Clin Neuroradiol (2019) 29:605–614
https://doi.org/10.1007/s00062-018-0727-8

ORIGINAL ARTICLE

Improved Automatic Morphology-Based Classification of Parkinson's Disease and Progressive Supranuclear Palsy

Aron S. Talai[1] · Zahinoor Ismail[2] · Jan Sedlacik[3] · Kai Boelmans[4] · Nils D. Forkert[1]

Received: 4 April 2018 / Accepted: 25 August 2018 / Published online: 14 September 2018

Abstract

Objectives The overlapping symptoms of Parkinson's disease (PD) and progressive supranuclear palsy—Richardson's syndrome (PSP-RS) often make a correct clinical diagnosis difficult. The volume of subcortical brain structures derived from high-resolution T1-weighted magnetic resonance imaging (MRI) datasets is frequently used for individual level classification of PD and PSP-RS patients. The aim of this study was to evaluate the benefit of including additional morphological features beyond the simple regional volume, as well as clinical features, and morphological features of cortical structures for an automatic classification of PD and PSP-RS patients.

Material and Methods A total of 98 high-resolution T1-weighted MRI datasets from 76 PD patients, and 22 PSP-RS patients were available for this study. Using an atlas-based approach, the volume, surface area, and surface-area-to-volume ratio (SA:V) of 21 subcortical and 48 cortical brain regions were calculated and used as features for a support vector machine classification after application of a RELIEF feature selection method.

Results The comparison of the classification results suggests that including all three morphological parameters (volume, surface area and SA:V) can considerably improve classification accuracy compared to using volume or surface area alone. Likewise, including clinical patient features in addition to morphological parameters also considerably increases the classification accuracy. In contrast to this, integrating morphological features of other cortical structures did not lead to improved classification accuracy. Using this optimal set-up, an accuracy of 98% was achieved with only one falsely classified PD and one falsely classified PSP-RS patient.

Conclusion The results of this study suggest that clinical features as well as more advanced morphological features should be used for future computer-aided diagnosis systems to differentiate PD and PSP-RS patients based on morphological parameters.

Keywords Magnetic resonance imaging · T1 image sequences · Computer-assisted image Analysis · Parkinson's disease · Progressive supranuclear palsy

Introduction

The identification of patients with a Parkinsonian syndrome and related disorders, and the accurate diagnosis of the specific entity causing the syndrome is an important clinical issue. Due to differences in disease progression, rate of cognitive decline, cognitive and neuropsychiatric symptom profile, quality of life, and responses to specific treatment strategies, such as levodopa or deep brain stimulation, diagnosis support systems can assist clinicians in choosing appropriate interventions earlier in the disease course [1–3].

This is especially a critical issue for patients with progressive supranuclear palsy—Richardson's syndrome (PSP-RS) due to the lower life expectancy, and the poor response to standard Parkinson's disease (PD) therapies. Diagnostic guidelines for PD [4] and PSP [5, 6] are mainly based

✉ Nils D. Forkert
nils.forkert@ucalgary.ca

1 Department of Radiology and Hotchkiss Brain Institute, Faculty of Medicine, University of Calgary, 3330 Hospital Drive NW, AB T2N 4N1 Calgary, Canada

2 Departments of Psychiatry, Clinical Neurosciences, and Community Health Sciences, and Hotchkiss Brain Institute, University of Calgary, Calgary, Canada

3 Department of Diagnostic and Interventional Neuroradiology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

4 Department of Neurology, University Hospital Würzburg, Würzburg, Germany

on clinical features, resulting in high misdiagnosis rates due to overlapping symptoms, particularly in early disease stages. More precisely, failure rates up to 25% have been reported in neuropathologically proven studies [7]. Thus, clinical symptom-based nosologies may be inadequate on their own for early detection and diagnosis.

To overcome this challenge and improve the diagnostic precision of PD and PSP differentiation, a wide variety of computer-aided diagnostic approaches based on electro-oculography [8], gait analyses [9], or imaging-based [10–14] methods have been suggested with varying degrees of success. Within this context, the usage of morphological parameters extracted from high-resolution T1-weighted magnetic resonance imaging (MRI) datasets seem especially promising due to the modality's comparably high discriminative power and widespread availability. In recent years, numerous studies utilizing information derived from structural T1-weighted images have been conducted to identify differences between Parkinsonian syndromes. In terms of group-wise studies, most previous studies have reported overall increased atrophy signs in PSP compared to PD. Messina et al. [15], for example, showed significant volume reductions in the cerebellum, putamen, thalamus, pallidum, brainstem, and hippocampus in PSP vs. PD. Cerebellar gray matter and mesencephalon white matter loss have been reported by Focke et al. [16] in PSP compared to PD. Alternatively, previously proposed methods for an individual level morphology-based classification of PD and PSP patients employing machine learning techniques focused mostly on the volume of specific regions of interests (ROI), such as the brainstem and other regions, which are known to be significant discriminators between PD and PSP [14, 17, 18].

Utilizing the volumetric values of these known ROIs as input features, Scherfler et al. [19], Sarica et al. [20], and Focke et al. [16] were able to obtain comparably high classification accuracies of more than 90%. However, recent studies suggest that various other brain regions, apart from the deep gray matter are also affected by atrophy and could hold additional informative value as input features in classification routines [15, 18, 21, 22].

In addition to plain volume of brain structures, previous research also suggests that the corresponding surface area of the brain structures might also have some additional predictive value for the classification of neurological diseases [23–26]. Worker et al. [27] for example, found an increased surface area in the pericalcarine cortex along with general patterns of cortical thinning as well as volume loss in the superior frontal gyrus in PSP compared to PD. Nevertheless, the surface area and combination of volume and surface area in terms of the surface area-to-volume ratio (SA:V) have rarely been used for classification despite their potential to enhance the classification accuracy.

