NeuroImage: Clinical 27 (2020) 102272

Contents lists available at ScienceDirect

NeuroImage: Clinical

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

Striatal shape alteration as a staging biomarker for Parkinson's Disease☆

Maxime Peralta[a], John S.H. Baxter[a], Ali R. Khan[b], Claire Haegelen[a,c], Pierre Jannin[a,*]

[a] INSERM, LTSI – UMR 1099, University of Rennes, Rennes, France
[b] Imaging Research Laboratories, Robarts Research institute, Western University, London, Canada
[c] CHU Rennes, Rennes, France

ARTICLE INFO

Keywords:
Parkinson's disease
Morphometric biomarkers
Machine learning
Staging biomarker
Medical imaging

ABSTRACT

Parkinson's Disease provokes alterations of subcortical deep gray matter, leading to subtle changes in the shape of several subcortical structures even before the manifestation of motor and non-motor clinical symptoms. We used an automated registration and segmentation pipeline to measure this structural alteration in one early and one advanced Parkinson's Disease (PD) cohorts, one prodromal stage cohort and one healthy control cohort. These structural alterations are then passed to a machine learning pipeline to classify these populations. Our workflow is able to distinguish different stages of PD based solely on shape analysis of the bilateral caudate nucleus and putamen, with balanced accuracies in the range of 59% to 85%. Furthermore, we compared the significance of each of these subcortical structure, compared the performances of different classifiers on this task, thus quantifying the informativeness of striatal shape alteration as a staging bio-marker for PD.

1. Introduction

Parkinson's Disease (PD) is a neurodegenerative disease resulting from the degeneration of the dopamine-producing areas of the basal ganglia. Although originally perceived as a primarily motor disease, its symptoms are now known to extend far beyond motricity including cognitive and neuropsychiatric symptoms (Pfeiffer, 2016). Some non-motor clinical symptoms may be detectable before the appearance of the more distinctive motor symptoms (Mahlknecht et al., 2015). This prodromal stage of PD is crucial for medical treatment but difficult to diagnose. There are a variety of treatments for PD, including pharmocological treatment such as the dopamine precursor levodopa or interventional treatments such as Deep Brain Stimulation (DBS), which are proposed to obstruct disease progression and enhance the patient's quality of life. The choice of treatment remains under debate, and the high heterogeneity of the disease (Thenganatt and Jankovic, 2014) renders the space of treatment options highly variable and patient-specific (Connolly et al., 2014; Limousin et al., 2008). Finally, there are unanswered questions regarding the long term effects of these treatments and the evolution of the disease (Limousin and Foltynie, 2019). For all of these reasons, PD is unanimously considered in pressing need of biomarkers both for improved diagnosis and treatment monitoring (Delenclos et al., 2015).

In terms of the underlying neuroanatomy of Parkinson's disease symptomology, promising results have been found in the bilateral putamen and caudate (the two major components of the dorsal striatum). It is well known that these structures are central to the progression of PD, which greatly alters their behavior. The lack of dopamine in the putamen as a result of PD is considered the direct cause of motor dysfunction, while the lack of dopamine in the caudate is more related to alterations in cognitive function (Kish et al., 1988). Griffiths et al. (1994) reported an alteration of the density of some neurotransmitter receptors in both of these structures in post mortem brains of PD subjects. Kish et al. (1988) observed a large loss of dopamine in both putamen and caudate nucleus, with reductions of over 99% in caudal portions of the putamen.

In clinic, the diagnosis of PD, staging of the disease, and selection of the treatment are made primarily using clinical biomarkers such as UPDRS scores or Hoehn and Yahr scale. Although brain alteration is prior to any clinical symptom, brain morphometry via Magnetic Resonance Imaging (MRI) is considered a marginal source of information in clinic PD (Péran et al., 2018; Paul, 2016). With increasing availability and resolution of MRI, the research community is moving towards finding reliable imaging biomarkers for the diagnosis and monitoring of PD.

☆ This document is the results of the research project funded by the Fondation pour la Recherche Médicale.
* Corresponding author.
E-mail address: pierre.jannin@univ-rennes1.fr (P. Jannin).

https://doi.org/10.1016/j.nicl.2020.102272
Received 20 December 2019; Received in revised form 20 April 2020; Accepted 21 April 2020
Available online 19 May 2020

1.1. Subcortical morphological biomarkers

Shape alterations in subcortical structures aside from volume or thickness reduction has long been identified as a potential area of analysis. Voxel Based Morphometry (VBM) has been for a long time the favoured method due to its fine granularity, although with defined limitations (Loretxu Bergouignan et al., 2009; Davatzikos, 2004). Most modern methods fall into the category of Surface-based Morphometry (SBM) which is a two-stage process. The first stage involves the segmentation and registration of subcortical structures from the MRI into a template space, and the second stage involves the representation and analysis of the segmented surface's shape with respect to that of the population. While there are a variety of segmentation methods in the first stage, the second almost unanimously extracts vertex-wise boundary displacements on a template mesh, representing each surface as a large, but constant-sized, vector. Among the most used, the Bayesian Appearance Model (BAM) is now a native part of the FSL library using the FIRST implementation. This method proposes to add a Bayesian framework to the Active Appearance Model (AAM), which incorporates intensity information on top of the shape deformation model. The authors claims that this Bayesian framework allows to capture more subtle shape deformation than the other methods, even when the training data amount is low. Other widely used methods for shape analysis are single-atlas segmentation or multi-atlas label fusion, followed by Large Deformation Diffeomorphic Metric Mapping (LDDMM) (Faisal Beg et al., 2005) for registration.

Amongst the techniques explored to analyze the brain morphometric variations, shape analysis has proved itself more reliable and consistent than volume analysis, thickness analysis, and voxel-based morphometry (Péran et al., 2018; Berner et al., 2019; van den Bogaard et al., 2011; Gerig et al., 2001; Menke et al., 2014; Tang et al., 2014; Wade et al., 2015) and is considered as a promising source of insight into various neurological and psychiatric disorders.

1.2. Subcortical shape analysis in Parkinson's Disease

Several studies have tried to evaluate the significance of subcortical shape displacements as a bio-marker for PD, leading to often contradictory results with the literature showing the impact of cohort and methodology on the relevance of subcortical structures as diagnosis biomarkers. Garg et al. (2015) show that the morphology of the bilateral caudate nucleus as well as the putamen is discriminant in Nemmi et al. (2015) found significant results only in the left caudate and putamen, whereas Owens-Walton et al. (2018) didn't find any correlation in shape information outside of gross volumetric differences.

