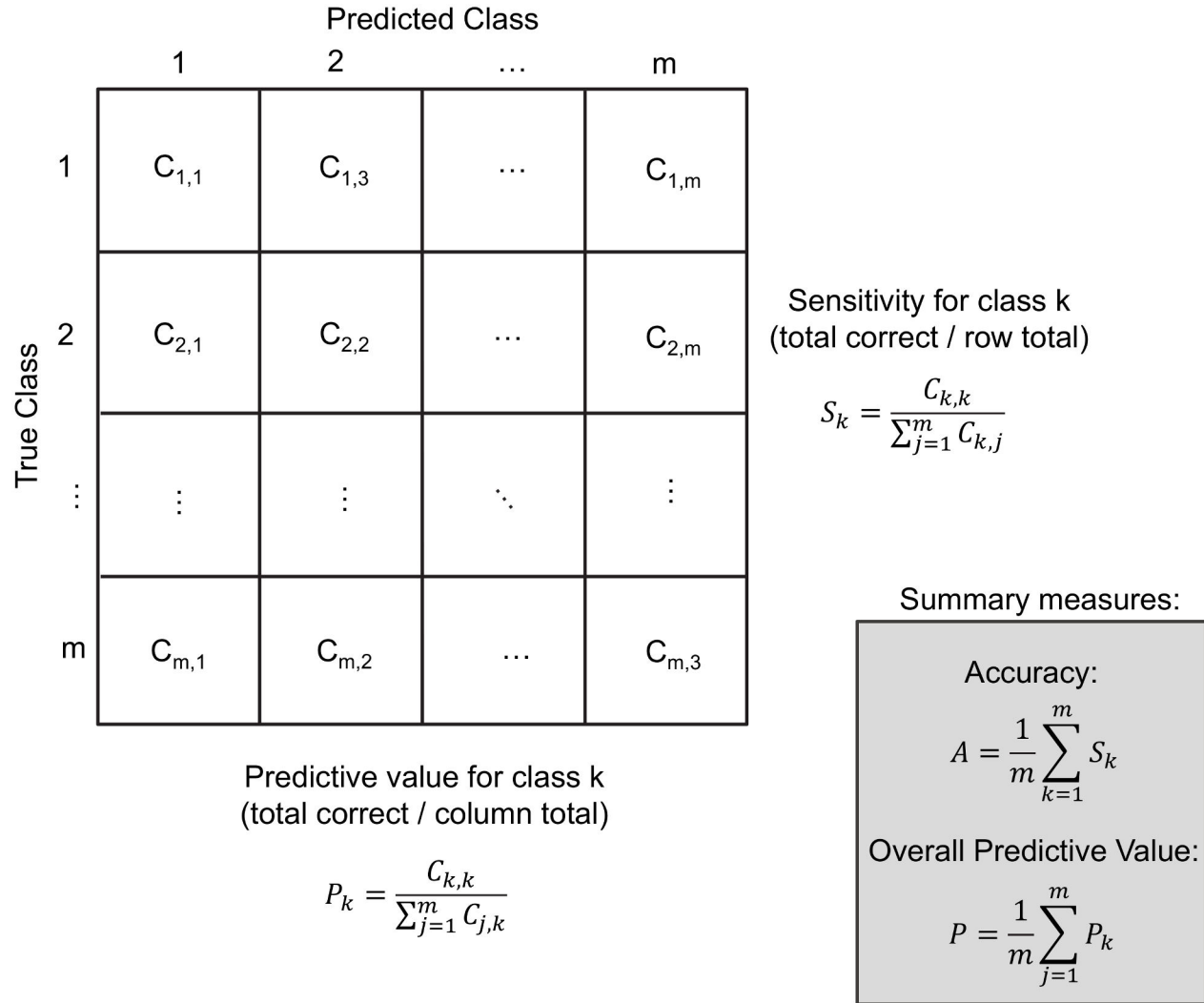
Predicted Class

Sensitivity for class k
(total correct / row total)

$$S_k = \frac{C_{k,k}}{\sum_{j=1}^{m} C_{k,j}}$$

Summary measures:

Accuracy:

$$A = \frac{1}{m} \sum_{k=1}^{m} S_k$$

Overall Predictive Value:

$$P = \frac{1}{m} \sum_{j=1}^{m} P_k$$

Predictive value for class k
(total correct / column total)

$$P_k = \frac{C_{k,k}}{\sum_{j=1}^{m} C_{j,k}}$$

Figure 1. Example confusion matrix for an m-class classification problem. $C_{i,j}$ denotes the number of predictions in row i, column j. The sensitivity and predictive value measure the performance of each class. The accuracy and overall predictive value are constructed by averaging the sensitivity and predictive value over all classes. Note that the accuracy and overall predictive value are balanced in that they avoid potential bias arising from variable numbers of samples in each class.
doi: 10.1371/journal.pone.0069237.g001

P = 91.7% sensitivity). All MSA-C cases were correctly classified (MSA-C = 100% sensitivity).

Classifier III exceeded chance sensitivity and PV for all classes (Figure 2). Notably, MSA-P and MSA-C were accurately discriminated, although there was some overlap between them (Figure 3). The MSA-P class was relatively poorly discernable, being frequently mistaken for PSP or MSA-C (Figure 3). For all three classifiers described above, all pathologically confirmed cases were correctly classified.