Volumetric differences are typically considered to be related to the atrophic processes, which naturally take place as part of the aging process but is often accelerated by neurological diseases [22]. While age is a confounding factor for atrophy assessment, it is only infrequently used as an additional feature for classification purposes. Similarly, very few image-based studies have included additional patient information such as gender or clinical test scores as input features within the classification process [28, 29]. Consequently, the investigation of the morphological profile such as volume, surface area, and SA:V of brain regions within and beyond deep gray matter, as well as the incorporation of cognitive and neuropsychiatric symptom test scores and patient demographics may further increase the diagnostic accuracy of an automatic morphology-based classification of PD and PSP-RS patients. Indeed, the combination of multi-modal imaging features has been shown to improve classification accuracy compared to using single source information for classification [30–35].

The aim of the present study was to utilize a wide range of morphological features in combination with clinical test scores to develop and evaluate an automatic method to differentiate PD and PSP-RS patients using a support vector machine classifier. Moreover, the significance of adding cortical features to subcortical features in this morphology-based classification was further investigated.

Material and Methods

Patients and MR Protocol

A total of 98 MR datasets were available for the present study, including 76 PD and 22 PSP-RS patients, which were previously described by Boelmans et al. [36]. Information about demographic and clinical features is given in Table 1.

Institutional ethics committee (Ärztekammer Hamburg) approval (ethics number PV3436) was obtained and all subjects provided written informed consent consistent with the Declaration of Helsinki prior to study participation. Patients included in this single center study were recruited from the movement disorder outpatient clinic of the Neurology Department of the University Medical Center Hamburg-Eppendorf between July 2009 and September 2010. All patients were seen at least twice by two movement disorder specialist and only patients with a probable diagnosis were included in this study. No FDG-PET results were available for the study participants.

The clinical diagnosis of PD [4] and PSP [5, 6] were made according to established consensus criteria at the time of patient recruitment, with PSP patients presenting as classical RS. In detail, probable PSP-RS patients with axial rigidity, early falls, balance instability, and vertical

Table 1 Demographic and clinical characteristics of study participants

	Parkinson's disease	Progressive supranuclear palsy
Number of patients	76	22
Gender, M/F	53/23	10/12
Age at examination, years, mean ± SD (range)	63.3 ± 9.0	70.9 ± 5.5
	(40–77)	(59–79)
Disease duration, years, mean ± SD (range)	12.2 ± 6.2	5.9 ± 3.3
	(0.5–30.2)	(1.2–12.6)
Hoehn and Yahr, mean ± SD (range)	2.7 ± 0.8	2.5 ± 0.8
	(1–4)	(1–4)
UPDRS motor score (OFF condition), mean ± SD (range)	36.8 ± 11.8	32.8 ± 11.4
	(14–63)	(9–52)
UPDRS motor score (ON condition), mean ± SD (range)	20.3 ± 8.8	28.7 ± 10.3
	(5–52)	(6–48)
MMSE, mean ± SD (range)	28.2 ± 1.3	24.8 ± 3.1
	(23–30)	(17–29)

MMSE mini-mental state examination; *UPDRS* unified Parkinson's disease rating scale; *SD* standard deviation

palsy were included in the study, whereas PSP-RS patients who demonstrated noticeable freezing, significant clinical response to levodopa, as well as asymmetric clinical features were excluded. Patients with a possible level of evidence for diagnosis, or presenting with other neurological, psychiatric, or systemic disorders were excluded from this study.

All MRI datasets were acquired using a 3 T Skyra scanner (Siemens, Erlangen, Germany). A high-resolution T1-weighted MPRAGE [37] dataset was acquired for each patient with TR = 1900 ms, TE = 2.46 ms, flip angle = 9°, TI = 900 ms, image in-plane resolution of 0.94 mm^2, and 0.94 mm slice thickness.

Image Processing

The automatic brain parcellation, which is required for the subsequent regional brain morphometry analysis, was performed by registration of the 152 MNI brain atlas [38] to each patient dataset (Fig. 1). For this purpose, an affine registration of the atlas to the patient dataset was performed first using the block-matching approach described by Ourselin et al. [39]. The resulting affine transformation was then used as an initialization for the subsequent non-linear registration using a free-form deformation as implemented in the NiftyReg software package [40].

The calculated non-linear transformation for each patient was used to warp the Harvard-Oxford sub-cortical and cortical atlas brain regions to each patient employing a nearest-neighbor interpolation. More precisely, the Harvard-Oxford subcortical atlas defined in the MNI reference space consists of 21 brain regions such as the thalamus, caudate, hippocampus, and brainstem and the Harvard-Oxford cortical atlas consists of 48 brain regions such as the insular cortex, precentral gyrus, and temporal pole.

In addition to the segmented atlas brain regions, the non-linear deformation field was also used for transforming a binary segmentation of the total intra-cranial volume defined in the atlas space to each patient dataset, which was used for volumetric normalization of the single brain regions to account for differences regarding the general head anatomy. Apart from the volume, the segmented brain regions were also used to determine the surface area of each brain region as well as the SA:V. For calculation of the SA:V, the raw regional volumes instead of the volumes corrected for the full intra-cranial volume were used as the ratio calculation represents a normalization for the individual head geometry by itself.

Classification Pipeline

The morphological analysis results in a total of 207 imaging features consisting of 69 features for each of the volume, brain surface area, and SA:V categories. In detail, 21 sub-cortical and 48 cortical structures were determined per patient in each category as described above. Furthermore, seven clinical features were recorded and available for classification in this study, which included age, gender, disease duration, Hoehn and Yahr score, unified PD rating scale (UPDRS; ON and OFF condition), and the Mini-Mental State Examination scale (MMSE). Thus, a total of 214 image and non-image features are available for each patient.

