Available online at www.sciencedirect.com

ScienceDirect

Biomedical Journal

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


Original Article

# Diffusion tensor imaging for the differential diagnosis of Parkinsonism by machine learning

*Chih-Chien Tsai* [a,1], *Yao-Liang Chen* [b,c,1], *Chin-Song Lu* [d,e,f], *Jur-Shan Cheng* [g,h], *Yi-Hsin Weng* [e,i,j], *Sung-Han Lin* [k], *Yi-Ming Wu* [c,l], *Jiun-Jie Wang* [a,b,j,l,m,*]

[a] *Healthy Aging Research Center, Chang Gung University, Taoyuan, Taiwan*
[b] *Department of Diagnostic Radiology, Chang Gung Memorial Hospital, Keelung, Taiwan*
[c] *Department of Medical Imaging and Intervention, Chang Gung Memorial Hospital, Linkou, Taoyuan, Taiwan*
[d] *Professor Lu Neurological Clinic, Taoyuan, Taiwan*
[e] *Division of Movement Disorders, Department of Neurology, Chang Gung Memorial Hospital, Linkou, Taoyuan, Taiwan*
[f] *Department of Neurology, Landseed International Hospital, Taoyuan, Taiwan*
[g] *Clinical Informatics and Medical Statistics Research Center, Chang Gung University, Taoyuan, Taiwan*
[h] *Division of Hepatology, Department of Gastroenterology and Hepatology, Chang Gung Memorial Hospital, Taoyuan, Taiwan*
[i] *School of Medicine, Chang Gung University, Taoyuan, Taiwan*
[j] *Neuroscience Research Center, Chang Gung Memorial Hospital, Linkou, Taoyuan, Taiwan*
[k] *Advanced Imaging Research Center, University of Texas Southwestern Medical Center, Dallas, TX, USA*
[l] *Department of Medical Imaging and Radiological Sciences, Chang Gung University, Taoyuan, Taiwan*
[m] *Institute for Radiological Research, Chang Gung University/Chang Gung Memorial Hospital, Linkou, Taoyuan, Taiwan*

ARTICLE INFO

*Article history:*
Received 7 June 2021
Accepted 27 May 2022
Available online xxx

*Keywords:*
Diffusion tensor imaging
Machine learning
Differential diagnosis
Idiopathic Parkinson's disease
Parkinson-plus syndromes

ABSTRACT

*Background:* There are currently no specific tests for either idiopathic Parkinson's disease or Parkinson-plus syndromes. The study aimed to investigate the diagnostic performance of features extracted from the whole brain using diffusion tensor imaging concerning parkinsonian disorders.

*Methods:* The retrospective data yielded 625 participants (average age: 61.4 ± 8.2, men/women: 313/312; healthy controls/idiopathic Parkinson's disease/multiple system atrophy/progressive supranuclear palsy: 219/286/51/69) between 2008 and 2017. Diffusion-weighted images were obtained using a 3T MR scanner. The 90th, 50th, and 10th percentiles of fractional anisotropy and mean/axial/radial diffusivity from each parcellated brain area were recorded. Statistical analysis was evaluated based on the features extracted from the whole brain, as determined using discriminant function analysis and support vector machine. 20% of the participants were used as an independent blind dataset with 5 times cross-verification. Diagnostic performance was evaluated by the sensitivity and the F1 score.

* Corresponding author. Department of Medical Imaging and Radiological Sciences, Chang Gung University, 259 WenHua 1st Rd., Taoyuan 333, Taiwan.
E-mail address: jwang@mail.cgu.edu.tw (J.-J. Wang).
Peer review under responsibility of Chang Gung University.
[1] These authors contributed equally to this work.
https://doi.org/10.1016/j.bj.2022.05.006
2319-4170/© 2022 Chang Gung University. Publishing services by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Please cite this article as: Tsai C-C et al., Diffusion tensor imaging for the differential diagnosis of Parkinsonism by machine learning, Biomedical Journal, https://doi.org/10.1016/j.bj.2022.05.006

*Results:* Diagnoses were accurate for distinguishing idiopathic Parkinson's disease from healthy control and Parkinson-plus syndromes (87.4 ± 2.1% and 82.5 ± 3.9%, respectively). Diagnostic F1 scores varied for Parkinson-plus syndromes with 67.2 ± 3.8% for multiple system atrophy and 71.6 ± 3.5% for progressive supranuclear palsy. For early and late detection of idiopathic Parkinson's disease, the diagnostic performance was 79.2 ± 7.4% and 84.4 ± 6.9%, respectively. The diagnostic performance was 68.8 ± 11.0% and 52.5 ± 8.9% in early and late detection to distinguish different Parkinson-plus syndromes.
*Conclusions:* Features extracted from diffusion tensor imaging of the whole brain can provide objective evidence for the diagnosis of healthy control, idiopathic Parkinson's disease, and Parkinson-plus syndromes with fair to very good diagnostic performance.

## At a glance commentary

### *Scientific background on the subject*

Different extents of cortical involvement might be observed in subtypes of Parkinson Disease. Differential diagnosis can be possible based on such topological difference. Diffusion MRI provides functional information which can be more sensitive than conventional morphological images. Machine learning algorithm is necessary for the reduction of the number of features and for subsequent classification.

### *What this study adds to the field*

Our algorithm showed the best classification in comparison was by using three-step procedure. The F1 score in the blind validation was 87.4% when separating patients from healthy controls. The F1 score was 82.5 % in separating idiopathic Parkinson's disease from atypical PD. The classification using diffusion MRI showed better performance than using T1 images. The performances were similar between using support vector machine and discriminant function.

## Introduction

Idiopathic Parkinson's disease (PD) is a neurodegenerative disorder characterized by the loss of dopaminergic neurons in the substantia nigra [1,2]. In Parkinson-plus syndrome, the typical symptoms are accompanied by additional features and progress much more quickly. Several subtypes have been identified, including progressive supranuclear palsy (PSP) and multiple system atrophy (MSA). PSP is an extrapyramidal disorder characterized by supranuclear vertical gaze palsy and postural instability [3]. Patients with MSA present with autonomic dysfunction and cerebellar ataxia.
There are currently no specific tests for Parkinsonism, which are diagnosed in accordance with generally accepted criteria based on clinical findings. Although patients with idiopathic PD typically respond well to dopaminergic treatment, patients with Parkinson-plus syndrome exhibit transient or no obvious responses. One of the aims of our study is to test the possibility of the proposed algorithm as applied close to real clinical scenario, rather than limited to a study specific, single imaging protocol in a scanner from a medical center. Because therapeutic interventions vary based on the subtype of the disease, early, efficient, and accurate methods for differential diagnosis are imperative for ensuring appropriate and timely intervention.