Classifier IV displayed similar characteristics to the other classifiers: all disease groups except IPD were discriminated above chance (Figure 2) and misclassifications were mainly between either IPD and HCs or MSA-P and MSA-C/PSP (Figure 3). For this classifier, all pathologically confirmed cases

were correctly classified except one PSP case (predicted as MSA-C).

Classification performance: regional classifiers

Classifier II exceeded chance accuracy and OPV in all regions except the nuclei accumbens (Table 3). The region producing the most accurate predictions overall was the midbrain/brainstem, achieving only slightly lower accuracy (-1.7%) and OPV (-2.0%) than the subcortical motor network (Table 3). For PSP, all regions were predictive (Figure 4). For MSA, the cerebellum and midbrain/brainstem were highly predictive and the putamina were moderately predictive. The cerebellum and midbrain/brainstem were predictive of both variants of MSA (cerebellum: MSA-P = 83.3% sensitivity, MSA-
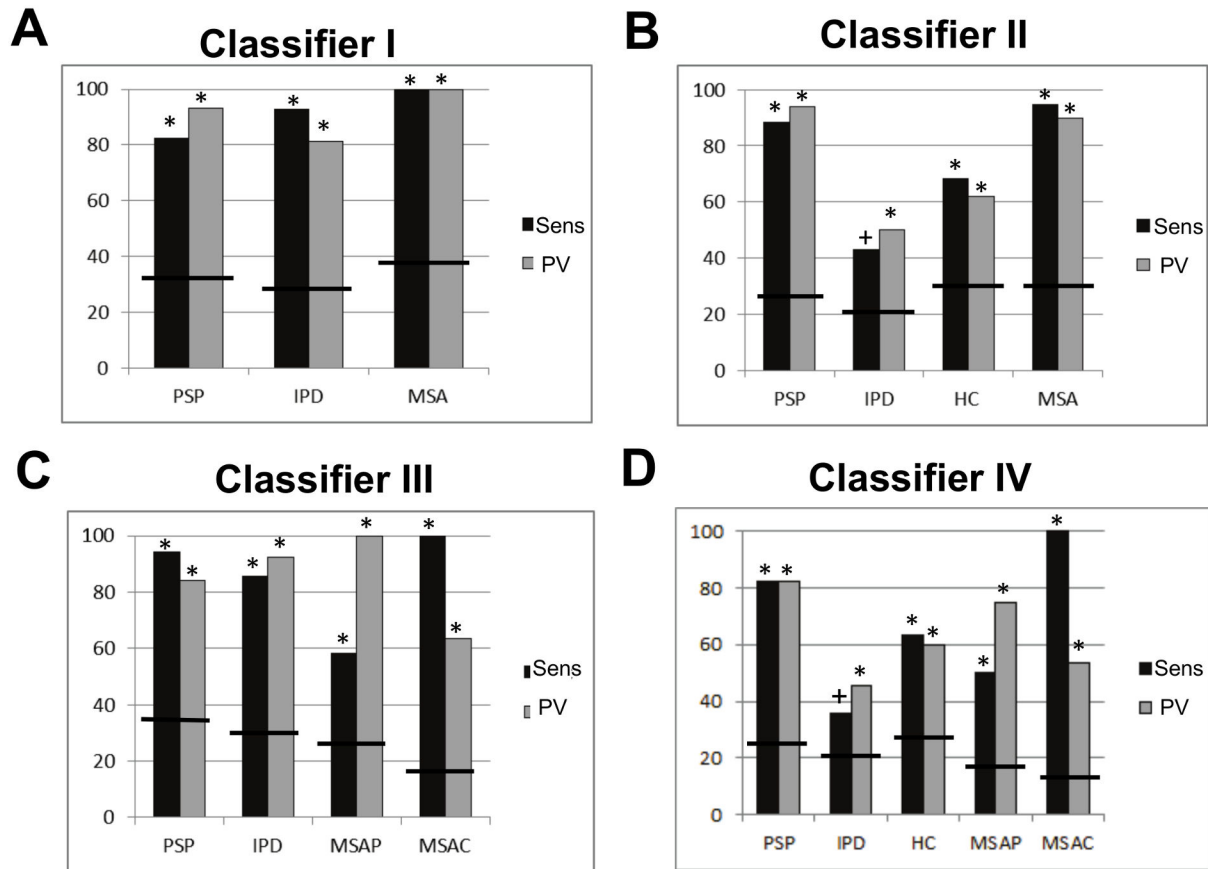
Figure 2. Sensitivity (Sens) and predictive value (PV) for each class within each diagnostic classifier based on the subcortical motor network features (classifiers I–IV in panels A-D respectively). Bars denote the chance levels determined by the proportion of samples in the training set. * = p < 0.01, # = p < 0.05 + = p < 0.1.

doi: 10.1371/journal.pone.0069237.g002

Table 2. Balanced accuracy and overall predictive value (OPV) for all classifiers trained using voxels derived from the subcortical motor network. * p < 0.01, # = p < 0.05. Values in brackets are 95% confidence intervals for the accuracies, derived by an obvious multiclass generalization of the method presented in [47].