The classification pipeline consists of a feature selection method followed by a classification algorithm. It is clear that many of the available features are likely to be non-infor-

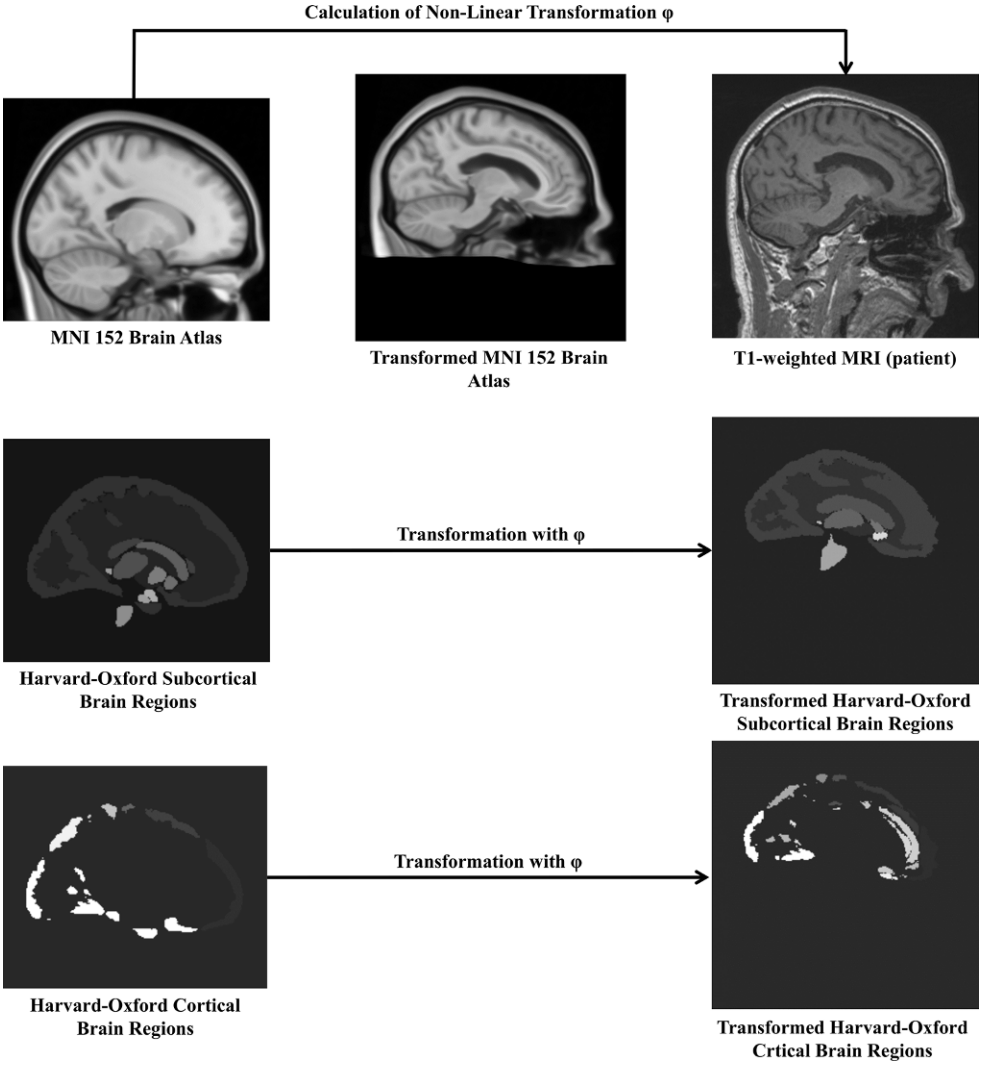

Fig. 1 Schematic representation of the regional brain volume quantification procedure using non-linear registration of the MNI atlas to the T1-weighted datasets

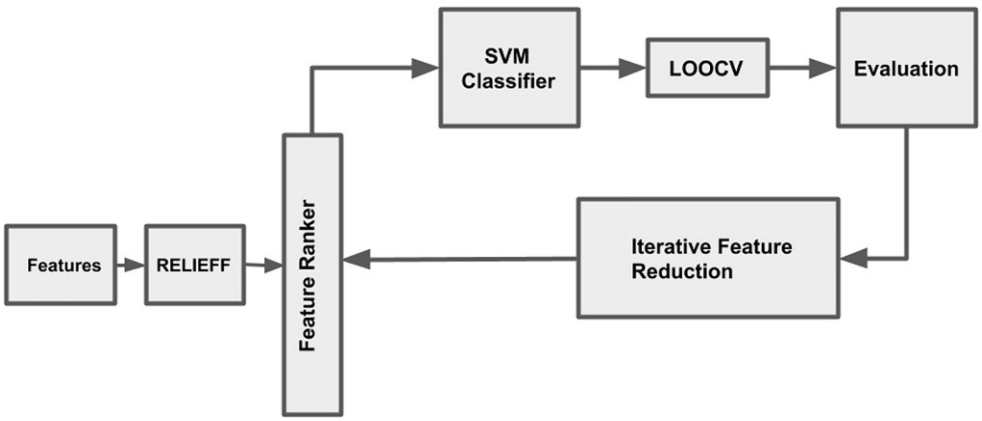

Fig. 2 Schematic representation of the classification pipeline used for this study (*LOOCV* leave-one-out cross validation, *SVM* support vector machine, *LOOCV* leave one out cross validation

mative, which may downgrade the classification accuracy if included in the feature pool. Therefore, the RELIEFF feature selection method [41], with 10 nearest neighbors was employed to reduce the number of redundant features. The RELIEFF method was chosen due to its ability to recognize non-linear relationships between features and the class labels.