While previous structural MRI studies [4] have demonstrated that patients with idiopathic PD and Parkinson-plus syndrome exhibit distinct types of atrophy in specific regions, potentially allowing for discrimination of the two conditions, the authors reported only moderate accuracy [5]. Diffusion tensor imaging (DTI) has revealed that patients with idiopathic PD exhibit reduced fractional anisotropy (FA) in the frontal lobes [6] and cerebellar/orbitofrontal cortices [7]. FA change also correspond to neuropathological change in specific brain regions related to Parkinson-plus syndrome, such as accumulation of $\alpha$-synuclein in temporal cortices and cerebellum in patients with MSA [8], frontal lobe dysfunction in PSP [9]. In the light of this, differences in tensor change could be located in specific cortical regions among different parkinsonism, but not limited in basal ganglia [10,11].

As the different subtypes of Parkinsonism are associated with differences in medical history, clinical expression, and pathogenesis, we hypothesized that each subtype would be associated with a different extent of cortical involvement. Therefore, a comprehensive analysis of all potentially involved regions in the whole brain would be required for an accurate, imaging-based differential diagnosis. In the present study, we designed a serial of classification procedures that aimed to evaluate the best diagnostic performance of DTI concerning the various subtypes of Parkinsonism.

## Materials and methods

This retrospective study re-analyzed images collected from 3 prospective studies during 2008–2017. All studies were approved by the local Institutional Review Board and conducted following the Declaration of Helsinki. The informed consent was waived owing to the retrospective nature of the current study.

A total of 655 participants were recruited from three retrospective studies between June 2008 and May 2017.

Please cite this article as: Tsai C-C et al., Diffusion tensor imaging for the differential diagnosis of Parkinsonism by machine learning, Biomedical Journal, https://doi.org/10.1016/j.bj.2022.05.006

Twenty-six participants failed to pass the eligibility criteria (1 tumor, 1 brain surgery, 3 memory loss, 2 infarct, 2 claustrophobia, 10 failed in physical/neurological examination, and 7 corticobasal syndrome). Four participants met the exclusion criteria of poor image quality. The study sample consisted of 625 patients (men/women: 313/312, mean age: 61.4 ± 8.2 years; mean disease duration: 4.9 ± 3.6 years; Healthy control (HC)/ idiopathic PD/MSA/PSP: 219/286/51/69).

Diagnoses of PD was made by two senior neurologists (28 and 21 years of experience) by the diagnostic criteria outlined by the National Institute of Neurological Disorders and Stroke (NINDS) of USA. Diagnoses were further supported by nuclear medicine examinations (TRODAT). Diagnoses of PSP and MSA were further made with criteria specified by Litvan et al. [12] and Gilman et al. [13], respectively. Exclusion criteria were as follows: (1) moderate/severe dementia; (2) severe dyskinesia; (3) significant major systemic disease; (4) brain abnormality by MRI or $^{18}$FDG PET; (5) history of intracranial operation; (6) significant physical/neuropsychiatric disorder; (7) medication pass through blood-brain-barrier or chronic medication use >10 years, with the exception of medication for parkinsonism; (8) general MRI exclusion criteria. To avoid the effects of drugs, all anti-parkinsonian medications were withdrawn at least 12 h before clinical assessment. The clinical severity was evaluated by using Unified Parkinson Disease Rating Scale (UPDRS) [14]. The disease staging was assessed by Modified Hoehn and Yahr Staging (MHY) [15]. Table 1 summarizes the demographic characteristics and study protocols for participants. The structural MR images were independently read by two neuro-radiologists (19 and 10 years of experience).

### MRI acquisition and image processing

Images were acquired using a 3T MR scanner (Trio, Siemens, Germany) with a 12-channel head matrix coil. Contiguous axial T1-weighted images were acquired using an MPRAGE sequence (echo time/repetition time = 2.63 ms/2000 ms, flip angle = 9°, field-of-view = 224 mm/256 mm, matrix size = 224 × 256, voxel size = 1 × 1 × 1, acquisition time = 4 min 08 s). Diffusion-weighted images were obtained using a spin-echo echo-planar-imaging sequence under different acquisition protocols as listed in Table 1. Only images of b-value of 0 and 1000 s/mm$^2$ were included in the final analysis if multiple b values were acquired. All data reconstruction was performed followed the protocol by Tsai et al. [16]. Diffusion tensor was reconstructed by using Diffusion Kurtosis Estimator [17] from diffusion-weighted images. The tensor derived index used in the current study included axial/radial/mean diffusivity (AD, RD, MD), and fractional anisotropy (FA). T1-weighted images were first co-registered to non-diffusion weighted B0 images. The co-registered T1-weighted images were subsequently normalized to a structural template (Montreal Neurological Institute152 [18]) using affine transformation. The transformation matrix was subsequently inversed transformed and applied to the Automatic Anatomic Labelling [19] template. The warped template was used as a mask to obtain region of interest from the tensor. Each parcellated region of interest was extracted from the individual map of tensor derived index.

From each region in each tensor derived index, the values of 90th, 50th, and 10th percentiles were extracted. This procedure resulted in 1392 features, i.e. combination of 4 tensor-derived indices, 116 regions, and 3 percentiles.

Features extracted from conventional structural images included cortical surface area and thickness, which were calculated from T1-weighted images using FreeSurfer 5.3 [20,21]. Thirty-four cortical regions based on the Desikan-Killiany atlas [21] were calculated in each hemisphere, which leads to 136 features in each participant.

### Statistical analysis

Statistical analyses were performed using SPSS (Version22, SPSS Inc, USA). Differences in age, disease duration, and clinical severity (the total and the motor subscale of Unified Parkinson's Disease Rating Scale) were evaluated via ANOVA, followed by *post hoc* analysis (least significant difference). Differences of diffusion tensor indices/structural features between each patient group and HC were evaluated by Mann-Whitney U test. The $p$ value <0.05 with Bonferroni correction of multiple comparison (0.05/1392 for the diffusion tensor indices; 0.05/136 for the structural features) was regarded as significant.

The possible interaction between the acquisition protocols (Protocol A, B, and C) and disease status (each patient group and healthy control) for diffusion tensor indices was examined by two-way ANOVA test [22]. The acquisition protocols and the disease status were set as categorical explanatory variables. The two-way ANOVA test examined the effect from interaction between the acquisition protocols and the disease status for the diffusion tensor indices. The $p$-value <0.05 with Bonferroni correction of multiple comparisons (0.05/1392 for the diffusion tensor indices) was regarded as significant.