Classifier	Classes	Region	Accuracy [95% C.I.]	OPV
I	PSP, IPD, MSA	Subcortical network	91.7%* [77.8–94.5]	91.5%*
II	PSP, IPD, HCs, MSA	Subcortical network	73,6%* [61.9–80.2]	73,9%*
III	PSP, IPD, MSA-P, MSA-C	Subcortical network	84.5%* [68.7–88.2]	85.0%*
IV	PSP, IPD, HCs, MSA-P, MSA-C	Subcortical network	66.2%* [53.7–72.8]	63.3%*

C = 100%; midbrain/brainstem: MSA-P = 66.7%, MSA-C = 100%), but the putamina were only predictive of MSA-P (MSA-P: 50.0%, MSA-C: 0.0%). The only region that discriminated

IPD from HCs with high sensitivity and PV was the midbrain/brainstem and was thus the only region that simultaneously discriminated all disease classes from one another and HCs (Figure 4). Overall, the patterns of predictive weights are congruent with the effects described above (materials S1).

Classification performance: whole-brain

While all whole-brain classifiers exceeded chance accuracy and OPV (p < 0.001), they were consistently poorer predictors than the subcortical motor network (mean difference of 12.1% accuracy and 14.3% OPV) and were also consistently poorer across classes (materials S1). Thus, they will not be considered further.

Comparison of MSA subtypes

As described, the sensitivity and PV for MSA-P were consistently higher when MSA-P and MSA-C were considered to be the same class (Table 2 Figure 2). Although the sensitivity for MSA-C was 100% for all classifiers, the PV for MSA-C was also consistently improved by considering MSA-P and MSA-C together.
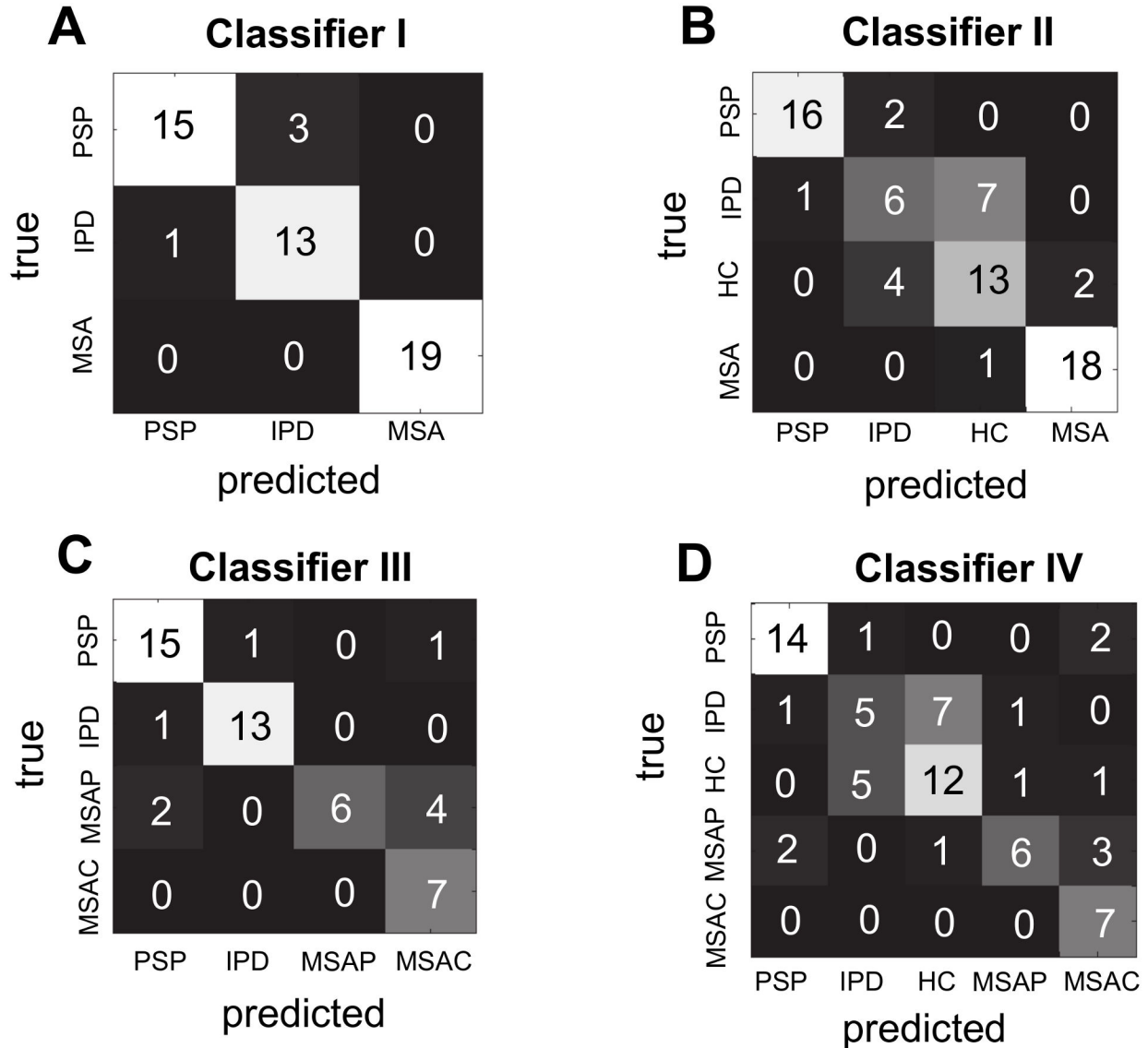
Figure 3. Confusion matrices for each diagnostic decision (classifiers I–IV in panels A-D respectively). Numbers in each cell describe the total number of predictions.

doi: 10.1371/journal.pone.0069237.g003

Discussion

In this study, we employed multi-class PR for single-subject classification of Parkinsonian disorders using structural MRI. In contrast to voxel-wise approaches that describe focal group-level effects across brain regions, PR predicts disease state at the single subject level using distributed patterns of atrophy. This provides the advantages that it is objective, fully automated and free from operator bias. We demonstrated nearly perfect diagnostic classification of PSP, MSA and IPD using a subcortical motor network. Our approach produced only four misclassifications from 50 predictions (91.7% accuracy, 91.5% OPV) and accurately discriminated all disease

classes. To our knowledge, this provides the first demonstration of accurate simultaneous discrimination between these disorders at an individual patient level using MRI measures.