After feature ranking, a linear kernel support vector machine (SVM) as implemented in the libSVM toolbox [42] with a cost function value of C = 1, which controls the trade-

off between misclassification and error minimization, was first employed using the entire feature set. The linear kernel SVM and default parameter C = 1 were selected to reduce the risk of data over-fitting. The optimal number of high ranking features used for training and testing of the classifier was then systematically optimized by iteratively removing the lowest ranked feature from training and testing. After training, the SVM can be used to classify a new feature sample with unknown class label not part of the training set. A nested leave-one-out cross-validation (LOOCV; leave one out cross validation) strategy was employed to evaluate the performance of the SVM classifiers by calculating the overall accuracy and group-specific evaluation metrics. To prevent a double-dipping bias, the feature selection was also included in the training of each leave-one-out iteration in a nested loop so that the composition and rank of features included in the optimal subset may be different for each training and testing iteration (Fig. 2).

The same process described above was conducted six times in two phases with different input feature set-ups. In the first phase, only subcortical features were used for the classification process. In the first case of phase 1, only the volumetric features ($n = 21$) were used for training and testing of the classifier. In the second case, the 7 clinical features were added to the volume-based subcortical features leading to a total of 28 features. In the third case, the brain surface area features instead of the volumetric features were employed for classification, whereas in the fourth case clinical features were added to the surface area features, again leading to 28 features. Finally, all subcortical morphological image-based features, consisting of volume, surface area, and SA:V ($n = 63$), were used in the classification pipeline and ultimately, the entire feature set including all image-based subcortical, morphological, and clinical features ($n = 70$) was employed for classification.

In the second phase, the same 6-stage classification set-ups described above were repeated but using subcortical as well as cortical features in each category. The primary rational behind this approach was to investigate the potential benefits of adding less investigated cortical features to traditionally used subcortical features.

Results

The RELIEFF and linear kernel SVM pipeline, followed by LOOCV, yielded a total of 12 classification paradigms, which were evaluated and compared. Table 2 contains details regarding the classification performance for each case using the subcortical features only while Table 3 shows the corresponding results for the classifiers using the cortical morphological features in addition to the subcortical morphological and clinical features.

Overall, the results show that subcortical volume and surface area features alone lead to rather similar classification results, with the surface area features slightly outperforming the volumetric features (84.7% vs. 85.7% accuracy); however, the volumetric features led to a better classification accuracy for PD patients (three misclassifications) compared to the subcortical surface area features (four misclassifications), which resulted in an improved classification of PSP-RS patients (10 vs. 12 misclassifications). This finding suggests that these two morphological parameters contain complementary information valuable for classification of the two patient groups, which is further supported by the higher classification accuracy of 88.8% obtained by the classifier including all three types of subcortical morphological features (only 5 PD patients falsely classified as PSP-RS and 6 PSP-RS patients falsely classified as PD).

Including the clinical features in addition to the subcortical morphometric features led to considerably higher classification accuracy in all three cases, with the classifier using all three morphological parameters and the clinical features achieving the best classification accuracy of 98% compared to 94.9% for the volume and clinical feature classifier and 95.6% for the surface area and clinical feature classifier. Using all subcortical morphological and clinical features led to only one falsely classified PD and one falsely classified PSP-RS patient.

The inclusion of the other cortical morphological parameters in addition to the subcortical structures only improved the classification accuracy slightly for three of the classifiers (1: volume only, 2: volume and clinical features, and 3: volume, surface area, and SA:V); however, including the morphological analysis of the cortical structures did not improve the accuracy of the classifier using the three subcortical morphological parameters and clinical features for training and testing of the classifier.

Discussion

This is one of the first studies to investigate the potential benefits of less frequently used morphological parameters such as the brain surface area and surface area to volume ratio as well as clinical test scores and patient demographics in an automatic classification routine. The main findings of this study with implications for future developments of morphological classification methods for PSP-RS vs. PD differentiation are three-fold. First, inclusion of clinical features and clinical test scores considerably improves the classification accuracy in all investigated cases. Second, the inclusion of the surface area and SA:V provides complimentary information and leads to slightly better classification results compared to using volumetric features only. Third, adding the morphological profiles of cortical structures for classi-

Table 2 Extended classification performance by class (76 PD, 22 PSP-RS) in each case by only using subcortical based features. Confusion matrix is denoted in *italics*

Volumetric features										
Class	TP rate	FP rate	Precision	Recall	F-measure	MCC	ROC Area	*PD*	*PSP-RS*	Accuracy
PD	0.961	0.545	0.859	0.961	0.907	0.511	0.708	*73*	*3*	84.69%
PSP-RS	0.455	0.039	0.769	0.455	0.571			*12*	*10*	
Volumetric and clinical features										
Class	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	*PD*	*PSP-RS*	Accuracy
PD	0.974	0.136	0.961	0.974	0.967	0.851	0.919	*74*	*2*	94.89%
PSP-RS	0.864	0.026	0.905	0.864	0.884			*3*	*19*	
Surface area features										
Class	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	*PD*	*PSP-RS*	Accuracy
PD	0.947	0.455	0.878	0.947	0.911	0.556	0.746	*72*	*4*	85.71%
PSP-RS	0.545	0.053	0.750	0.545	0.632			*10*	*12*	
Surface area and clinical features										
Class	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	*PD*	*PSP-RS*	Accuracy
PD	0.961	0.045	0.986	0.961	0.973	0.888	0.958	*73*	*3*	95.91%
PSP-RS	0.955	0.039	0.875	0.955	0.913			*1*	*21*	
Surface area, volumetric, and surface area to volume ratio features										
Class	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	*PD*	*PSP-RS*	Accuracy
PD	0.934	0.273	0.922	0.934	0.928	0.673	0.831	*71*	*5*	88.77%
PSP-RS	0.727	0.066	0.762	0.727	0.744			*6*	*16*	
Surface area, volumetric, surface area to volume ratio, and clinical features										
Class	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	*PD*	*PSP-RS*	Accuracy
PD	0.987	0.045	0.987	0.987	0.987	0.941	0.971	*75*	*1*	97.95%
PSP-RS	0.955	0.013	0.955	0.955	0.955			*1*	*21*	

TP true positive, *FP* false positive, *MCC* Matthews correlation coefficient, *ROC* receiver operating characteristic, *PD* Parkinson's disease, *PSP-RS* progressive supranuclear palsy—Richardson's syndrome

fication does not improve the classification accuracy compared to using all three subcortical morphological features and clinical scores and, thus, can be neglected.