Statistical analysis was conducted by statistician (7 years of experience). The original data were randomly splitted into training dataset (80% n = 498; HC/PD/PSP/MSA: 175/228/55/40) and independent blind dataset (20% n = 127; HC/PD/PSP/MSA: 44/58/14/11) with five times cross-verified. Patients in the blind dataset were further divided into the group with long (>5 years) and short (≦ 5 years) disease duration [23,24].

### Differential diagnosis process

The differential diagnosis process consisted of feature reduction and classification. We examined the diagnostic performance from three different processes: a), single-step: HC, PD, MSA and PSP; b), two-step; step 1: HC vs Patients; step 2: 3 groups: PD, MSA and PSP; c), three-step; step 1: HC vs Patients; step 2: PD vs Parkinson-plus syndromes; step 3: PSP vs MSA.

The diagnostic performance for each classification was further assessed using the leave-one-out validation in the training dataset, followed by blind validation in the independent dataset. Figures described the process [Fig. 1] and summarized the primary result of F1 score, i.e. the harmonic mean of precision and recall [Fig. 2: three-step, Supplementary Fig. 1A: two-step, Supplementary Fig. 1B: single-step].

#### Feature reduction

Feature reduction was conducted in order to minimize the chance of potential model overfitting. The number of the features after reduction was expected to meet the following two criteria. 1.)

Please cite this article as: Tsai C-C et al., Diffusion tensor imaging for the differential diagnosis of Parkinsonism by machine learning, Biomedical Journal, https://doi.org/10.1016/j.bj.2022.05.006

**Table 1 General characteristics of participants.**

| | Protocol A | Protocol B | Protocol C | Final analysis |
|---|---|---|---|---|
| | TE: 108 ms, TR:5700 ms | TE: 96 ms, TR:8200 ms | TE: 83 ms, TR:7800 ms | |
| **Voxel size** | 2*2*3 | 2*2*2 | 2*2*2 | |
| **Gradient** | 30 | 64 | 64 | |
| **b value** | 1000/2000/3000 | 1000 | 1000/2000 | |
| **Matrix size** | 96*96 | 128*128 | 128*128 | |
| **Slice** | 40 | 64 | 64 | |
| **Scan time** | 9m09s | 9m20s | 8m57s | |
| | HC/PD/MSA/PSP | PD/MSA/PSP | HC/PD/MSA/PSP | HC/PD/MSA/PSP |
| **Number** | 96/150/20/25 | 27/13/15 | 123/109/18/29 | 219/286/51/69 |
| **Age** | 59.7 ± 7.1[a,b]/61.6 ± 6.2[c]/61.0 ± 7.8/69.6 ± 7.4 | 63.7 ± 3.5/65.5 ± 7.7/64.6 ± 6.1 | 62.1 ± 7.1/56.8.4 ± 6.1/62.3 ± 11.5/65.7 ± 4.4 | 61.4 ± 7.5/59.9 ± 9.5/62.6 ± 6.9/66.8 ± 6.4 |
| **Duration** | NA/9.2 ± 6.1[d]/3.7 ± 1.9/4.5 ± 3.0 | 5.6 ± 3.0/4.4 ± 2.9/4.7 ± 2.7 | NA/11.3 ± 7.5[d]/4.6 ± 2.9/4.5 ± 2.1 | NA/8.8 ± 6.5[d]/4.1 ± 2.6/4.7 ± 3.7 |
| **Gender** | | | | |
| Male | 48/82/9/10 | 17/2/5 | 52/66/10/12 | 100/165/21/27 |
| Female | 48/68/11/15 | 10/11/10 | 71/43/8/17 | 119/121/30/42 |
| **MHY** | | | | |
| ≦2 | NA/63/1/0 | NA/NA/NA | NA/32/0/1 | NA/100/1/3 |
| >2 | NA/73/14/15 | NA/NA/NA | NA/38/14/7 | NA/114/32/44 |
| **PET positive** | NA/150/20/25 | 27/13/15 | NA/109/18/29 | NA/286/51/69 |
| **UPDRS** | NA/38.1 ± 24.8[e]/61.6 ± 25.5/43.3 ± 24.0 | NA/NA/NA | NA/39.4 ± 25.9[d]/59.8 ± 24.0/62.6 ± 24.0 | NA/38.2 ± 24.8[d]/61.0 ± 22.9/46.3 ± 24.8 |
| **UPDRS_III** | NA/22.8 ± 15.2[e]/37.4 ± 17.7/23.6 ± 14.9 | NA/NA/NA | NA/22.8 ± 16.1[d]/36.6 ± 19.0/38.9 ± 17.0 | NA/22.7 ± 15.2[e]/38.3 ± 18.2/34.0 ± 17.0 |

Abbreviations: TR: Repetition time; TE: Echo time; MSA: Multiple System Atrophy; PSP: Progressive Supranuclear Palsy; PD: Parkinson's Disease; HC: Healthy control; MHY: Modified Hoehn and Yahr Stage; UPDRS: Unified Parkinson's Disease Rating Scale; UPDRS_III: Motor subscale of UPDRS; NA: Not available.
Final is the result presented in the study.
[a] HC versus PSP.
[b] HC versus PD.
[c] PD versus PSP.
[d] PD versus MSA and PSP.
[e] PD versus MSA.

Please cite this article as: Tsai C-C et al., Diffusion tensor imaging for the differential diagnosis of Parkinsonism by machine learning, Biomedical Journal, https://doi.org/10.1016/j.bj.2022.05.006

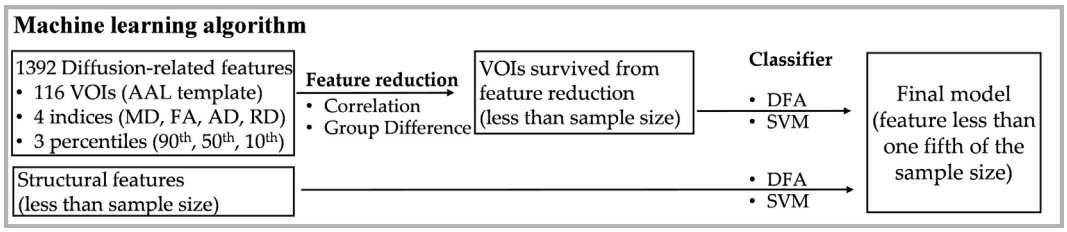


Fig. 1 **Graphical description for the procedures of feature reduction and classification.** (A) Features comprising diffusion parameters from different brain regions or percentiles were reduced by the correlation (Pearson Correlation) and group difference (Kruskal-Wallis test or Mann-Whitney U test). Survived features were then entered into the classifiers (discriminant function analysis/support vector machine). Abbreviations: VOI: volume of interested; MD: mean diffusivity; FA: fractional anisotropy; AD: axial diffusivity; RD: radial diffusivity.