All disease classes were accurately discriminated from one another with predictive performance that can be considered excellent relative to: (i) rMRI and semi-automated VBM studies [7,9,12], (ii) measures derived from DTI [15] and (iii) studies applying PR to structural MRI, to which they are most directly comparable [20,21,36]. Amongst these latter studies, one study reported accurate discrimination between typical and atypical Parkinsonian syndromes after pooling MSA and PSP but did not attempt to discriminate between PSP and MSA [20].

Table 3. Balanced accuracy and overall predictive value (OPV) for the four-class classifiers trained to discriminate PSP, IPD, HCs and MSA (Classifier II) using voxels derived from each constituent region. All regions were defined bilaterally using anatomical masks (see supplementary material). * = p < 0.01, # = p < 0.05. Values in brackets are 95% confidence intervals for the accuracies, derived by an obvious multiclass generalization of the method presented in [47].

Classifier	Classes	Region	Accuracy [95% C.I.]	OPV
II	PSP, IPD, HCs, MSA	Cerebellum	60.0%* [49.3–69.1]	60.7%*
II	PSP, IPD, HCs, MSA	Midbrain/Brainstem	71.7%* [59.2–79.1]	71.9%*
II	PSP, IPD, HCs, MSA	Caudate	38.6%* [30.8–49.6]	37.3%#
II	PSP, IPD, HCs, MSA	Putamen	46.7%* [37.0–57.6]	45.8%*
II	PSP, IPD, HCs, MSA	Pallidum	40.1%* [32.6–50.3]	36.8%*
II	PSP, IPD, HCs, MSA	Accumbens	37.1% [27.3–45.6]	32.3%

Another study aimed to discriminate MSA-P, PSP, IPD and HCs in a pair-wise manner, reporting: (i) high accuracy (66-97%) discrimination of PSP from HCs and IPD; (ii) marginal discrimination of MSA-P from HCs and IPD and (iii) no discrimination of other classes [21]. In future studies, it will also be important to validate performance of the classifier in the presence of other disorders that have similar symptoms (e.g. corticobasal degeneration), although MSA and PSP are more common than CBD, accounting for 80% of cases misdiagnosed with IPD [43,44].

An important feature of our approach is that it provides estimates of how accurately each model will make predictions for new cases, which is of direct diagnostic relevance. This was achieved through the cross-validation approach that we employed, which is well known to provide approximately unbiased estimates of the true generalizability. This provides a more appropriate assessment of diagnostic value than simply postulating a discriminatory cut-off using the same data that was used to construct the model (which yields overly optimistic estimates of generalisability).

We acknowledge that a limiting factor in our study is the modest number of patients for whom pathological confirmation of diagnosis could be obtained (eight out of 50 cases). This proportion of patients where diagnosis could be confirmed pathologically is comparable to or greater than in most previous neuroimaging studies (e.g. [7–14,20,21] and references in [15]). In all eight patients where diagnosis was pathologically confirmed, the model accurately predicted the diagnosis. In our study, lack of pathological diagnosis occurred as some patients did not consent to autopsy and some are still living. This is a problem frequently encountered in neuroimaging studies. Although the modest rate of pathological confirmation must be taken into consideration when interpreting our results, we do not believe that this invalidates our findings. Each patient had the typical clinical syndrome for their particular diagnosis, fulfilling stringent clinical diagnostic criteria. Another potential limitation is the moderate overall sample size, which motivates future replication of these findings in a larger sample. This

sample is smaller than many pattern recognition studies in other disorders (e.g. dementia), but is nevertheless substantially larger than nearly all published studies investigating Parkinsonian disorders with MRI (reviewed in [15]).

For PSP accurate predictions were derived from all subcortical regions, reflecting the known distribution of pathology in cerebellum, midbrain and basal ganglia [2]. Of these regions, predictions with the highest sensitivity and PV were derived from the midbrain/brainstem, caudate nuclei and pallidum. Midbrain atrophy is the most consistent finding in VBM studies of PSP [12,13], and atrophy of the caudate nuclei has been reported in some [13,24] but not all studies [12]. Indeed, the magnitude of focal effects in the basal ganglia were modest in relation to those in the midbrain (materials S1), but the overall pattern in each region was nevertheless highly predictive of PSP. The cerebellum was a poorer predictor of PSP than the other regions, which is surprising considering the use of SCP atrophy for identifying PSP in rMRI [7,9]. This is probably attributable to the small size of the SCP relative to the voxel size of MRI, making it less suited to detection by automated approaches, although atrophy of the decussations of the SCP (which are larger and contained within the brainstem mask) are probably more useful and were assigned high predictive weight (materials S1). However, when these single structures were considered together within the subcortical motor network, this yielded superior sensitivity and PV to every component region, (and to the whole-brain classifier), indicating that a network approach is better suited than single regions for detecting PSP.