Some tried to use subcortical shape displacement for other purposes than classifying PD patients from healthy controls. Mak et al. (2014) and Foo et al. (2017) attempted to correlate subcortical shape displacement with Mild Control Impairment (MCI) in PD, the first not showing significant results while the second found significant information only in the left caudate. Both Nemmi et al. (2015) and Garg et al. (2015) tried to correlate subcortical shape displacement with disease severity (using left, right and global UPDRS and disease duration for the first, and global UPDRS only for the second). The only significant result obtained, on these two studies, was between the right UPDRS and the left putamen shape (Nemmi et al., 2015). Owens-Walton et al. (2018) failed to show discriminating shape features in the putamen and caudate nucleus to classify different stages of PD, getting only significant results through volume analysis. Ultimately, the literature shows that correlation between specific clinical symptoms and subcortical shape displacements in PD remains a difficult task.

1.3. Contributions

To the best of our knowledge, no cross-validated studies has been done to quantify the predictive power of subcortical shape displacements in PD. Moreover, little work has been done to compare the relevance of classifiers and to compare the informativeness of the structures. Finally, most of the studies focus on diagnosing PD against Healthy Control (HC), without differentiating intermediate stages, and the need to assess the relevance of morphometric biomarkers to diagnose PD prodromal phase has been pointed out (Garg et al., 2015).

Our proposed method is the first fully-automated, cross-validated pipeline to classify different stages of PD. This data-driven pipeline uses the state of the art of machine learning and data analysis to quantify the relevance of shape displacements of putamen and caudate nucleus as diagnostic and staging biomarkers, benchmarking different classifiers and structures. We have extended clinical knowledge by investigating subcortical shape displacements and determining those which are relevant for diagnosing prodromal stages of PD, therefore responds to the necessity raised by Péran et al. (2018) to extend the morphometric biomarkers research on prodromal stage of the disease.

We also shown that the MDS-UPDRS3 score of PD patients can be predicted from subcortical shape displacements with our method, showing weak, yet significant (at $p < 0.001$) results.

2. Materials and methods

2.1. Proposed method

The pipeline proposed is composed of three steps, as presented in Fig. 1, which include the extraction of displacement vectors, their compression, and finally the classification of the patient based on this information.

The first step is to extract the displacement fields of the four striatal structures from T1-weighted MRI sequences, compared to a fixed atlas template described in Section 2.3 using the method presented in Section 2.1.1. The output of this process is a vector of vertex-wise signed displacement magnitudes, with positive values indicating displacement outwards from the template surface and negative indicating an inwards displacement. We chose not to analyze the full displacement vectors themselves, as we hypothesize that taking into account the direction of the vectors would not add substantial information, while increasing the vector size by a factor of three.

As the vector of displacements magnitudes is high dimensional (6362, 6391, 6490, and 6510 elements for left and right caudate and left and right putamen respectively), a compression step was performed prior to the final classification step. Each structure has been compressed through Principal Component Analysis (PCA). The number of principal component kept has been chosen through Hyper-parameters Optimization (HPO), as explained in Section 2.6.

Finally, a classification is performed using the concatenation of the compressed displacements magnitudes vectors as input. These methods are explained in more detail in the subsequent sections.

2.1.1. Image processing pipeline

We used a single-atlas based segmentation followed by LDDMM registration and label propagation to automatically extract subcortical shape displacements. More details about this image processing pipeline can be found in Khan et al. (2019), and the full scripts have been made available by the authors (https://github.com/khanlab/diffparc-sumo).

The original version of this method was proposed by Khan et al. (2008) and shown a high degree of robustness for the segmentation of the caudate nucleus and putamen with a Dice overlap measure of 81% and 83%. As the reliability of this method has already been proven many times (Garg et al., 2015; Khan et al., 2019; Khan et al., 2008; Garg et al., 2014; Wang et al., 2009; Ansari, 2010), we did not perform any extensive segmentation quality check at this stage.

To summarize, this pipeline first registered the T1-weighted MRI scan into MNI152NLin2009cAsym space. Then, a Region of Interest (RoI) around the desired structure of interest on T1-weighted MRI scans

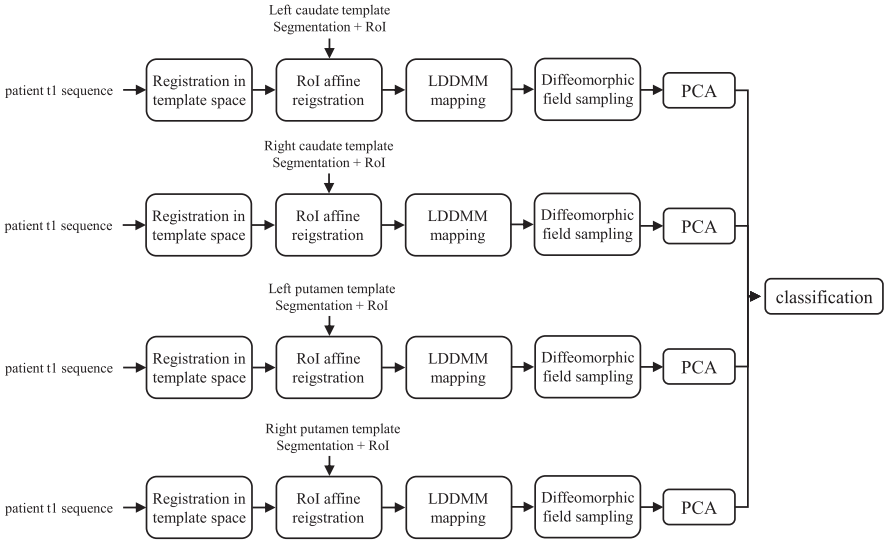

Fig. 1. Pipeline proposed and tested in this study.

was coarsely aligned with a template segmentation using an initial affine registration step. Finally, this mapping is deformably refined using LDDMM. The output of this process is a smooth diffeomorphic vector field, which can be sampled at the surface vertex locations in order to create the surface displacement vector. The magnitude of each vertex displacement is then saved for each surface.