Before being entered into a classification process, the number should be reduced to less than the sample size [25]. 2.) After the classification analysis, the number of the features survived should be less than one fifth of the sample size [26,27]. The feature reduction process removed the redundant features until the first criteria was met. After the classification process, the survived features were examined to make sure of the second criteria. In our study, redundant features were identified through the following.

First, the correlation between any two features was examined by using Pearson Correlation. The 1392 features in our study resulted in total 968832 correlations (1392 * 1392/2). A threshold of high correlation was defined if Pearson Correlation coefficient $r \geq 0.8$ [28]. If a feature involved in 3 high correlations, this feature was excluded.

Participants

| | HC | Patient |
|---|---|---|
| DTI | 83.9 (1.0) / 78.4(2.1)*** | 90.4 (0.7) / 87.4(1.5)*** |
| T1 | 63.4 (3.4) / 55.7(5.8) | 74.8 (2.8) / 70.2(2.6) |

| | PD | PM |
|---|---|---|
| DTI | 84.1 (2.8) / 82.5(3.9)* | 65.7 (4.3) / 62.4(6.9) |
| T1 | 82.7 (1.1) / 75.8(5.0) | 63.9 (1.8) / 53.5 (5.8) |

| | MSA | PSP |
|---|---|---|
| DTI | 68.8 (3.2) / 67.2(3.8)* | 73.8 (3.7) / 71.6(3.5)*** |
| T1 | 64.7 (6.0) / 49.4 (11.1) | 72.1 (4.6) / 57.5 (4.5) |

Fig. 2 **The diagnostic performance by discriminant function analysis in the classification procedures of three-step.** The diagnostic performance by discriminant function analysis was reported along with F1 scores (training/blind) in five-fold cross-validation with the classification procedures of three-step. Comparisons of diagnostic performance using features from diffusion MRI and conventional structural images: *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$. Abbreviations: HC: healthy control; PD: idiopathic Parkinson's disease; MSA: multiple system atrophy; PSP: progressive supranuclear palsy; DTI: Diffusion-tensor imaging; T1: T1-weighted imaging.

Secondly, the between-group difference was examined from those features which survived the correlation examination. The difference between two groups was examined by using Mann-Whitney U test (HC and Patient, PD and Parkinson-plus syndromes, PSP and MSA). The difference examined by Kruskal-Wallis test included a.) among 4 groups of participants in the single-step analysis, b.) among 3 groups of patients in the step 2 of the two-step analysis. The difference was regarded as statistically significant if $p < 0.05/n$, where n is the number of the features that entered the respective test. Only significant features would enter the subsequent classification analysis. Non-significant features were excluded.

*Classification*

The machine learning classifiers in our study was by discriminant function analysis. It was performed using SPSS version 22 (SPSS Inc, Chicago, IL, USA). The discriminant function examined the reduction of Wilk's lambda [29] at each feature's inclusion/exclusion by the probability of the F test ($p < 0.05$, feature entered into the classification model; $p > 0.1$, feature was removed from the classification model). The process stopped when no feature meets the criterion for entry and removal of the analysis [29].

To further support the algorithm, a second classifier which was widely used, support vector machine, was implemented. Support vector machine was performed in MATLAB (Classification Learner toolbox, Matlab version R2018a, The MathWorks, Natick, MA). Principal components were entered into the classification, which was extracted from the survived features in order to reduce the dimensionality. The number of principal components was set as the same as the number of features in the discriminant function.

Differences of diagnostic performance between the leave-one-out cross validation in the training and blind dataset were evaluated by the Mann-Whitney U test. Differences in diagnostic performance among diffusion MRI and conventional structural images were evaluated by the Mann-Whitney U test. The diagnostic performance from combined features was compared with diffusion tensor or structural images alone by the Mann-Whitney U test. The $p$-value < 0.05 was regarded as significant.

Please cite this article as: Tsai C-C et al., Diffusion tensor imaging for the differential diagnosis of Parkinsonism by machine learning, Biomedical Journal, https://doi.org/10.1016/j.bj.2022.05.006

## Results

### *Demographic characteristics*

The entire cohort consisted of 625 participants [Table 1, HC: 219; idiopathic PD: 286; MSA: 51; PSP: 69]. Significant differences in disease duration ($p < 0.001$) and total/motor scores of the UPDRS were observed ($p < 0.001$) among the groups. Post hoc analyses revealed that the patients with idiopathic PD exhibited longer disease duration (relative to patients with MSA and PSP, all $p < 0.001$), lower total UPDRS scores (MSA, $p < 0.001$; PSP, $p = 0.011$), and lower UPDRS motor scores (MSA, $p < 0.001$) than the patients with Parkinson-plus syndromes within a protocol.

### *Differential diagnosis by classifier*

The diagnostic performance from discriminant function analysis was shown in Fig. 2 and Supplementary Fig. 1. To separate patients from HC, the F1 score was 87.4 ± 1.5% using either three- or two-step procedures [Fig. 2 and Supplementary Fig. 1A], and 86.7 ± 3.1% in single-step [Supplementary Fig. 1B and Table 2]. To separate PD patients from those with Parkinson-plus syndromes, the F1 score varied between 82.5 ± 3.9% (three-step) to 63.0 ± 4.6% (two-step). The F1 score can be good for the diagnosis between PSP (71.6 ± 3.5%) and MSA (67.2 ± 3.8%). The diagnostic performance from support vector machine was shown in Supplementary Fig. 2. The sensitivity, specificity, precision, accuracy, and F1 score from discriminant function analysis were reported in Table 2. During the feature reduction process, the survived features from the original 1392 imaging parameters were all reduced to less than the sample size of each validation dataset in all classification processes [Supplementary Table 1].

When compared to the structural imaging, features extracted from diffusion MRI showed improved diagnostic performance, noticeably between patients and HC and among those with Parkinson-plus syndromes [Fig. 2]. The diagnostic performance by combining features from both diffusion and structural images was examined. The F1 score was 84.8 ± 2.8% to separate patients from HC, using either three- or two-step procedures. When separating PD patients from those with Parkinson-plus syndromes the F1 score was 84.7 ± 4.8%. The F1 scores for the diagnosis between PSP and MSA were 61.8 ± 7.8% and 51.7 ± 12.6%, respectively.

Figure 3 showed the distribution of the discriminant scores for distinguishing HC/Patients [Fig. 3A], PD/Parkinson-plus [Fig. 3B], and MSA/PSP [Fig. 3C] in the first training and the blind dataset using three-step procedure. In the blind dataset, the mean discriminant scores for HC/Patients were: −1.45 ± 1.27/1.04 ± 1.35; for PD/Parkinson-plus: 0.47 ± 0.85/-0.64 ± 1.67; and for MSA/PSP: 0.34 ± 0.99/-0.42 ± 0.77.