The cerebellum and brainstem were highly predictive of MSA, in accordance with: (i) their degree of pathological involvement in MSA [3], (ii) their utility as markers in rMRI [8,9] and (iii) VBM studies that report extensive pontocerebellar damage in MSA-C and MSA-P [14]. Accordingly, the pontocerebellar degeneration we observed was widespread and severe in MSA (materials S1). Our results suggest that to optimally discriminate MSA, a focussed subcortical network containing the cerebellum, brainstem and putamen may be better suited than the more extensive subcortical network that optimally predicts PSP. The ability of the model to predict either MSA-P or MSA-C was improved when they were considered together. This suggests that the characteristics of MSA-P and MSA-C may overlap sufficiently at the network level for it to be advantageous for them to be considered together when building an analysis model for automated discrimination using MRI.

While IPD could be accurately discriminated from MSA and PSP, it was only possible to discriminate IPD from HCs using the midbrain/brainstem. This was expected, given that early- and mid-stage IPD pathology is largely restricted to the midbrain [4], and the brains of IPD patients usually appear normal in rMRI [15]. VBM studies have only reported subtle focal differences in early or mid-stage IPD relative to HCs [13,23] although more extensive cortical damage may occur in late-stage or demented IPD patients [45]. Our results accord with these findings and indicate that although midbrain/brainstem changes in IPD are subtle, they are sufficiently
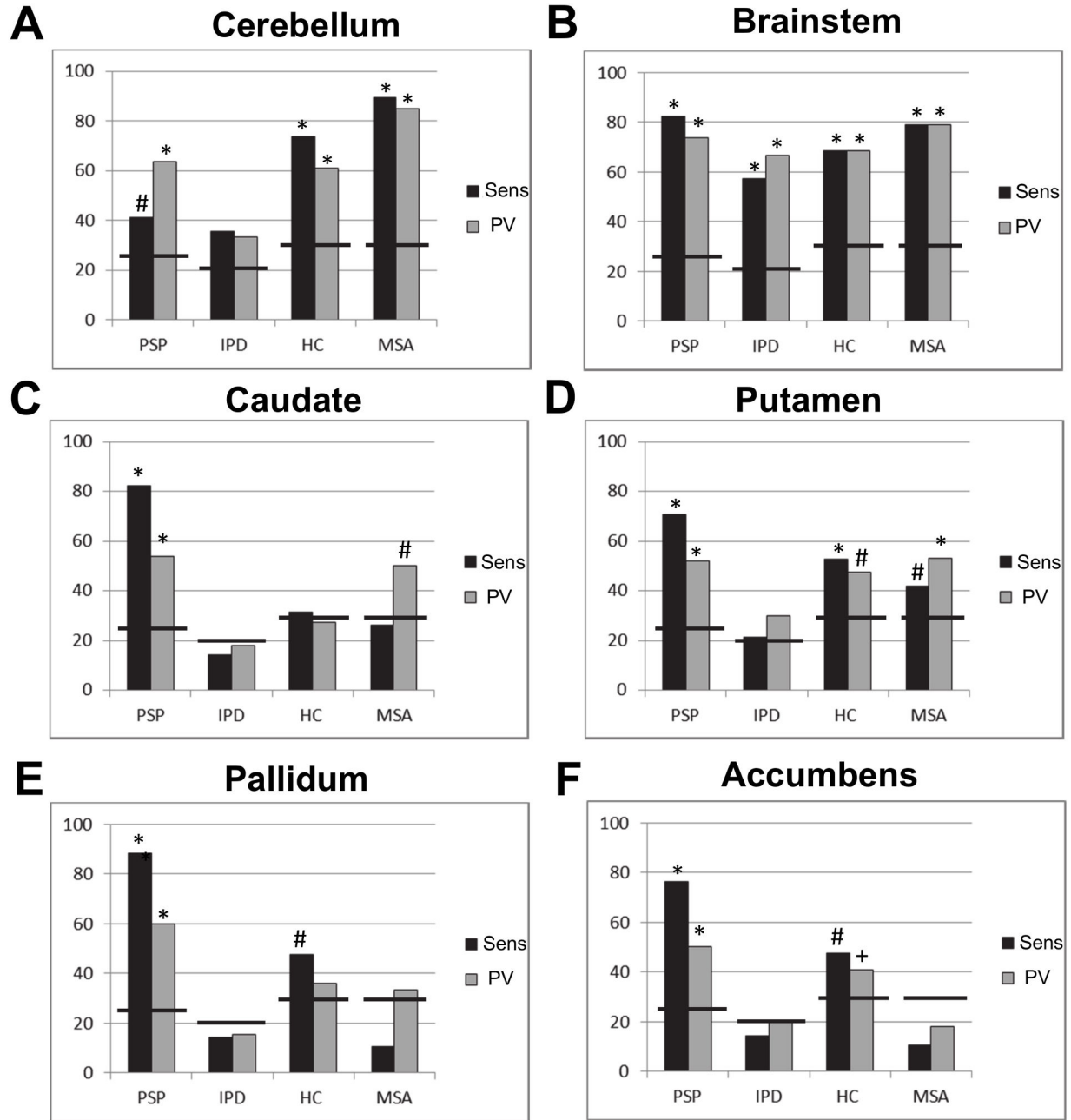
Figure 4. Sensitivity (Sens) and predictive value (PV) for each region in the subcortical motor network for the four-class classifier contrasting PSP, IPD, HC and MSA (Classifier II). A: cerebellum; B: brainstem; C: caudate; D: putamen; E: pallidum; F: accumbens. Bars denote the chance levels determined by the proportion of samples in the training set. * = p < 0.01, # = p < 0.05 + = p < 0.1.

doi: 10.1371/journal.pone.0069237.g004

informative to accurately discriminate IPD from all other classes. Accordingly, our results suggest that a region-of-interest approach restricted to midbrain/brainstem may be better suited to discriminate IPD than a network approach.