2.1.2. Compression

We compressed displacements vectors through PCA. By keeping only the first few principal components which are orthogonal and therefore decorrelated, the data is compressed while keeping most of the information. On top of that, it often eliminates noise in the data, as noise (assuming it is of lower variation than the data and is, independent of it) is often relegated to the least-significant principal components. This triple advantage of compression, decorrelation, and denoising makes PCA useful for downstream statistical or machine learning analysis, especially when the original dimensionality of the data is high compared to the number of available samples.

2.1.3. Classification

We used and compared four different classifiers in this study: Support Vector Machine (SVM) with linear and radial basis kernels, Random Forests (RF) and Ensemble Learning (EL) through stacking classifiers. Stacking classifiers combines the strengths of different classifiers. The meta-classifier may be trained solely on the predictions of the base classifiers or to rather extend the original pool of features with them. Although the former is simpler and has much lower parameterization, the latter allows the meta-classifier to use the underlying data to determine which classifiers are likely to be more accurate for any particular datapoint. In this study, we chose the later option in order to give the meta-classifiers the theoretical ability to dynamically adapt the weight of base-classifier, depending on the inputs. We used logistic regression as a meta-classifier.

2.2. Data

2.2.1. Cohort description

Data for this article come from two databases. The first database consists of PD patients from the cohort of DBS patients recruited at Rennes University Hospital (from now refered as 'DBS PD'). These patients are all candidates for DBS, indicating that they experience sufficiently advanced motor symptoms to warrant such an intervention and are thus in an advanced phase of the disease. All patients of this cohort have been informed and gave their consent to be included in this study, which have been approved the institution's ethics committee (trial code "35RC19_4001_PSCP"). The second database is derived from the Parkinson's Progression Markers Initiative (PPMI), which is a program sponsored by the Michael J. Fox Foundation for Parkinson's Research. It is an observational clinical study which tracks cohorts of subjects with different forms of Parkinson disease for up to 8 years, with the goal of identifying biomarkers of disease progression using MR imaging, biologic sampling as well as clinical and behavioural assessments.

The PPMI database is used to define three separate cohorts. The first consists of HC subjects without any neurologic disorder and who do not have an immediate relative with PD. The second is a prodromal stage cohort. Prodromal PD patients are at risk of developing PD and display some characteristic clinical symptoms but not the more diagnostic motor symptoms (Mahlknecht et al., 2015). The third is the early PD cohort, composed of subjects with a diagnosis of PD for two years or less, and not taking PD medications.

Fig. 2 shows the normalized distribution of UPDRS3 score, the main motor assessment for PD, for the different cohorts. The UPDRS3 score used for the DBS PD cohort has been converted to the MDS-UPDR3 score used in PPMI with the equation proposed by Hentz et al. (2015). The three PPMI cohorts follow well-defined exponential or Gaussian distributions, whereas the Rennes cohort is more heterogeneous. This heterogeneity is expected as DBS is now sometimes proposed quite early during the disease at the Rennes University Hospital. More information about the inclusion criteria of the PPMI cohorts can be found on the PPMI study protocol, following the link https://www.ppmi-info.org/study-design/research-documents-and-sops/.

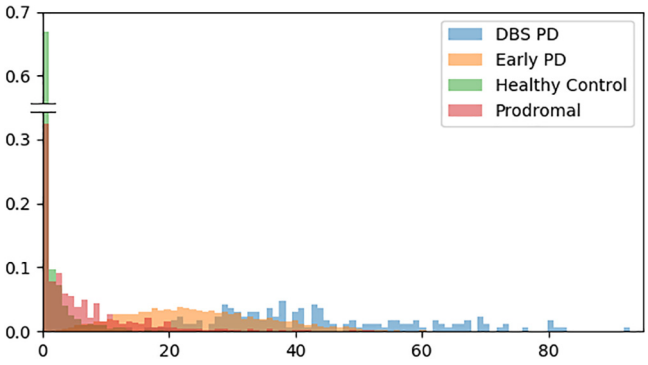

Fig. 2. UPDRS-3 normalized distribution of the cohorts used in this study.

Table 1
Statistics of the cohorts used in this study.

Cohort	F/M (total)	Age (range)	Mean UPDRS (range)
HC	64/113 (177)	67.8 ± 11.2 (39–90)	1.73 ± 3.35 (0–34)
Prodromal	9/32 (41)	74.9 ± 6.93 (56–91)	6.73 ± 8.60 (0–52)
Early PD	127/241 (368)	68.4 ± 9.73 (40–98)	25.4 ± 11.7 (0–90)
DBS PD	76/104 (180)	65.1 ± 9.48 (26–85)	43.4 ± 16.1 (14–92)

Table 1 shows the UPDRS mean and standard deviation of each cohort. From this, one can observe a natural progression of the disease in terms of motor symptoms: first healthy, then prodromal stage, then early-stage Parkinson and finally late-stage Parkinson. One can also note that the standard deviation of each cohorts is quite high, indicating that this information alone is not optimal and underscoring the importance of finding other biomarkers.

2.2.2. Image acquisition

All patients from the Rennes database had one preoperative 3T T1-weighted MRI scan (1 mm × 1 mm × 1 mm, Philips Medical Systems). All sequences were acquired prior to DBS electrode placement.

For the PPMI database, all T1-weighted sequences (e.g. MPRAGE or SPGR) were required to have a total scan time between 20 and 30 min and to have a slice thickness of 1.5 mm or less with no interslice gap.

2.3. Atlas

Probabilistic segmentation of striatal structures were obtained with the MNI PD25 atlas (Xiao et al., 2017). This atlas was built by averaging 3T MRI scans (T1w (FLASH and MPRAGE), T2*w, T1-T2* fusion, phase, and an R2* map) of 25 PD patients, making it an atlas of choice for studying PD patients in the MNI152NLin2009cAsym space. Eight subcortical structures have been segmented in this atlas, including the caudate nucleus and the putamen. This atlas is freely available through the link http://nist.mni.mcgill.ca/?p = 1209.

As noted in Section 1.0.2, the striatum has been of particular interest in morphological image analysis for PD. Thus, the segmentation of the left and right caudate and putamen were extracted for use in our study.