### *Model performance in Parkinsonism with different duration*

We further evaluated the diagnostic performance under different disease durations [Table 3] and validated in all the 5 different blind datasets. When disease duration is longer than 5 years, the sensitivity and F1 score in the diagnosis of PD patients were 93.0 ± 5.1% and 84.4 ± 6.9%, respectively. In contrast, patients at early stage of the disease (disease duration less than 5 years), the sensitivity and F1 score was 78.2 ± 7.4% and 79.2 ± 7.4%. In the diagnosis of patients with Parkinson-plus syndromes, the sensitivity was 73.9 ± 18.0% if with the short duration and 49.8 ± 29.0% if long. The corresponding F1 score was 68.8 ± 11.0% and 52.5 ± 8.9%, respectively.

**Table 2 Diagnostic performance with discriminant function analysis for all participants from single-, two-, and three-step classification procedures. Classification parameters were calculated from confusion matrix. BOLD value means combining and re-calculating the classification result. Data are reported as means (standard deviations) from 5-times cross validation.**

| Group | Parameters | Procedure | HC | PA[a] | PA | | PM | |
|---|---|---|---|---|---|---|---|---|
| | | | | | PD | PM[a] | MSA | PSP[a] |
| **ALL** | **Sensitivity (%)** | **Three** | – | 84.6 (2.9) | – | 65.6 (8.3) | – | 68.6 (3.9) |
| | | **Two**[b] | – | – | 76.2 (4.5) | **70.0 (2.3)** | 58.2 (16.5) | 48.6 (9.3) |
| | | **Single** | 79.5 (8.5) | **84.8 (3.1)** | 66.5 (1.0) | - | 52.7 (1.8) | 45.7 (3.6) |
| | **Specificity (%)** | **Three** | – | 83.2 (3.4) | – | 80.7 (6.0) | – | 70.9 (4.1) |
| | | **Two**[b] | – | – | 65.6 (10.0) | **75.0 (4.1)** | 91.1 (2.3) | 84.6 (5.6) |
| | | **Single** | 84.8 (2.5) | **78.3 (8.4)** | 81.4 (3.6) | - | 95.2 (1.5) | 91.0 (1.0) |
| | **Accuracy (%)** | **Three** | 84.1 (1.7) | 84.1 (1.7) | 76.1 (4.9) | 76.1 (4.9) | 69.6 (3.6) | 69.6 (3.6) |
| | | **Two**[b] | – | – | 73.0 (3.1) | **73.5 (3.4)** | 86.7 (1.5) | 78.6 (3.1) |
| | | **Single** | 82.9 (4.2) | **83.0 (4.0)** | 74.6 (2.3) | - | 91.5 (1.2) | 86.0 (1.2) |
| | **Precision (%)** | **Three** | 74.2 (3.3) | 90.5 (1.7) | 84.5 (3.4) | 60.1 (8.5) | 64.0 (3.9) | 75.0 (3.3) |
| | | **Two**[b] | – | – | 83.9 (3.4) | **54.9 (4.5)** | 49.9 (6.0) | 39.9 (4.1) |
| | | **Single** | 73.5 (4.8) | **87.9 (4.6)** | 75.2 (3.8) | - | 50.3 (7.0) | 39.4 (4.5) |
| | **F1 score (%)** | **Three** | 78.4 (2.1) | 87.4 (1.5) | 82.5 (3.9) | 62.4 (6.9) | 67.2 (3.8) | 71.6 (3.5) |
| | | **Two**[b] | – | – | 79.8 (2.6) | **63.0 (4.6)** | 53.0 (9.2) | 43.2 (1.7) |
| | | **Single** | 76.4 (6.3) | **86.7 (3.1)** | 70.5 (2.1) | - | 50.8 (12.5) | 42.2 (3.9) |

Abbreviations: HC: healthy control; PA: patients; PD: idiopathic Parkinson's Disease; PM: Parkinson-plus syndrome; MSA: multiple system atrophy; PSP: progressive supranuclear palsy.

[a] Sensitivity of PA/PM/PSP by definition is the same as the specificity of HC/PD/MSA and vice versa in three-step procedure.

[b] Same classification results between HC and PA in two- and three-step procedures.

Please cite this article as: Tsai C-C et al., Diffusion tensor imaging for the differential diagnosis of Parkinsonism by machine learning, Biomedical Journal, https://doi.org/10.1016/j.bj.2022.05.006

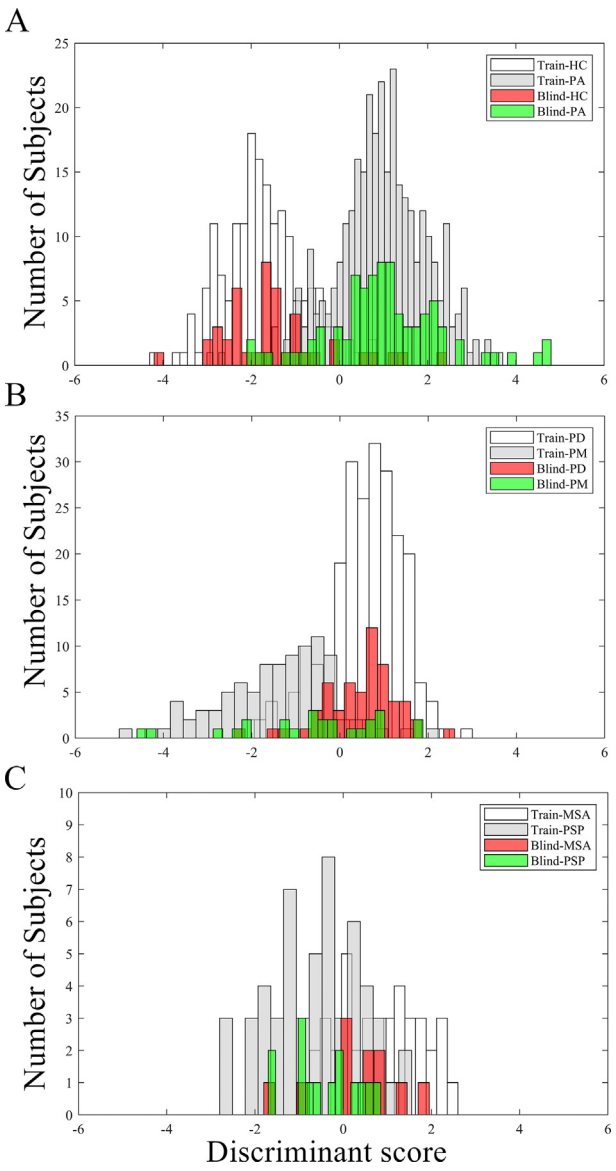


Fig. 3 **Frequency distribution of classification results for the three-step procedure by discriminant function analysis.** This figure depicts the feature from the three-step procedure using the discriminant functions as the coordinates. Distinct differences of discriminant score can be observed between (A) healthy control versus patients, (B) idiopathic Parkinson disease versus Parkinson-plus syndromes, and (C) multiple system atrophy versus progressive supranuclear palsy. Abbreviations: HC: healthy control; PA: patients; PD: idiopathic Parkinson's disease; PM: Parkinson-plus syndromes; MSA: multiple system atrophy; PSP: progressive supranuclear palsy.