Individual level classification of PD vs. PSP-RS resulted in an overall accuracy of 98%, when all subcortical morphological and clinical features were employed. The obtained results of this classification method are in the top range of previously reported classification approaches, although it needs to be highlighted that the results are not directly comparable since different databases were used for the development and evaluation of the classifiers. The 98 patients included in this study were recruited prospectively to set up a representative clinical cohort, which denotes a relatively large number of participants compared to most previous studies.

The finding that the inclusion of clinical features allows higher classification accuracies is well in line with previous research in this domain. Planetta et al. [28], for example, reported that adding clinical measures to diffusion-based image features improved the overall classification metrics differentiating PD patients and patients with various atypical Parkinsonian syndromes. Similarly, Hirschauer et al. [29] showed an overall increase in classification accuracy for differentiating PD patients from those patients with scans without evidence of dopaminergic deficit (SWEDD) when clinical and image-based features were combined. Considering the ubiquity of these tests in clinical practice, as well as their positive contribution to classification performance, the integration of such measures within future computer-aided diagnosis systems for Parkinsonian syndrome classification seems strongly suggested. This finding is especially relevant as most image-based classification methods described in the past did not even include simple clinical variables such as age.

The second major finding of this study is that the combination of brain surface area, volume, and area to volume ratio features outperforms the individual features sets in terms of classification accuracy. Several studies in the past have illustrated the point that the combination of multimodal imaging data (e.g. regional volumetric and diffusion features) compared to unimodal feature usage can improve classification accuracy [30, 31, 43]. This is an intuitive aspect considering that multimodal feature sets elucidate a more comprehensive overview of a given entity compared to single-source features. Within this context, it should be highlighted that the morphometric features used in this study were obtained from a single high resolution T1-weighted dataset and, thus, are strictly speaking not mul-

Table 3 Extended classification performance by class (76 PD, 22 PSP) in each case using subcortical and cortical features. Confusion matrix is denoted in *italics*

Volumetric features										
Class	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	*PD*	*PSP-RS*	Accuracy
PD	0.961	0.364	0.901	0.961	0.930	0.658	0.798	*73*	*3*	88.77%
PSP-RS	0.636	0.039	0.824	0.636	0.718			*8*	*14*	
Volumetric and clinical features										
Class	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	*PD*	*PSP-RS*	Accuracy
PD	0.987	0.091	0.974	0.987	0.980	0.911	0.948	*75*	*1*	96.93%
PSP-RS	0.909	0.013	0.952	0.909	0.930			*2*	*20*	
Surface area features										
Class	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	*PD*	*PSP-RS*	Accuracy
PD	0.947	0.455	0.878	0.947	0.911	0.556	0.746	*72*	*4*	85.71%
PSP-RS	0.545	0.053	0.750	0.545	0.632			*10*	*12*	
Surface area and clinical features										
Class	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	*PD*	*PSP-RS*	Accuracy
PD	0.961	0.045	0.986	0.961	0.973	0.888	0.958	*73*	*3*	95.91%
PSP-RS	0.955	0.039	0.875	0.955	0.913			*1*	*21*	
Surface area, volumetric, and surface area to volume ratio features										
Class	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	*PD*	*PSP-RS*	Accuracy
PD	0.961	0.318	0.913	0.961	0.936	0.692	0.821	*73*	*3*	89.79%
PSP-RS	0.682	0.039	0.833	0.682	0.750			*7*	*15*	
Surface area, volumetric, surface area to volume ratio, and clinical features										
Class	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	*PD*	*PSP-RS*	Accuracy
PD	0.987	0.045	0.987	0.987	0.987	0.941	0.971	*75*	*1*	97.95%
PSP-RS	0.955	0.013	0.955	0.955	0.955			*1*	*21*	

TP True Positive, *FP* False Positive, *MCC* Matthews correlation coefficient, *ROC* Receiver operating characteristic

timodal; however, the classification results suggest that the volume, surface area, and SA:V indeed describe different aspects of the regional brain morphometry that are not fully captured by one of these morphometric parameters alone. This hypothesis generated based on the classification results is further supported by the features selected and used for the optimal classification procedure, which generally included a mixture of volumetric, surface area, SA:V, and clinical features in the top ranked features; however, more research is needed to investigate how these different morphological features might describe different atrophic processes in PD and PSP-RS patients, which is beyond the scope of this work.

Finally, the comparison of the results obtained by the different classifier set-ups suggests that the inclusion of cortical morphological brain features in addition to the subcortical regions does not or only marginally improves the classification accuracy. This might be because cortical regions are less affected than subcortical areas in PD syndromes [44]. It is worth noting that no additional classification was performed using just the corresponding cortical features; however, it is highly likely that the performance would be weaker than what is explicitly reported in this study.

It is worth highlighting that in the proposed classification pipeline, brain surface area features result in slightly better PSP-RS differentiation compared to volumetric features when subcortical or the combination with cortical features are employed. This finding further encourages the combination of multimodal features in classification tasks.