2.4. Accuracy and loss metrics

As cohort sizes are uneven, we used balanced accuracy (BACC) (Eq. 1) as a classification performance and comparison metric. This balanced accuracy is necessary to prevent systemic bias towards the larger cohort during the classification step. For compression, we used the reconstruction mean standard error as a compression performance and comparison metric. Each classifier was trained with class weighting, giving a weight to training sample inversely proportional to the representation of the belonging class in the training set.

$$BACC = \frac{1}{2}\left(\frac{TP}{P} + \frac{TN}{N}\right) \quad (1)$$

2.5. Training and validation

For each test, we used a stratified 10-fold Cross-Validation (CV). Each model has been trained 10 times separately, using one fold as a validation data and the remaining nine as training data. Each model has been trained and evaluated with the same folds, removing this as a potential source of variation. Additionally, each fold contains approximately the same number of samples from each cohort. In the event that a patient has multiple MRI acquisitions, all the acquisitions are assigned to the same fold, in order not to ensure the folds are independent of each other and there is no training-evaluation set corruption. To address the issue of class-balance in training, each classifier has been trained weighting samples by the inverse of the class size.

2.6. Hyper-parameter optimization

The hyper-parameters of each classifier have been optimized for each CV fold through Bayesian optimization using a Gaussian processes as a surrogate model with expected improvement as the criterion. The number of points tested was the number of hyper-parameters to optimize squared plus one.

2.7. Statistical analysis

The resulting BACC's were analyzed statistically using multi-factorial ANOVA in order to estimate the effect size and possible contributions of different classifiers and combinations of structures, our literature review suggesting these as potential sources of variability.

2.8. Software environment

The Scikit-learn implementation of PCA as well as each classifier was used, and the Python code-base used to perform the different experiments is available at https://github.com/m-prl/ParDi. Statistical analysis was performed using IBM SPSS Statistics.

3. Results

3.1. Compression performance

Fig. 3 presents the PCA compression mean squared errors for all four structures, and with different numbers of components kept.

3.2. Classification results

We performed a total of 1200 binary classifications (10 folds times 6 cohort-pairs times 5 combinations of structures times 4 different classification algorithms) each involving an independent hyper-parameter optimization process. We reported the best balanced accuracy configuration of 10-fold classification. Table 2 presents the multi-factorial ANOVA analysis of the classification results, showing that every variable contributed to the results with a high level of significance.

As expected, the Problem (pair of cohorts being distinguished) factor had a more defined effect over the Structs (the combination of

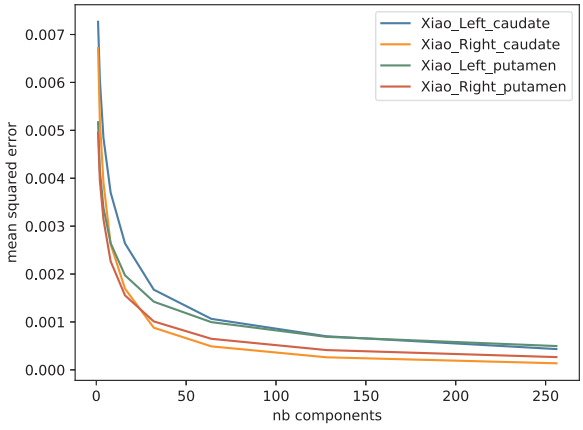

Fig. 3. Reconstruction mean squared error of PCA compression on test set for left caudate nucleus (blue), right caudate nucleus (orange), left putamen (green) and right putamen (red), with various number of components kept. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 2
Multivariate ANOVA test of the BACC across methods organized by problem, combination of structures used, and classification algorithm.

Source	T.III SS	df	Mean Sqr	F	Sig.	ηp^2
Corr. Model	7.31[a]	12	0.61	260.7	0.00	0.81
Problem	6.80	5	1.4	832.5	0.00	0.78
Structs	0.31	4	0.08	47.12	0.00	0.14
Structs*Problem	0.52	20	0.03	15.93	0.00	0.22
Algo	0.21	3	0.07	42.11	0.00	0.10
Fold	0.36	9	0.04	24.66	0.00	0.16
Error	1.89	1158	0.02			
Corr. Total	10.08	1199				

[a] R Squared = 0.812 (Adjusted R Squared = 0.806)

structures used for classification) and Algo (classification algorithm) factors. We can also note the significance of the Structure*Problem factor, indicating that the structures of interest vary through the stage of the disease.

In order to further investigate the effects of the structures and classifiers used, follow-up analysis using Tukey's Honestly Significant Difference (HSD) test was performed. This test partitions the values for each factor (e.g. Structure, Classifier, etc…) into a series of clusters that are statistically significantly different from each other. This allows for us to infer which structures or classifiers have improved performance given that significance under the ANOVA test.

3.2.1. Classification performance between problems

In order to develop a reference of the relative difficulty between the problems, a preliminary classification using only the clinical symptomatology was performed. Table 3 presents the classification metrics obtained on the training base for the 6 different binary classification problems, based solely on the MDS-UPDRS3 using a naive Bayes classifier. It is to note that the dataset is heavily class unbalanced between the DBS cohort and the others, as in the PPMI program the MDS-UPDRS3 is assessed at multiple visits, thus explaining the low f1 scores and sensitivity of the last three rows of the Table. The accuracy is expected to be higher using clinical symptomatology than via the use of morphological information as clinical symptomatology is currently relied upon for the diagnosis and staging of PD (Mahlknecht et al., 2015). Imaging, let alone morphological analysis, is not always used in the diagnosis of PD. This is particularly relevant for defining Early and Late stage PD in which the motor symptoms are a defining characteristic of the disorder (Mahlknecht et al., 2015). Thus, one would expect nearly perfect accuracy between these two stages and Prodromal or HC based solely on motor symptomatology, although this gives little information about the underlying etiology of Parkinson's disease, only the clinical workflow used to diagnose it.

Table 4 presents different performance metrics for all the six problems using all structures as input and Ensemble Learning as the highest performing classifier. Further statistical analysis (i.e. Tukey's HSD test) could not be performed as the cohort size plays a preponderant role in the analysis of these results. It shows that logically, the DBS cohort was the easiest to classify, as it is composed of patients with the most advanced state of the disease. The three other cohorts were more difficult to distinguish between themselves.

Table 3
Results of different binary problems using MDS-UPDRS3 score as input and a naive Bayes classifier.