To separate patients of PSP from MSA, the sensitivity and the F1 score were 88.3 ± 16.2% and 77.7 ± 13.7%, respectively, when at long disease duration. However, if the disease duration is less than 5 years, the sensitivity was 67.2 ± 7.1% in MSA and 70.1 ± 5.0% in PSP. The corresponding F1 score was 69.6 ± 5.4% and 66.5 ± 8.1%, respectively.

### *Components of classification model*

The identified features in our classification model were located in 24, 11, and 3 brain regions during each stage of the 3 steps, respectively. The selected regions between patients and HC were mostly located in cingulum, the frontal gyrus (superior, middle and inferior part), insula, and superior parietal gyrus. Axial Diffusivity (AD) was the most frequently selected diffusion index (12 times), when compared to either RD (6), MD (5), and FA (7). The selected features were reported in Table 4, including the location, the diffusion index of interest, the corresponding coefficient, and Wilk's lambda of the discriminant function.

### *Differential involvement in Parkinsonism*

For each brain region, we examined differences in the diffusion index in each patient group, relative to HC. In MD, different pattern of cortical involvement could be found in each patient group compared with HC. Significant increases in MD was observed in the triangular part of the supplementary motor area, rectus, calcarine, cuneus, superior parietal, and supramarginal, precuneus in patients with idiopathic PD [Fig. 4A, all $p$ values $\leq 0.000431$]. In patients with MSA, regions exhibiting increased MD were located in the cerebellum, such as superior semilunar, alae, quadrangular lobules, culmen, declive, pyramid, and uvula [Fig. 4B]. In contrast, patients with PSP [Fig. 4C] exhibited increased MD in the superior, middle, and inferior frontal gyrus, rolandic operculum, hippocampus, superior occipital, putamen, and thalamus. The involved regions in AD were visualized in Supplementary Fig. 3. However, significant increase in AD can be noticed in widespread brain regions. Significant change in FA and RD were shown in Supplementary Figs. 4–5.

The difference of cortical surface area [Supplementary Fig. 6] and thickness [Supplementary Fig. 7] were shown for each patients group when compared to HC. For patients with idiopathic PD and MSA, the difference was mainly reduced cortical thickness. The overlapped regions in both comparisons included the temporal lobe (superior, middle, and inferior gyrus), and supramarginal gyrus [Supplementary Fig. 7A and 7B, respectively]. The other regions with reduced thickness in MSA were located in medial part of orbital frontal gyrus and anterior cingulate gyrus.

In patients with PSP, significant reduction was found in cortical surface area [Supplementary Fig. 6C] and thickness [Supplementary Fig. 7C]. The involved regions were located in the frontal lobe (superior and middle), temporal lobe (superior, middle, and inferior), supramarginal, precentral gyrus, postcentral gyrus, cingulate gyrus, cuneus, calcarine and lateral occipital gyrus.

## Discussion

### *Differential diagnosis based on whole-brain DTI*

In the present study, we examined the potential of DTI for the differential diagnosis of PD using a multivariate approach.

Please cite this article as: Tsai C-C et al., Diffusion tensor imaging for the differential diagnosis of Parkinsonism by machine learning, Biomedical Journal, https://doi.org/10.1016/j.bj.2022.05.006

**Table 3 Diagnostic performance with discriminant function analysis for patients with different disease duration from single-, two-, and three-step classification procedures. Classification parameters (duration >5/duration ≦5) were calculated from confusion matrix. Data are reported as means (standard deviations) from 5-times cross validation.**

| Group | Parameters | Procedure | HC | PA | PD | PA | PM | PM | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | MSA | PSP |
| Duration>5/≦5 | Sensitivity | Three steps | – | 81.8(2.9)/81.4(4.3) | 93.0(5.1)/78.2(7.4) | 49.8(29.0)/73.9(18.0) | | 56.7(25.3)/67.2(7.1) | 88.3(16.2)/70.1(5.0) |
| | | Two steps | – | 81.8(2.9)/81.4(4.3) | 79.7(16.3)/68.5(24.1) | – | | 63.3(41.5)/57.0(14.0) | 66.4(26.8)/55.7(11.2) |
| | | Single step | 78.6(7.6)/78.6(7.6) | – | 71.7(16.7)/59.2(22.0) | – | | 46.7(27.4)/57.8(12.0) | 50.3(24.6)/41.2(17.9) |
| | Specificity | Three steps | – | 84.5(4.9)/82.9(5.4) | 49.8(29.0)/73.9(18.0) | 93.0(5.1)/78.2(7.4) | | 88.3(16.2)/70.1(5.0) | 56.7(25.3)/67.2(7.1) |
| | | Two steps | – | 84.5(4.9)/82.9(5.4) | 85.8(14.2)/73.0(9.8) | – | | 89.6(4.3)/91.9(2.7) | 86.4(14.9)/75.0(21.6) |
| | | Single step | 81.8(2.8)/86.9(5.1) | – | 88.5(7.1)/84.7(7.6) | – | | 94.2(2.6)/94.6(3.4) | 91.4(4.0)/88.8(8.2) |
| | Accuracy | Three steps | – | 82.9(2.6)/82.0(3.5) | 76.8(9.0)/75.2(8.5) | 76.8(9.0)/75.2(8.5) | | 72.8(16.2)/68.4(5.3) | 72.8(16.2)/68.4(5.3) |
| | | Two steps | – | 82.9(2.6)/82.0(3.5) | 79.4(13.6)/74.0(7.5) | – | | 88.4(3.1)/83.8(6.5) | 81.4(10.3)/72.1(12.2) |
| | | Single step | 80.2(4.7)/82.6(5.0) | – | 81.6(9.8)/80.0(5.1) | – | | 93.2(2.5)/90.6(4.5) | 87.5(6.1)/83.5(6.1) |
| | Precision | Three steps | – | 83.4(7.1)/82.0(5.8) | 78.3(13.5)/82.0(14.4) | 73.1(21.0)/67.0(13.5) | | 83.3(23.6)/72.6(7.1) | 70.7(16.7)/63.9(11.7) |
| | | Two steps | – | 83.4(7.1)/82.0(5.8) | 81.1(31.7)/71.8(12.5) | – | | 24.7(16.3)/66.1(9.0) | 51.8(31.8)/49.6(12.9) |
| | | Single step | 81.9(3.9)/86.6(4.3) | – | 67.7(20.3)/57.5(7.1) | – | | 19.0(12.6)/59.4(13.2) | 37.8(29.8)/37.0(7.7) |
| | F1 | Three steps | – | 82.4(3.4)/81.6(3.6) | 84.4(6.9)/79.2(7.4) | 52.5(8.9)/68.8(11.0) | | 64.8(20.9)/69.6(5.4) | 77.7(13.7)/66.5(8.1) |
| | | Two steps | – | 82.4(3.4)/81.6(3.6) | 74.6(21.1)/69.4(19.1) | – | | 44.1(13.8)/60.8(11.6) | 41.5(25.6)/37.5(9.3) |
| | | Single step | 80.2(5.4)/82.3(5.2) | – | 65.8(10.0)/56.7(16.3) | – | | 33.0(7.7)/58.3(12.0) | 41.5(25.6)/37.5(9.3) |