It should be noted that morphological parameters have been used rather frequently to classify PD and PSP subjects compared to other image sequences with varying classification results [16, 19, 45, 46]. This is most likely a result of the widespread availability of this image sequence in retrospectively collected study cohorts as a high-resolution T1-weighted MRI sequence is typically part of any brain MRI protocol and the regional brain volume is a rather intuitive parameter that is related to known atrophic processes in Parkinsonian syndromes, e. g. the brainstem. These aspects highlight the utility of this specific MRI sequence for Parkinsonian syndrome classification; however, it also needs to be highlighted that this is the first work to systemically investigate the potential benefits of several technical aspects of machine learning models using features extracted from high-resolution T1-weighted MRI sequences. It might also be argued that at least some of the findings of this study also hold true for machine learning models

using different image sequences (e. g. the aspect of inclusion of clinical features improving classification accuracy and focus on subcortical structures rather than whole brain analyses). Within this context, it should also be noted that only support vector machines with a standard set-up, which are typically viewed as very powerful, were used in this study. While other machine learning approaches such as random forests or neural nets might be able to improve the classification accuracies to some extent, preliminary tests conducted with other machine learning approaches suggest that the general conclusions from this study still hold true.

A few limitations are present in this study. First, the ground truth classifications were determined by expert movement disorder specialists according to established consensus criteria without neuropathological proof. Thus, there may still be a minor level of uncertainty left regarding the ground truth classification used for training and evaluation of the classifier. Within this context, it should also be mentioned that due to the retrospective analysis of the prospectively collected datasets, the most recent guidelines for PD and PSP-RS diagnosis were not available at that time and the previous guidelines were used. The new Movement Disorder Society Clinical Diagnostic Criteria for PD [47] aim to increase the diagnostic accuracy with the emphasis on exclusion criteria, red flags, and supportive criteria. Likewise, the new clinical criteria for PSP [48] recognize early forms of PSP and operationalize a broader spectrum of PSP clinical phenotypes; however, the main findings should also apply if these criteria were used for patient selection and all patients included in this study were seen by two movement disorder specialists at least two times to ensure highly probable ground truth classifications. Second, the proposed method was only evaluated using a leave-one-out validation. Thus, further validation using a completely different prospectively collected database should be performed in the next step. Finally, while the sample size available for this study is considerably large compared to previous similar studies, an even larger sample population would improve the generalizability of the classification model. Consequently, a large database of standardized MR images obtained from multiple MR manufacturers will be helpful in reaching a complete general model needed for assisted clinical diagnosis. Within this context, it should also be pointed out that the two patient cohorts differed considerably regarding the age distribution, which might bias the classification results to some extent; however, while the PSP patients were considerably older, which should lead to more pronounced atrophy effects, the average disease duration was also considerably shorter compared to the PD patients, which might counterbalance the expected increased atrophy due to age compared to the PD group.

Further improvements of the method could be possible by including multispectral MRI features in the classification method such as susceptibility weighted imaging (SWI), diffusion tensor imaging (DTI), or T2 prime MR datasets, which have also been used in previous classification models. The presented model has the potential of integrating subcortical and cortical features of other less prevalent atypical PSP subtypes such as PSP-parkinsonism, PSP-corticobasal syndrome, PSP-progressive non-fluent aphasia, and others in a multilevel classification paradigm provided sufficient syndrome representative datasets are available. Likewise, it would be possible to also include patients with other neurological diseases that are associated with atrophy and healthy subjects in the presented classification framework; however, in this case the morphological fingerprint of the other cortical structures, which did not seem to be relevant for the classification of PD and PSP-RS patients, might become more informative.

Conclusion

In summary, the results of this study show that combination of subcortical morphological features such as regional brain volumes, brain surface area, and brain surface area to volume ratio has the potential to discriminate PSP-RS and PD subjects with high accuracy. Moreover, including the morphology of other cortical brain regions does not lead to considerably better classification results, whereas, the addition of clinical and demographic features considerably enhances the predictive power of imaging biomarkers, reflecting the importance of nosological constructs in an automatic classification paradigm.

Acknowledgements This work was supported by Parkinson Alberta.

Conflict of interest A.S. Talai, Z. Ismail, J. Sedlacik, K. Boelmans and N.D. Forkert declare that they have no competing interests.

References

1. Tolosa E, Wenning G, Poewe W. The diagnosis of Parkinson's disease. Lancet Neurol. 2006;5:75–86.
2. Olanow CW, Hauser RA, Jankovic J, Langston W, Lang A, Poewe W, Tolosa E, Stocchi F, Melamed E, Eyal E, Rascol O. A randomized, double-blind, placebo-controlled, delayed start study to assess rasagiline as a disease modifying therapy in Parkinson's disease (the ADAGIO study): rationale, design, and baseline characteristics. Mov Disord. 2008;23:2194–201.
3. Pellicano C, Assogna F, Cellupica N, Piras F, Pierantozzi M, Stefani A, Cerroni R, Mercuri B, Caltagirone C, Pontieri FE, Spalletta G. Neuropsychiatric and cognitive profile of early Richardson's syndrome, progressive supranuclear Palsy-parkinsonism and Parkinson's disease. Parkinsonism Relat Disord. 2017;45:50–6.
4. Hughes AJ, Daniel SE, Ben-Shlomo Y, Lees AJ. The accuracy of diagnosis of parkinsonian syndromes in a specialist movement disorder service. Brain. 2002;125:861–70.
5. Litvan I, Agid Y, Calne D, Campbell G, Dubois B, Duvoisin RC, Goetz CG, Golbe LI, Grafman J, Growdon JH, Hallett M, Jankovic