Problem	BACC	Sens.	Spec.	F1
Early vs HC	95%	99%	82%	97%
Early vs Pro.	82%	98%	27%	86%
Pro. vs HC	64%	69%	77%	46%
DBS vs Pro.	94%	85%	98%	90%
DBS vs Early	72%	11%	98%	19%
DBS vs HC	99%	85%	100%	92%

Table 4
Results of different binary problems, with Ensemble Learning as a classifier, and all the structures in input. The first class of the problem is considered the positive class.

Problem	BACC	Sens.	Spec.	F1
Early vs HC	59%	77%	38%	67%
Early vs Pro.	64%	93%	23%	88%
Pro. vs HC	69%	42%	88%	49%
DBS vs Pro.	78%	89%	62%	86%
DBS vs Early	80%	64%	91%	68%
DBS vs HC	85%	87%	87%	83%

An interesting result was that the use of morphological information for classifying between Prodromal stage PD and healthy controls appears to outperform the use of motor clinical symptomatology. This is likely as the motor symptomatology does not capture the other symptoms that are more indicative of Prodromal stage PD which may be related to morphological changes detectable from imaging.

It however is unexpected that the HC versus Early PD problem was more difficult than the Prodromal versus Early PD, as the prodromal stage is widely considered anterior than the early stage, in the course of the disease. Yet, prodromal stage is considered as very heterogeneous in terms of symptomatology (Mahlknecht et al., 2015).

3.2.2. Classifier comparison

Table 5 shows the BACC's across the different classification algorithms as well as the results of a Tukey's HSD test. It shows three clusters of classifiers with algorithms in different clusters having statistically significant differences in performance. The first cluster is composed of RF alone indicating that it has significantly worse performance of all the methods investigated. The second cluster is composed of SVM with a linear kernel indicating that although it significantly outperformed RF, it still underperformed SVM with a non-linear kernel and ensemble learning. The final, highest-performing cluster is composed of the SVM with radial basis kernels and Ensemble Learning (EL) indicating that these two methods are roughly equivalent in performance, neither outperforming the other significantly.

The fact that EL does not have significant better results than SVMr can be explained by the fact that it often relied solely on the prediction of SVMr, which is the highest accuracy base classifier. Indeed, the three base classifiers mostly agree on the classification outcomes. This lack of variety in the predictions between base classifiers limits the interest of stacking these classifiers together. Nevertheless, stacking classifiers can be considered as a safer choice in general as it theoretically and experimentally does not under-perform any base classifier. Overall, our results indicate that although different classification algorithms delivered statistically significantly different performance, this effect is marginal.

Table 5
Tukey's HSD test to compare classifier performances. Mean BACC for each classifier is also displayed. Alpha = 0.05.

Classifier	N	Cluster		
		1	2	3
RF	300	68.0%		
SVMl	300		69.3%	
SVMr	300			70.9%
EL	300			71.3%
Within-group Sig.		1	1	0.674

Table 6
Tukey's HSD test to compare structures' performances. Mean BACC for each structure combination is also displayed. Alpha = 0.05.

Structures	N	Cluster		
	1	2	3	
Left caudate	240	68.3%		
Left putamen	240	68.7%		
Rigth caudate	240	69.3%		
Right putamen	240		70.6%	
All structures	240			72.7%
Within-group Sig.		0.224	1.000	1.000

3.2.3. Structure comparison

Table 6 shows Tukey's HSD test results to compare the informativeness of the four different structures studied identifying three clusters. The first and lowest performing cluster included the left and right caudate as well as the left putamen, implying that these structures provide an equivalent amount of information to the classification algorithm. The second cluster contained the right putamen, which was statistically significantly more informative than the other individual structures. The final and highest performing cluster reflects when the classifier had information regarding all structures simultaneously.

It is unsurprising that taking all four structures together to make predictions is statistically significantly better than taking any individual one (Table 6). This demonstrates that there exists some non-overlapping information across structures, i.e. that the effect of PD cannot be isolated solely to a single structure, and that striatal structure deterioration is not uniform between all structures. Subcortical structure deterioration in PD is likely to be specific to each structure, with these specificities being differently informative. It was unexpected, however, that right structures seem to be more informative overall than the left ones which is investigated in the following section.

3.3. Laterality significance

As shown in Section 3.2, right structures appeared to be more informative than the left ones when considered in isolation. This asymmetry could be due to two sources: a consistent right-left difference in the algorithm or underlying disease progression, or a contralateral-ipsilateral difference which is exposed from a left–right bias in the underlying population. A possible technical reason for the first source could be the higher compressibility of right structures, shown in Section 3.1. As fewer principal components are needed to embed more information about the shape displacements for right structures, it is possible that classifiers favour these structures more than the left ones.

The second source is partially confirmed by the slight over representation of right-side PD in our DBS PD cohort (81 left side start against 86 right side start). In this cohort, we could extract information regarding the symptom progression of PD, determining whether or not the symptoms began on the right or left side. Indeed, if the disease starts evolving earlier on one side of the brain, the subcortical structures of this side may be in a more advanced phase of deterioration, and thus are more informative to diagnosis the disease.

To verify this hypothesis, we ran a new experiment to compare the significance of left against right structures for both left sided PD patients and right sided PD patients. We ran a 10-fold CV, using Ensemble Learning as a classifier, as it is the most performant, as shown in Table 5. Two problems have been tested, HC versus Left sided PD and HC versus Right sided PD. For each problem, we reported the BACC of the 10-fold CV, using in one case the left caudate nucleus and the left putamen, and in the other case the right caudate nucleus and right putamen. For each new problem, a paired-sample t-test has been performed to assess if the side ipsilateral to disease progression is giving statistically better accuracy results than the contralateral side. Results

Table 7
BACC for Ensemble Learning between HC and DBS cohorts, the latter having laterality information regarding PD progression.

Struct. side	Left struct.	Right struct.	p-value
HC vs Left PD	81.8% ± 0.5%	78.5 ± 3.4%	0.009
HC vs Right PD	76.5% ± 1.2%	87.9 ± 2.1%	< 0.001

are reported in Table 7. We can see that, for left-sided PD patients, the left structures were consistently more informative than the right ones ($p = 0.009$). For right-sided PD patients, right structures were consistently more informative than left ones ($p < 0.001$).

This experiment confirms the hypothesis that the starting side of PD is an important source of diagnostic information, and should be taken into account when interpreting medical images. This suggests that the left–right difference observed in Table 6 is a result of a difference in population size of left-sided and right-sided PD in which there is a contra- vs. ipsi-lateral difference in terms of the information content of each structure. However, more longitudinal studies would be necessarily to determine if this difference persists in the Early and Prodromal cohorts in which this laterality may not have yet manifested.