For all disorders, the whole-brain approach yielded lower performance than using only the core network. This does not

exclude the possibility that cortical pathology is predictive of any of the disorders if the component regions are more carefully specified a priori, but indicates that if anatomical hypotheses cannot be clearly formulated it is preferable to focus classification on a smaller network of core regions where degeneration is known to occur rather than employ an

exploratory classification approach. Similar findings have been reported for dementia, where PR approaches are also more accurate using a set of core regions relative to the whole brain, despite widespread cortical involvement [46]. An advantage of the multi-class approach employed here is that an independent predictive function is used to model each class, so the framework accommodates distinct sets of features for identifying each disease.

In summary, we demonstrated highly accurate, fully automated single subject classification of MSA, PSP and IPD from one another and from healthy controls using a conventional MRI sequence that could easily be obtained as part of a clinical protocol. We identified different sets of regional features optimal for predicting each disorder, which are important because (i) they define an objective set of biomarkers predictive of disease state and (ii) can guide future studies aiming to automatically classify these disorders using MRI. The next step is to validate these findings in a larger sample of patients at an earlier stage in the disease process with histological confirmation of diagnosis.

Supporting Information

Materials S1. Balanced accuracy and overall predictive value (OPV) for the four-class classifiers trained to discriminate PSP,

IPD, HCs and MSA (Classifier II) using voxels derived from each constituent region. All regions were defined bilaterally using anatomical masks (see supplementary material). * = p < 0.01, # = p < 0.05. Values in brackets are 95% confidence intervals for the accuracies, derived by an obvious multiclass generalization of the method presented in [47].
(PDF)

References

1. Litvan I, Bhatia KP, Burn DJ, Goetz CG, Lang AE et al. (2003) SIC Task Force appraisal of clinical diagnostic criteria for Parkinsonian disorders. Mov Disord 18: 467-486. doi:10.1002/mds.10459. PubMed: 12722160.
2. Hauw JJ, Daniel SE, Dickson D, Horoupian DS, Jellinger K et al. (1994) Preliminary NINDS neuropathologic criteria for Steele-Richardson-Olszewski syndrome (progressive supranuclear palsy). Neurology 44: 2015-2019. doi:10.1212/WNL.44.11.2015. PubMed: 7969952.
3. Papp MI, Lantos PL (1994) The distribution of oligodendroglial inclusions in multiple system atrophy and its relevance to clinical symptomatology. Brain 117: 235-243. doi:10.1093/brain/117.2.235. PubMed: 8186951.
4. Braak H, Braak E (2000) Pathoanatomy of Parkinson's disease. J Neurol 247: 3-10. doi:10.1007/s004150050002. PubMed: 10701890.
5. Deuschl G, Schade-Brittinger C, Krack P, Volkmann J, Schäfer H et al. (2006) A randomized trial of deep-brain stimulation for Parkinson's disease. N Engl J Med 355: 896-908. doi:10.1056/NEJMoa060281. PubMed: 16943402.
6. Shih LC, Tarsy D (2007) Deep brain stimulation for the treatment of atypical parkinsonism. Mov Disord 22: 2149-2155. doi:10.1002/mds.21648. PubMed: 17659638.
7. Paviour DC, Price SL, Stevens JM, Lees AJ, Fox NC (2005) Quantitative MRI measurement of superior cerebellar peduncle in progressive supranuclear palsy. Neurology 64: 675-679. doi:10.1212/01.WNL.0000151854.85743.C7. PubMed: 15728291.
8. Nicoletti G, Fera F, Condino F, Auteri W, Gallo O et al. (2006) MR imaging of middle cerebellar peduncle width: Differentiation of multiple system atrophy from Parkinson disease. Radiology 239: 825-830. doi:10.1148/radiol.2393050459. PubMed: 16714464.
9. Quattrone A, Nicoletti G, Messina D, Fera F, Condino F et al. (2008) MR imaging index for differentiation of progressive supranuclear palsy from Parkinson disease and the Parkinson variant of multiple system atrophy. Radiology 246: 214-221. doi:10.1002/mds.10502. PubMed: 17991785.
10. Rolland Y, Verin M, Payan CA, Duchesne S, Kraft E et al. (2011) A new MRI rating scale for progressive supranuclear palsy and multiple system atrophy: validity and reliability. J Neurol Neurosurg, Psychiatry 82: 1025-1032. doi:10.1136/jnnp.2010.214890.
11. Brenneis C, Seppi K, Schocke MF, Müller J, Luginger E et al. (2003) Voxel-based morphometry detects cortical atrophy in the Parkinson variant of multiple system atrophy. Mov Disord 18: 1132-1138. doi:10.1002/mds.10502. PubMed: 14534916.