Abbreviations: HC: healthy control; PA: patients; PD: idiopathic Parkinson's Disease; PM: Parkinson-plus syndrome; MSA: multiple system atrophy; PSP: progressive supranuclear palsy.

Features based on the diffusion indices were measured from the parceled regions, which covered the whole brain. A machine learning algorithm was implemented, which consisted of the reduction of the number of features and a classification process. The results from two comparable classification methods were presented, discriminant function analysis and support vector machine. Using a cohort of 625 subjects, both yielded similar conclusion.

The differential diagnostic performance using features extracted from diffusion tensor was reported. The discriminant function analysis and support vector machine used in our algorithm shown the similar diagnostic performance (F1 score for separating patients from HC: DFA, 87.4 ± 1.5%; SVM, 85.4 ± 2.5%). The features extracted from diffusion MRI had improved diagnostic performance than that with conventional structural images in both discriminant function analysis and support vector machine methods. Our result showed that the best diagnostic strategy for differential diagnosis can be achieved through three binary steps, i.e. by distinguishing patients from healthy controls, subsequently by patients with Parkinson-plus syndromes from idiopathic PD, and finally between patients with PSP and MSA. The diagnostic performance, as reflected by the F1 score, varied between 82.5% (PD) and 67.2% (MSA). Early detection (less than 5 years) is possible, where best sensitivity is 70.1% for the diagnosis of MSA and PSP. At the early stage of the disease, the F1 score can be as good as 79.2% (PD). The current study represented the satisfactory performance with the ROI features from three different imaging protocols.

It is of interest to early detect the patients with Parkinson-plus syndrome because they might be prone to faster disease progression in either motor or non-motor functions when compared with idiopathic PD [23,24]. Our result showed the sensitivity of 78.2 ± 7.4% in the detection of this group of patients from those with idiopathic PD when the disease duration is less than 5 years. If early detected and diagnosed, the patients might benefit significantly from the available early intervention. With an increase in the disease duration (longer than 5 years), the diagnosis of PD is more likely to be correct (93.0 ± 5.1% of sensitivity and 84.4 ± 6.9% of F1 score). Between PSP and MSA, both the sensitivity and F1 score are comparable between short and long disease duration. Although neuroimaging cannot replace the clinical diagnosis, it may aid in providing supporting evidence for a suspected diagnosis, decreasing the overall burden on patients due to repeated visits.

### *Diagnostic performance by combined diffusion and structural-related features*

In this study, we examined the diagnostic performance of PD by combining multiple features as extracted from diffusion imaging. For the purpose of comparison, features of cortical thickness and surface area, as extracted from structural images, were used as a conventional standard. We further examined the diagnostic performance by combining features from both diffusion and structural images in an effort to improve the performance. In the diagnosis of either patients from HC or PD patients from Parkinson-plus syndromes, the difference in the F1 scores was not significant between using diffusion related features alone and combined with structural

Please cite this article as: Tsai C-C et al., Diffusion tensor imaging for the differential diagnosis of Parkinsonism by machine learning, Biomedical Journal, https://doi.org/10.1016/j.bj.2022.05.006

**Table 4 Discriminant function coefficients and Wilks' Lambda for binary differential diagnostic model.**

| Brain area for HC/PA | Coefficient | Wilks' Lambda |
|---|---|---|
| SuperiorParietal_L(AD50) | 5.85 | 0.719 |
| Cingulum, _Middle_R(MD50) | 5.292 | 0.623 |
| SupraMarginal_L(FA50) | 12.234 | 0.598 |
| SuperiorFrontal, MedialOrbital_R(MD50) | −5.134 | 0.568 |
| Precentral_R(AD50) | 9.169 | 0.545 |
| Calcarine_R(FA50) | 4.853 | 0.530 |
| SuperiorParietal_R(AD50) | 4.592 | 0.515 |
| MiddleFrontal_R(RD50) | −7.987 | 0.503 |
| Caudate_L(AD50) | 1.552 | 0.493 |
| SuperiorTemporal_R(RD50) | −1.99 | 0.479 |
| MiddleFrontal, Orbital_R(MD50) | 1.293 | 0.470 |
| Flocculus_R(AD50) | −0.94 | 0.463 |
| InferiorFrontal, Triangular_R(FA50) | 12.689 | 0.456 |
| SuperiorFrontal, Medial_L(FA50) | −11.404 | 0.450 |
| Lingual_L(AD50) | 5.475 | 0.442 |
| InferiorParietal_L(FA50) | −3.027 | 0.437 |
| InferiorFrontal, Triangular_L(AD50) | −1.284 | 0.432 |
| Thalamus_R(AD50) | −3.748 | 0.426 |
| InferiorFrontal, Triangular_R(MD50) | 1.602 | 0.423 |
| SuperiorParietal_R(AD50) | −1.332 | 0.417 |
| MiddleTemporalPole_R(AD50) | −3.574 | 0.413 |
| Insula_L(AD50) | 8.943 | 0.410 |
| Insula_L(RD50) | −6.396 | 0.404 |
| SuperiorFrontal, Orbital_R(AD50) | −10.309 | 0.400 |
| SuperiorFrontal, Orbital_R(MD50) | 8.401 | 0.395 |
| InferiorFrontal, Orbital_R(RD50) | −1.025 | 0.392 |
| RolandicOperculum_R(RD50) | 0.767 | 0.388 |
| Cingulum, Anterior_R(FA50) | −5.041 | 0.385 |
| Cingulum, Anterior_R(FA50) | 11.099 | 0.381 |
| MiddleFrontal, Orbital_R(RD50) | −2.586 | 0.378 |
| Constant | −16.664 | |
| **Brain area for PD/PM** | **Coefficient** | **Wilks' Lambda** |
| Pyramid(AD50) | −5.552 | 0.827 |
| Thalamus_L(FA50) | 7.884 | 0.749 |
| Insula_L(AD50) | −4.43 | 0.708 |
| Precuneus_L(FA50) | 6.848 | 0.665 |
| SuperiorTemporalPole_L(RD50) | 4.118 | 0.647 |
| Nodule(FA50) | 2.183 | 0.631 |
| Uvula(RD50) | 4.221 | 0.618 |
| Flocculus_R(FA50) | 1.676 | 0.599 |
| Putamen_L(RD50) | −3.501 | 0.583 |
| SuperiorFrontal, MedialOrbital_L(AD50) | −1.923 | 0.573 |
| Caudate_L(FA50) | −2.625 | 0.562 |
| Constant | 5.082 | |
| **Brain area for MSA/PSP** | **Coefficient** | **Wilks' Lambda** |
| Alae_L(MD90) | 1.243 | 0.764 |
| Quadrangular_R(AD90) | 1.403 | 0.704 |
| CentralLobule(RD90) | 2.854 | 0.663 |
| Constant | −7.067 | |