3.4. MDS-UPDRS3 prediction

In order to quantify the predictive power of striatal shape morphology for the motor symptomatology of the patients, we used an analogous pipeline as in the previous experiment, with RF as a regressor instead of a classifier, to predict the MDS-UPDRS3 score of the patients of all cohorts, besides HC. We decided to remove the HC cohort from this experiment, as the task of predicting their MDS-UPDRS3 is not relevant, as it is concentrated at zero, thus biasing the model's prediction towards zero. Fig. 4 presents the results of a 10-fold CV, after having ran an HPO. We compared the performance of our system with a baseline predicting always the cohorts' mean MDS-UPDRS3 score. The performance of both systems are shown on Table 8. The correlation coefficient R of the regression is equal to 0.215, which can be considered as weak, yet significant ($p < 0.001$). This may indicate that the investigate subcortical shape alterations are co-caused along with motor symptomatology, rather than being in direct causation.

4. Discussion

When comparing the baseline classification results of Table 3

Fig. 4. 10-fold CV MDS-UPDRS3 prediction with proposed method, for cohorts DBS PD, Early PD and Prodromal. Red curve is the linear regression line of the predictions. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 8
Statistics for 10-fold CV MDS-UPDRS3 prediction with our method and a mean-prediction baseline, for cohorts DBS PD, Early PD and Prodromal.

Method	MSE	MAE	R
Proposed method	224.4 ± 380.4	11.64 ± 9.434	0.215
Mean baseline	234.9 ± 433.1	11.77 ± 9.816	−0.160

exclusively based on motor symptomatology, and the results of Table 4 obtained with our system, we can observe that, in 4 cases out of 6, the results are better for the baseline. We can explain that by the fact that the stage of the disease is derived from these clinical measurements, including the motricity. The most interesting exception is for the prodromal versus healthy control problem, where our systems performs better. It confirms that brain alteration in PD's prodromal stage is prior to and globally more informative than any motor symptoms, and seem to be correlated with the apparition of non-motor symptoms. Secondarily, it indicates that information from subcortical structures shape alteration is somewhat orthogonal to the clinical motor symptomatology of the patients, and that both should be taken into account when monitoring the evolution of the disease. This conclusion, coherent with the literature (Garg et al., 2015; Nemmi et al., 2015), is supported by the results presented in Fig. 4, which shows that motor symptomatology of the patients is largely orthogonal to striatal shape alterations.

Using only subcortical shape analysis, our system shows, with each structure, consistent results for classifying PD patients from healthy controls, whether they are on early or advanced phases, contrary to most of the literature presented in the Introduction, where the results are more subtle, and often contradictory (Nemmi et al., 2015; Owens-Walton et al., 2018). We are also the first to successfully differentiate healthy controls and Prodromal subjects using subcortical shape analysis. The main difference between our study and the current state of the art is that we are using larger cohorts and more advanced machine learning algorithms, which consider non-linear relations along all the surface mesh, whereas the literature usually considers linear relations, one point at a time. This observation shows that subcortical shape alteration is complex and non-linear, and that a supplementary methodological effort has to be done when working with subcortical imaging data in this context.

An unexpected result from our study is that Prodromal patients seem to be easier to distinguish from healthy controls than early PD patients. Although we don't have any clear explanation to provide, this observation illustrates that subcortical shape alteration is not uniform in time.

In term of compression, all four structures displayed the same expected behaviour (Fig. 3), although right structures appeared to be more easily compressed. This observation is not linked to the volume of the segmentations on the atlas. We cannot be sure if right structures have a more coherent shape deformation pattern, or if this observation is caused by a unexpectedly noisier output from the segmentation, registration and vertex-wise boundary displacements on the left structures.

We also tried a non-linear data compression method using a neural network and more specifically deep stacked residual autoencoder. In addition, we used feature selection with the k-best algorithm using different correlation scores. However, neither of these approaches yielded consistent better performance while greatly increasing its complexity, in the case of neural networks.

It should be noted that the different stages of PD exist in a continuum, that is, there is no clearly defined, objective delineation between stages. In addition, the Prodromal cohort do not all uniformly progress to full symptomatic PD, although it is increasingly been seen as a common pre-cursor (Mahlknecht et al., 2015). This has a distinct effect on the accuracy of the methods using either motor symptomology or striatal morphology, specifically for problems that can be seen as 'adjacent' such as classifying between early-stage and late-stage PD, or between healthy controls and the Prodromal cohort.

Our system was able to predict MDS-UPDRS3 score of various PD subjects. The performance is weak, yet significant ($p < 0.001$) as shown in Table 8. To us, it indicates that our methodology can be used in theory to predict the motor symptomology of a patient based on striatal morphology, but that they are either not strongly correlated or causally connected. The fact that cohort classification is an easier task than motor symptomatology prediction also supports our conclusion that PD striatal shape alteration is orthogonal to motor symptomatology.

4.1. Future work

By observing that the shape displacements of some structures are more relevant in some problems than others, we hypothesized that striatal structures would deform non–homogeneously through time. We attempted to find a pattern in the spreading of relevant clusters of points through the evolution of the disease, but couldn't find anything reliable. Further study would be needed in this direction. Additionally, we could not find a clear interpretation regarding the most relevant structures of every problem.

In this extent, the longitudinal observation of a cohort of patients would be a good follow up for this study. Indeed, tracking patient's disease on several years could allow to conclude more reliably on patterns in striatal shape alteration by removing the inter-population bias.

A second important direction would be to correlate subcortical shape alterations and the severity of non-motor symptoms, such as cognitive and neuropsychiatric decline. So far, most of the literature, including this manuscript, focuses essentially on motor symptomatology. Finding biomarkers hinting at the severity and the speed of cognitive and neuropsychatric decline would allow for greatly improved medical care, as the choice of treatment would benefit from this information. The ability of our system to classify healthy controls and prodromal subjects, two cohorts for which the motor symptomatology difference is marginal, highlights the relevance of subcortical shape alterations for non-motor symptomatology. Present work could be extended to study other subcortical structures, from other imaging sequences than t1-weighted MRI especially for a better understanding of Prodromal Parkinson's disease with its more heterogeneous symptoms.