12. Price S, Paviour D, Scahill R, Stevens J, Rossor M et al. (2004) Voxel-based morphometry detects patterns of atrophy that help differentiate progressive supranuclear palsy and Parkinson's disease. NeuroImage 23: 663-669. doi:10.1016/j.neuroimage.2004.06.013. PubMed: 15488416.
13. Cordato NJ, Duggins AJ, Halliday GM, Morris JGL, Pantelis C (2005) Clinical deficits correlate with regional cerebral atrophy in progressive supranuclear palsy. Brain 128: 1259-1266. doi:10.1093/brain/awh508. PubMed: 15843423.
14. Minnerop M, Specht K, Ruhlmann J, Schimke N, Abele M et al. (2007) Voxel-based morphometry and voxel-based relaxometry in multiple system atrophy - A comparison between clinical subtypes and correlations with clinical parameters. NeuroImage 36: 1086-1095. doi:10.1016/j.neuroimage.2007.04.028. PubMed: 17512219.
15. Mahlknecht P, Hotter A, Hussl A, Esterhammer R, Schocke M et al. (2010) Significance of MRI in Diagnosis and Differential Diagnosis of Parkinson's Disease. Neurodegener Dis 7: 300-318. doi:10.1159/000314495. PubMed: 20616565.
16. Orrù G, Pettersson-Yeo W, Marquand AF, Sartori G, Mechelli A (2012) Using Support Vector Machine to identify imaging biomarkers of neurological and psychiatric disease: A critical review. Neurosci Biobehav Rev 36: 1140-1152. doi:10.1016/j.neubiorev.2012.01.004. PubMed: 22305994.
17. Klöppel S, Abdulkadir A, Jack CR Jr., Koutsouleris N, Mourão-Miranda J et al. (2012) Diagnostic neuroimaging across diseases. NeuroImage 61: 457-463. doi:10.1016/j.neuroimage.2011.11.002. PubMed: 22094642.
18. Klöppel S, Stonnington CM, Chu C, Draganski B, Scahill RI et al. (2008) Automatic classification of MR scans in Alzheimers disease. Brain 131: 681-689. doi:10.1093/brain/awm319. PubMed: 18202106.
19. Klöppel S, Chu C, Tan GC, Draganski B, Johnson H et al. (2009) Automatic detection of preclinical neurodegeneration Presymptomatic Huntington disease. Neurology 72: 426-431. doi:10.1212/01.wnl.0000341768.28646.b6. PubMed: 19188573.
20. Duchesne S, Rolland Y, Vérin M (2009) Automated Computer Differential Classification in Parkinsonian Syndromes via Pattern Analysis on MRI. Acad Radiol 16: 61-70. doi:10.1016/j.acra.2008.05.024. PubMed: 19064213.
21. Focke NK, Helms G, Scheewe S, Pantel PM, Bachmann CG et al. (2011) Individual Voxel-Based Subtype Prediction can Differentiate Progressive Supranuclear Palsy from Idiopathic Parkinson Syndrome

and Healthy Controls. Hum Brain Mapp 32: 1905-1915. doi:10.1002/hbm.21161. PubMed: 21246668.

22. Ozawa T, Paviour D, Quinn NP, Josephs KA, Sangha H et al. (2004) The spectrum of pathological involvement of the striatonigral and olivopontocerebellar systems in multiple system atrophy: clinicopathological correlations. Brain 127: 2657-2671. doi:10.1093/brain/awh303. PubMed: 15509623.

23. Jubault T, Brambati SM, Degroot C, Kullmann B, Strafella AP et al. (2009) Regional Brain Stem Atrophy in Idiopathic Parkinson's Disease Detected by Anatomical MRI. PLOS ONE 4: e8247. PubMed: 20011063.

24. Josephs KA, Whitwell JL, Dickson DW, Boeve BF, Knopman DS et al. (2008) Voxel-based morphometry in autopsy proven PSP and CBD. Neurobiology of Aging 29: 280-289. doi:10.1016/j.neurobiolaging.2006.09.019. PubMed: 17097770.

25. Hughes AJ, Daniel SE, Kilford L, Lees AJ (1992) Accuracy of clinical-diagnosis of idiopathic Parkinson's disease - a clinicopathological study of 100 cases. J Neurol Neurosurg, Psychiatry 55: 181-184. doi:10.1136/jnnp.55.3.181.

26. Litvan I, Agid Y, Calne D, Campbell G, Dubois B et al. (1996) Clinical research criteria for the diagnosis of progressive supranuclear palsy. Steele: Richardson-Olszewski syndrome): Report of the NINDS-SPSP International Workshop. Neurology 47: 1-9

27. Gilman S, Low PA, Quinn N, Albanese A, Ben-Shlomo Y et al. (1999) Consensus statement on the diagnosis of multiple system atrophy. J Neurol Sci 163: 94-98. doi:10.1016/S0022-510X(98)00304-9. PubMed: 10223419.

28. Blain CRV, Barker GJ, Jarosz JM, Coyle NA, Landau S et al. (2006) Measuring brain stem and cerebellar damage in parkinsonian syndromes using diffusion tensor MRI. Neurology 67: 2199-2205. doi:10.1212/01.wnl.0000249307.59950.f8. PubMed: 17190944.

29. Bensimon G, Ludolph A, Agid Y, Vidailhet M, Payan C et al. (2009) Riluzole treatment, survival and diagnostic criteria in Parkinson plus disorders: The NNIPPS Study. Brain 132: 156-171. PubMed: 19029129.