J, Quinn NP, Tolosa E, Zee DS. Clinical research criteria for the diagnosis of progressive supranuclear palsy (Steele-Richardson-Olszewski syndrome): report of the NINDS-SPSP International Workshop. Neurology. 1996;47:1–9.
6. Bensimon G, Ludolph A, Agid Y, Vidailhet M, Payan C, Leigh PN; NNIPPS Study Group. Riluzole treatment, survival and diagnostic criteria in Parkinson plus disorders: the NNIPPS study. Brain. 2009;132:156–71.
7. Davie CA. A review of Parkinson's disease. Br Med Bull. 2008;86: 109–27.
8. Marx S, Respondek G, Stamelou M, Dowiasch S, Stoll J, Bremmer F, Oertel WH, Höglinger GU, Einhäuser W. Validation of mobile eye-tracking as novel and efficient means for differentiating progressive supranuclear palsy from Parkinson's disease. Front Behav Neurosci. 2012;6:1–11.
9. Egerton T, Williams DR, Iansek R. Comparison of gait in progressive supranuclear palsy, Parkinson's disease and healthy older adults. BMC Neurol. 2012;12:116.
10. Tang CC, Poston KL, Eckert T, Feigin A, Frucht S, Gudesblatt M, Dhawan V, Lesser M, Vonsattel JP, Fahn S, Eidelberg D. Differential diagnosis of parkinsonism: a metabolic imaging study using pattern analysis. Lancet Neurol. 2010;9:149–58.
11. Eckert T, Sailer M, Kaufmann J, Schrader C, Peschel T, Bodammer N, Heinze HJ, Schoenfeld MA. Differentiation of idiopathic Parkinson's disease, multiple system atrophy, progressive supranuclear palsy, and healthy controls using magnetization transfer imaging. Neuroimage. 2004;21:229–35.
12. Forkert ND, Schmidt-Richberg A, Treszl A, Hilgetag C, Fiehler J, Münchau A, Handels H, Boelmans K. Automated volumes-of-interest identification for classical and atypical parkinsonian syndrome differentiation using T2' MR imaging. Methods Inf Med. 2013;52:128–36.
13. Duchesne S, Rolland Y, Verin M. Automated computer differential classification in parkinsonian syndromes via pattern analysis on MRI. Acad Radiol. 2009;16:61–70.
14. Quattrone A, Nicoletti G, Messina D, Fera F, Condino F, Pugliese P, Lanza P, Barone P, Morgante L, Zappia M, Aguglia U, Gallo O. MR imaging index for differentiation of progressive supranuclear palsy from Parkinson disease and the Parkinson variant of multiple system atrophy. Radiology. 2008;246:214–21.
15. Messina D, Cerasa A, Condino F, Arabia G, Novellino F, Nicoletti G, Salsone M, Morelli M, Lanza PL, Quattrone A. Patterns of brain atrophy in Parkinson's disease, progressive supranuclear palsy and multiple system atrophy. Parkinsonism Relat Disord. 2011;17:172–6.
16. Focke NK, Helms G, Scheewe S, Pantel PM, Bachmann CG, Dechent P, Ebentheuer J, Mohr A, Paulus W, Trenkwalder C. Individual voxel-based subtype prediction can differentiate progressive supranuclear palsy from idiopathic Parkinson syndrome and healthy controls. Hum Brain Mapp. 2011;32:1905–15.
17. Gama RL, Távora DF, Bomfim RC, Silva CE, Bruin VM, Bruin PF. Morphometry MRI in the differential diagnosis of parkinsonian syndromes. Arq Neuropsiquiatr. 2010;68:333–8.
18. Price S, Paviour D, Scahill R, Stevens J, Rossor M, Lees A, Fox N. Voxel-based morphometry detects patterns of atrophy that help differentiate progressive supranuclear palsy and Parkinson's disease. Neuroimage. 2004;23:663–9.
19. Scherfler C, Göbel G, Müller C, Nocker M, Wenning GK, Schocke M, Poewe W, Seppi K. Diagnostic potential of automated subcortical volume segmentation in atypical parkinsonism. Neurology. 2016;86:1242–9.
20. Sarica A, Critelli C, Guzzi PH, Cerasa A, Quattrone A, Cannataro M. Application of different classification techniques on brain morphological data. Comput Med Syst (CBMS), 2013 IEEE 26th Int Symp on IEEE. 2013. pp. 425–8.
21. Lee JH, Han YH, Kang BM, Mun CW, Lee SJ, Baik SK. Quantitative assessment of subcortical atrophy and iron content in progressive supranuclear palsy and parkinsonian variant of multiple system atrophy. J Neurol. 2013;260:2094–101.
22. Sakurai K, Tokumaru AM, Shimoji K, Murayama S, Kanemaru K, Morimoto S, Aiba I, Nakagawa M, Ozawa Y, Shimohira M, Matsukawa N, Hashizume Y, Shibamoto Y. Beyond the midbrain atrophy: wide spectrum of structural MRI finding in cases of pathologically proven progressive supranuclear palsy. Neuroradiology. 2017;59:431–43.
23. Dotson VM, Szymkowicz SM, Sozda CN, Kirton JW, Green ML, O'Shea A, McLaren ME, Anton SD, Manini TM, Woods AJ. Age differences in prefrontal surface area and thickness in middle aged to older adults. Front Aging Neurosci. 2016;7:1–9.
24. Jubault T, Gagnon JF, Karama S, Ptito A, Lafontaine AL, Evans AC, Monchi O. Patterns of cortical thickness and surface area in early Parkinson's disease. Neuroimage. 2011;55:462–7.
25. Gerrits NJ, van Loenhoud AC, van den Berg SF, Berendse HW, Foncke EM, Klein M, Stoffers D, van der Werf YD, van den Heuvel OA. Cortical thickness, surface area and subcortical volume differentially contribute to cognitive heterogeneity in Parkinson's disease. PLoS ONE. 2016;11:1–14.
26. Dickerson BC, Feczko E, Augustinack JC, Pacheco J, Morris JC, Fischl B, Buckner RL. Differential effects of aging and Alzheimer's disease on medial temporal lobe cortical thickness and surface area. Neurobiol Aging. 2009;30:432–40.
27. Worker A, Blain C, Jarosz J, Chaudhuri KR, Barker GJ, Williams SC, Brown R, Leigh PN, Simmons A. Cortical thickness, surface area and volume measures in Parkinson's disease, multiple system atrophy and progressive supranuclear palsy. PLoS ONE. 2014;9:1–15.
28. Planetta PJ, Ofori E, Pasternak O, Burciu RG, Shukla P, DeSimone JC, Okun MS, McFarland NR, Vaillancourt DE. Free-water imaging in Parkinson's disease and atypical parkinsonism. Brain. 2016;139:495–508.
29. Hirschauer TJ, Adeli H, Buford JA. Computer-aided diagnosis of Parkinson's disease using enhanced probabilistic neural network. J Med Syst. 2015;39:179.
30. Long D, Wang J, Xuan M, Gu Q, Xu X, Kong D, Zhang M. Automatic classification of early Parkinson's disease with multi-modal MR imaging. PLoS ONE. 2012;7:1–9.
31. Péran P, Cherubini A, Assogna F, Piras F, Quattrocchi C, Peppe A, Celsis P, Rascol O, Démonet JF, Stefani A, Pierantozzi M, Pontieri FE, Caltagirone C, Spalletta G, Sabatini U. Magnetic resonance imaging markers of Parkinson's disease nigrostriatal signature. Brain. 2010;133:3423–33.
32. Morisi R, Cha M, Arafa M, Zagrouba E. Binary and multi-class parkinsonian disorders classification using support vector machines. In: Lecture notes in computer science. 2015. pp. 379–86.
33. Ota M, Nakata Y, Ito K, Kamiya K, Ogawa M, Murata M, Obu S, Kunugi H, Sato N. Differential diagnosis tool for parkinsonian syndrome using multiple structural brain measures. Comput Math Methods Med. 2013;2013:571289.
34. Segovia F, Illán IA, Górriz JM, Ramírez J, Rominger A, Levin J. Distinguishing Parkinson's disease from atypical parkinsonian syndromes using PET data and a computer system based on support vector machines and Bayesian networks. Front Comput Neurosci. 2015;9:1–8.
35. Prashanth R, Dutta Roy S, Mandal PK, Ghosh S. High-accuracy detection of early Parkinson's disease through multimodal features and machine learning. Int J Med Inform. 2016;90:13–21.
36. Boelmans K, Holst B, Hackius M, Finsterbusch J, Gerloff C, Fiehler J, Münchau A. Brain iron deposition fingerprints in Parkinson's disease and progressive supranuclear palsy. Mov Disord. 2012;27:421–7.