Finally, methodologically speaking, our study suffers from some biases. First of all, the sizes of the cohorts are very uneven, which keeps us from directly comparing the relevance of shape displacement for different diseases stages. Secondarily, PPMI being a multi-centre program, the imaging acquisition methodology and devices are not consistent. Although the registration and segmentation processes are designed to be somewhat agnostic to these differences, there is a possibility that they could have impacted the output of the shape analysis pipeline, and thus bias the results. Nonetheless, this study being mostly preliminary and comparative, the observations drawn in this paper are still relevant to the scientific community and give some insight into the progression pattern of PD.

5. Conclusions

In this paper we presented a pipeline able to successfully differentiate populations at different stages of PD, solely from the morphological shape alterations of the bilateral caudate nucleus and putamen. This is an important conclusion for the PD research, as MRI is only marginally used as a bio-marker for the disease, compared to clinical biomarkers. Imaging biomarkers also give more direct insight into disease progression, giving targeted information about the patient's neuroanatomy, as opposed to the more global behavioural view attained through measuring manifested motor symptoms. This is crucial for the understanding of prodromal and early stages of PD in which

symptoms are not confined solely to motor performance.

Our analysis framework was constructed as a series of binary classifiers distinguishing between different stages of PD. Subjects with advanced PD are reliably distinguished from the other cohorts, with a high sensitivity and specificity. Distinguishing between more similar stages of PD has lower performance, but still outperforms diagnosis based solely on symptomology indicating that shape alterations occur early in the disease and progress over time. We show that striatal structures shape displacements can be reliably used as a diagnosis biomarker for PD, even with relatively simple and well-validated machine learning tools.

The performance is more limited in distinguishing between the healthy control, prodromal PD and early PD groups. Yet, the results are significant, and we show that subtle shape displacements exists, are detected and are exploitable with this methodology. Subcortical shape displacement can thus be used as a staging biomarker. We notably showed, for the first time, that striatal structures shape displacements allows to diagnosticate PD's prodromal stage.

We showed that each of these structures are informative in a different way, as the best performance is given when using them all as input. Thus, the shape displacements of each structure carries different relevant information. Then, we investigated on the informativeness of left and right structures, and concluded that the starting side of the disease is directly correlated to it.

Finally, we shown that our system is hardly capable to predict motor symptomatology from striatal shape displacement, showing that both are either poorly, or difficultly correlated.

CRediT authorship contribution statement

Maxime Peralta: Conceptualization, Methodology, Software, Validation, Investigation, Data curation, Writing - original draft. **John S.H. Baxter:** Conceptualization, Methodology, Software, Validation, Investigation, Writing - review & editing. **Ali R. Khan:** Methodology, Software, Writing - review & editing. **Claire Haegelen:** Validation, Resources, Writing - review & editing. **Pierre Jannin:** Conceptualization, Validation, Writing - review & editing, Supervision.

Acknowledgments

Maxime Peralta's PhD is funded by the Fondation pour la Recherche Médicale (FRM). John Baxter is supported by a Post-Doctoral Fellowship from the Natural Sciences and Research Council of Canada (NSERC) and by the Institut des Neurosciences Cliniques de Rennes (INCR).

Part of the data used in the preparation of this article were obtained from the Parkinson's Progression Markers Initiative (PPMI) database (www.ppmi-info.org/data). For up-to-date information on the study, visit www.ppmi-info.org. PPMI - a public–private partnership - is funded by the Michael J. Fox Foundation for Parkinson's Research and funding partners, including AbbVie, Allergan, Avid Radiopharmaceuticals, Biogen, BioLegend, Bristol-Myers Squibb, Celgene, Denali Therapeutics, GE Healthcare, Genentech, GlaxoSmithKline, Lilly, Lundbeck, Merck, Meso Scale Discovery, Pfizer, Piramal, Prevail, Roche, Sanofi-Genzyme, Servier, Takeda, Teva, UCB, Verily, Voyager Therapeutics and Golub Capital.

References

Ansari, Shahab U., 2010. Validation of fs+ lddmm by automatic segmentation of caudate nucleus in brain mri. In: Proceedings of the 8th International Conference on Frontiers of Information Technology ACM, pp. 10.

Bergouignan, Loretxu, Chupin, Marie, Czechowska, Yvonne, Kinkingnéhun, Serge, Lemogne, Cédric, Le Bastard, Guillaume, Lepage, Martin, Garnero, Line, Colliot, Olivier, Fossati, Philippe, 2009. Can voxel based morphometry, manual segmentation and automated segmentation equally detect hippocampal volume differences in acute depression? Neuroimage 45 (1), 29–37.

Berner, Laura A, Wang, Zhishun, Stefan, Mihaela, Lee, Seonjoo, Huo, Zhiyong, Cyr, Marilyn, Marsh, Rachel, 2019. Subcortical shape abnormalities in bulimia nervosa. Biolog. Psychiatry: Cognitive Neurosci. Neuroimaging.

Connolly, Barbara S., Lang, Anthony E., 2014. Pharmacological treatment of parkinson disease: a review. Jama 311 (16), 1670–1683.

Davatzikos, Christos, 2004. Why voxel-based morphometric analysis should be used with great caution when characterizing group differences. Neuroimage 23 (1), 17–20.

Delenclos, Marion, Jones, Daryl R., McLean, Pamela J., Uitti, Ryan J., 2015. Biomarkers in parkinson's disease: advances and strategies. Parkinsonism Related Disorders.

Faisal Beg, M., Miller, Michael I, Trouvé, Alain, Younes, Laurent, 2005. Computing large deformation metric mappings via geodesic flows of diffeomorphisms. Int. J. Ccomputer Vision 61 (2), 139–157.

Foo, H., Mak, E., Yong, T.T., Wen, M.C., Chander, R.J., Au, W.L., Sitoh, Y.Y., Tan, L.C.S., Kandiah, N., 2017. Progression of subcortical atrophy in mild parkinson's disease and its impact on cognition. Eur. J. Neurol. 24 (2), 341–348.

Garg, Amanmeet, Wong, Darren, Popuri, Karteek, Poskitt, Kenneth J, Fitzpatrick, Kevin, Bjornson, Bruce, Grunau, Ruth E, Beg, Mirza Faisal, 2014. Manually segmented template library for 8-year-old pediatric brain mri data with 16 subcortical structures. J. Med. Imaging 1 (3), 034502.