30. Williams DR, de Silva R, Paviour DC, Pittman A, Watt HC et al. (2005) Characteristics of two distinct clinical phenotypes in pathologically proven progressive supranuclear palsy: Richardson's syndrome and PSP-parkinsonism. Brain 128: 1247–58. PubMed: 15788542.

31. Gilman S, Wenning GK, Low PA, Brooks DJ, Mathias CJ et al. (2008) Second consensus statement on the diagnosis of multiple system atrophy. Neurology 71: 670–6. PubMed: 18725592.

32. Hoehn MM, Yahr MD (1967) Parkinsonism - onset, progression and mortality. Neurology 17: 427–442&. doi:10.1212/WNL.17.5.427. PubMed: 6067254

33. Schwab RS, England AC (1968) Projection techniques for evaluating surgery in Parkinson's Disease. Third Symposium on Parkinson's Disease, Royal College of Surgeons in Edinburgh, May. E S Livingstone Ltd 20-22: 152-157.

34. Payan CAM, Vidailhet M, Lacomblez L, Viallet F, Borg M et al. (2002) Neuroprotection and Natural History in Parkinson Plus Syndromes (NNIPPS): Construction and validation of a functional scale for disease progression assessment in Parkinson Plus Syndromes, Progressive Supranuclear Palsy (PSP) and Multiple System Atrophy (MSA). Mov Disord 17: S256-S256.

35. Bronte-Stewart HM, Minn AY, Rodrigues K, Buckley EL, Nashner LM (2002) Postural instability in idiopathic Parkinson's disease: the role of medication and unilateral pallidotomy. Brain 125: 2100-2114. doi:10.1093/brain/awf207. PubMed: 12183355.

36. Filippone M, Marquand AF, Blain CRV, Williams SCR, Mourão-Miranda J et al. (2012) Probabilistic prediction of neurological disorders with a statistical assessment of neuroimaging data modalities. Annals Appl Statistics, 6(4): 1883-1905. doi:10.1214/12-AOAS562.

37. Singh N, Fletcher PT, Preston JS, Ha L, King R et al. (2010) Multivariate Statistical Analysis of Deformation Momenta Relating Anatomical Shape to Neuropsychological Measures. Medical Images Computing Computer-Assist Interv-MICCAI 6363: 529-537. PubMed: 20879441.

38. Marquand AF, Mourão-Miranda J, Brammer MJ, Cleare AJ, Fu CHY (2008) Neuroanatomy of verbal working memory as a diagnostic biomarker for depression. Neuroreport 19: 1507-1511. doi:10.1097/WNR.0b013e328310425e. PubMed: 18797307.

39. Mourão-Miranda J, Oliveira L, Ladouceur CD, Marquand A, Brammer M et al. (2012) Pattern Recognition and Functional Neuroimaging Help to Discriminate Healthy Adolescents at Risk for Mood Disorders from Low Risk Adolescents. PLOS ONE 7: e29482. PubMed: 22355302.

40. Marquand AF, O'Daly OG, De Simoni S, Alsop DC, Maguire RP et al. (2012) Dissociable effects of methylphenidate, atomoxetine and placebo on regional cerebral blood flow in healthy volunteers at rest: A multi-class pattern recognition approach. Neuroimage 60: 1015–24. PubMed: 22266414.

41. Mourão-Miranda J, Almeida JR, Hassel S, de Oliveira L, Versace A et al. (2012) Pattern recognition analyses of brain activation elicited by happy and neutral faces in unipolar and bipolar depression. Bipolar Disord 14: 451-460. doi:10.1111/j.1399-5618.2012.01019.x. PubMed: 22631624.

42. Marquand AF, De Simoni S, O'Daly OG, Williams SC, Mourão-Miranda J et al. (2011) Pattern classification of working memory networks reveals differential effects of methylphenidate, atomoxetine, and placebo in healthy volunteers. Neuropsychopharmacology 36: 1237-1247. doi:10.1038/npp.2011.9. PubMed: 21346736.

43. Hughes AJ, Ben-Shlomo Y, Daniel SE, Lees AJ (2001) What features improve the accuracy of clinical diagnosis in Parkinson's disease: A clinicopathologic study. Neurology 57: 1142–6. PubMed: 1603339.

44. Hughes AJ, Daniel SE, Ben-Shlomo Y, Lees AJ (2002) The accuracy of diagnosis of parkinsonian syndromes in a specialist movement disorder service. Brain 125.

45. Burton EJ, McKeith IG, Burn DJ, Williams ED, O'Brien JT (2004) Cerebral atrophy in Parkinson's disease with and without dementia: a comparison with Alzheimer's disease, dementia with Lewy bodies and controls. Brain 127: 791-800. doi:10.1093/brain/awh088. PubMed: 14749292.

46. Chu C, Hsu A-L, Chou K-H, Bandettini P, Lin C et al. (2012) Does feature selection improve classification accuracy? Impact of sample size and feature selection on classification using anatomical magnetic resonance images. Neuroimage 60: 59-70. doi:10.1016/j.neuroimage.2011.11.066. PubMed: 22166797.

47. Brodersen KH, Cheng Soon O, Stephan KE, Buhmann JM (2010) The balanced accuracy and its posterior distribution. Proceedings of the 20th International Conference on Pattern Recognition(ICPR 2010).