37. Fellner F, Holl K, Held P, Fellner C, Schmitt R, Böhm-Jurkovic H. A T1-weighted rapid three-dimensional gradient-echo technique (MP-RAGE) in preoperative MRI of intracranial tumours. Neuroradiology. 1996;38:199–206.
38. Mazziotta J, Toga A, Evans A, Fox P, Lancaster J, Zilles K, Woods R, Paus T, Simpson G, Pike B, Holmes C, Collins L, Thompson P, MacDonald D, Iacoboni M, Schormann T, Amunts K, Palomero-Gallagher N, Geyer S, Parsons L, Narr K, Kabani N, Le Goualher G, Boomsma D, Cannon T, Kawashima R, Mazoyer B. A probabilistic atlas and reference system for the human brain: International Consortium for Brain Mapping (ICBM). Philos Trans R Soc Lond, B, Biol Sci. 2001;356:1293–322.
39. Ourselin S, Roche A, Subsol G, Pennec X, Ayache N. Reconstructing a 3D structure from serial histological sections. Image Vis Comput. 2001;19:25–31.
40. Modat M, Ridgway GR, Taylor ZA, Lehmann M, Barnes J, Hawkes DJ, Fox NC, Ourselin S. Fast free-form deformation using graphics processing units. Comput Methods Programs Biomed. 2010;98:278–84.
41. Kononenko I, Šimec E, Robnik-Šikonja M. Overcoming the myopia of inductive learning algorithms with RELIEFF. Appl Intell. 1997;7:39–55.
42. Chang C, Lin C. LIBSVM: A Library for Support Vector Machines. ACM Trans Intell Syst Technol. 2011;2:27:1-27.
43. Du G, Lewis MM, Kanekar S, Sterling NW, He L, Kong L, Li R, Huang X. Combined diffusion tensor imaging and apparent transverse relaxation rate differentiate Parkinson disease and atypical parkinsonism. AJNR Am J Neuroradiol. 2017;38:966–72.
44. Dubois B, Pillon B. Cognitive deficits in Parkinson's disease. J Neurol. 1996;244:2–8.
45. Salvatore C, Cerasa A, Castiglioni I, Gallivanone F, Augimeri A, Lopez M, Arabia G, Morelli M, Gilardi MC, Quattrone A. Machine learning on brain MRI data for differential diagnosis of Parkinson's disease and progressive supranuclear palsy. J Neurosci Methods. 2014;222:230–7.
46. Cherubini A, Morelli M, Nisticó R, Salsone M, Arabia G, Vasta R, Augimeri A, Caligiuri ME, Quattrone A. Magnetic resonance support vector machine discriminates between Parkinson disease and progressive supranuclear palsy. Mov Disord. 2014;29:266–9.
47. Postuma RB, Berg D, Stern M, Poewe W, Olanow CW, Oertel W, Obeso J, Marek K, Litvan I, Lang AE, Halliday G, Goetz CG, Gasser T, Dubois B, Chan P, Bloem BR, Adler CH, Deuschl G. MDS clinical diagnostic criteria for Parkinson's disease. Mov Disord. 2015;30:1591–601.
48. Boxer AL, Yu JT, Golbe LI, Litvan I, Lang AE, Höglinger GU. Advances in progressive supranuclear palsy: new diagnostic criteria, biomarkers, and therapeutic approaches. Lancet Neurol. 2017;16:552–63.