Garg, Amanmeet, Appel-Cresswell, Silke, Popuri, Karteek, McKeown, Martin J., Faisal Beg, Mirza, 2015. Morphological alterations in the caudate, putamen, pallidum, and thalamus in parkinson's disease. Frontiers in Neuroscience 9, 101.

Gerig, Guido, Styner, Martin, Shenton, Martha E., Lieberman, Jeffrey A., 2001. Shape versus size: Improved understanding of the morphology of brain structures. In: International Conference on Medical Image Computing and Computer-Assisted Intervention Springer, pp. 24–32.

Griffiths, P.D., Perry, R.H., Crossman, A.R., 1994. A detailed anatomical analysis of neurotransmitter receptors in the putamen and caudate in parkinson's disease and alzheimer's disease. Neurosci. Lett. 169 (1–2), 68–72.

Hentz, Joseph G., Mehta, Shyamal H., Shill, Holly A., Driver-Dunckley, Erika, Beach, Thomas G., Adler, Charles H., 2015. Simplified conversion method for unified parkinson's disease rating scale motor examinations. Mov. Disord. 30 (14), 1967–1970.

Khan, Ali R., Wang, Lei, Beg, Mirza Faisal, 2008. Freesurfer-initiated fully-automated subcortical brain segmentation in mri using large deformation diffeomorphic metric mapping. Neuroimage 41 (3), 735–746.

Khan, Ali R., Hiebert, Nole M., Vo, Andrew, Wang, Brian T., Owen, Adrian M., Seergobin, Ken N., MacDonald, Penny A., 2019. Biomarkers of parkinson's disease: Striatal sub-regional structural morphometry and diffusion mri. NeuroImage: Clinical 21, 101597.

Kish, Stephen J, Shannak, Kathleen, Hornykiewicz, Oleh, 1988. Uneven pattern of dopamine loss in the striatum of patients with idiopathic parkinson's disease. N. Engl. J. Med. 318 (14), 876–880.

Limousin, Patricia, Foltynie, Tom, 2019. Long-term outcomes of deep brain stimulation in parkinson disease. Nature Rev. Neurol. 15 (4), 234–242.

Limousin, Patricia, Martinez-Torres, Irene, 2008. Deep brain stimulation for parkinson's disease. Neurotherapeutics 5 (2), 309–319.

Mahlknecht, Philipp, Seppi, Klaus, Poewe, Werner, 2015. The concept of prodromal parkinson's disease. J. Parkinson's Disease 5 (4), 681–697.

Mak, E., Bergsland, N., Dwyer, M.G., Zivadinov, R., Kandiah, N., 2014. Subcortical atrophy is associated with cognitive impairment in mild parkinson disease: a combined investigation of volumetric changes, cortical thickness, and vertex-based shape analysis. Am. J. Neuroradiol. 35 (12), 2257–2264.

Menke, Ricarda A.L., Szewczyk-Krolikowski, Konrad, Jbabdi, Sa.ad., Jenkinson, Mark, Talbot, Kevin, Mackay, Clare E., Michele, Hu, 2014. Comprehensive morphometry of subcortical grey matter structures in early-stage parkinson's disease. Human Brain Mapping 35 (4), 1681–1690.

Nemmi, Federico, Sabatini, Umberto, Rascol, Olivier, Péran, Patrice, 2015. Parkinson's disease and local atrophy in subcortical nuclei: insight from shape analysis. Neurobiol. Aging 36 (1), 424–433.

Owens-Walton, Conor, Jakabek, David, Li, Xiaozhen, Wilkes, Fiona A., Walterfang, Mark, Velakoulis, Dennis, Van Westen, Danielle, Looi, Jeffrey C.L., Hansson, Oskar, 2018. Striatal changes in parkinson disease: an investigation of morphology, functional connectivity and their relationship to clinical symptoms. Psychiatry Res.: Neuroimaging 275, 5–13.

Péran, Patrice, Nemmi, Federico, Barbagallo, Gaetano, 2018. Brain morphometry: Parkinson's disease. In: Brain Morphometry Springer, pp. 267–277.

Pfeiffer, Ronald F., 2016. Non-motor symptoms in parkinson's disease. Parkinsonism Related Disorders 22, S119–S122.

Xiaoying Tang, Dominic Holland, Anders M. Dale, Laurent Younes, Michael I. Miller, Alzheimer's Disease Neuroimaging Initiative. Shape abnormalities of subcortical and ventricular structures in mild cognitive impairment and alzheimer's disease: detecting, quantifying, and predicting. Human Brain Mapping, 35(8):3701–3725, 2014.

Thenganatt, Mary Ann, Jankovic, Joseph, 2014. Parkinson disease subtypes. JAMA Neurol. 71 (4), 499–504.

Tuite, Paul, 2016. Magnetic resonance imaging as a potential biomarker for parkinson's disease. Transl. Res. 175, 4–16.

van den Bogaard, Simon J.A., Dumas, Eve M., Ferrarini, Luca, Milles, Julien, van Buchem, Mark A., van der Grond, Jeroen, Roos, Raymund A.C., 2011. Shape analysis of subcortical nuclei in huntington's disease, global versus local atrophy–results from the track-hd study. J. Neurolog. Sci. 307 (1-2), 60–68.

Wade, Benjamin S.C., Joshi, Shantanu H., Pirnia, Tara, Leaver, Amber M., Woods, Roger P., Thompson, Paul M., 2015. Randall Espinoza,and Katherine L. Narr. Random forest classification of depression status based on subcortical brain morphometry following electroconvulsive therapy. In: 2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI) IEEE, pp. 92–96.

Wang, Lei, Khan, Ali, Csernansky, John G., Fischl, Bruce, Miller, Michael I., Morris, John C., Faisal Beg, M., 2009. Fully-automated, multi-stage hippocampus mapping in very mild alzheimer disease. Hippocampus 19 (6), 541–548.

Xiao, Yiming, Vladimir Fonov, M., Chakravarty, Mallar, Beriault, Silvain, Subaie, Fahd Al, Abbas Sadikot, G., Pike, Bruce, Bertrand, Gilles, Louis Collins, D., 2017. A dataset of multicontrast population-averaged brain mri atlases of a parkinson's disease cohort. Data in brief 12, 370–379.