Abbreviations: MD: mean diffusivity; FA: fractional anisotropy; AD: axial diffusivity; RD: radial diffusivity; L: left hemisphere; R: right hemisphere; HC: healthy control; PA: patient; PD: idiopathic Parkinson's disease; PM: Parkinson-plus syndromes; MSA: multiple system atrophy; PSP: progressive supranuclear palsy.

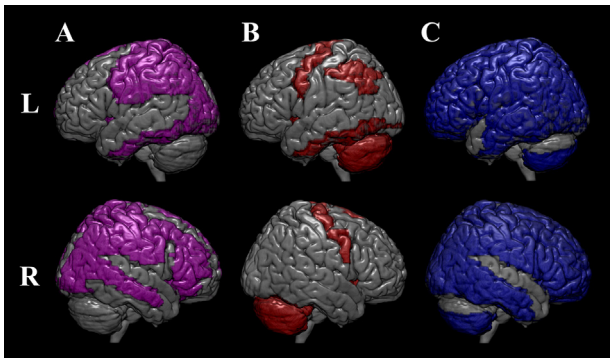


Fig. 4 **Difference in mean diffusivity between patients and healthy controls.** The figure is a three-dimensional visualization of regions exhibiting significant differences in mean diffusivity relative to healthy controls. Column A: idiopathic Parkinson's disease; Column B: multiple system atrophy; Column C: progressive supranuclear palsy.

features ($p$-value ranged from 0.075 to 0.796). However, both were significantly higher than that by structural features alone ($p$-value ranged from 0.02 to < 0.001).

Diffusion MRI was sensitive to the subtle changes in tissue microstructure in PD [30,31] even when there was no volumetric reduction or signal change on conventional MRI [32,33]. It was also noticed that the diffusion related features might improve the diagnostic performance when compared to that of structure related features. Our result showed that the performances of the combined classifier were at minimum identical to the diffusion alone. However, this observation might require further validation by using an independent blind dataset with increased number of subjects.

In the diagnosis of patients with either MSA or PSP, the reduction of the F1 scores in using combined or structural features, when compared to diffusion alone, can be significant (combined features: $p = 0.049$ for PSP and $p = 0.046$ for MSA; structural features: $p = 0.02$ for PSP and $p < 0.001$ for MSA). Because both MSA and PSP are rare diseases [34], the small cohorts in our dataset might result in possible statistical fluctuation. We cannot rule out the contribution from potential model overfit in our algorithm due to the relatively small number of patients. We believed that the future study using larger MSA and PSP cohorts (i.e. enrollment of new participants or collaboration with international database) might help to further improve the diagnostic performance from combining diffusion-related and structural features.

### *Relevance of involved regions*

Most DTI studies have focused on white matter abnormalities. However, pathologically, idiopathic PD is induced by the death of dopaminergic neurons in the gray matter (basal ganglia). Therefore, in the present study, we examined alterations in general brain parenchyma rather than limited in connecting regions. The etiology of movement disorders is often attributed to malfunctioning within neural circuits related to motor function, which likely consists of multiple brain regions

including the basal ganglia, supplementary motor area, premotor cortex, and sensorimotor cortex [35]. Additional studies have revealed that patients with PSP exhibit similar changes in the caudate and putamen [36], while those with MSA exhibit such changes in the cerebellum and pons [37]. Indeed, we observed differential involvement of brain regions based on the type of Parkinsonism [Fig. 3]. Our findings are consistent with previous studies that investigated alterations in metabolism or cortical thickness [8]. Our findings also highlight the role of the cerebellum in the differential diagnosis of Parkinsonism. The cerebellum plays an important role in motor control and may be associated with functions such as attention and language [38], which have been widely implicated in the pathophysiology of idiopathic PD [39]. These findings suggest that differential diagnosis is possible based on differential brain involvement, which can only be determined when whole-brain analyses are performed.

### *The advantage in the approach*

Our approach is advantageous because image normalization enabled the inclusion of multiple regions of interest throughout the brain. Conventionally, a representative slice is selected from the region of interest. However, pathological changes may only occur in the part of the structure and not necessarily be included in the selected slice. While voxel-based methods represent an alternative approach, the large amount of voxels contained within the images renders statistical interpretation difficult. Moreover, multivariate analysis is not straightforward because no structure/regional information can be directly derived. Although morphometric analysis, such as the magnetic resonance parkinsonism index [40], showed improved diagnostic performance in patients with PSP, it failed to separate HC and idiopathic PD. Furthermore, the process relies on manual selection of regions of interest, which is tedious and subjective, making it difficult to include many regions. The parcellation with the whole brain with machine learning algorithm further allowed identifying the specific brain structure to apply for the precision medicine of automated diagnosis in the future study.

Diffusion Tensor Imaging has been used in several studies of differential diagnosis in PD [30,31]. However, the findings can be controversial. Haller et al. reported a sensitivity between 90 and 94% in a voxel-based study using FA map and support vector machine [41]. Additional machine learning algorithm used elastic-net and receiver operating characteristic analysis showed that the sensitivity can be as good as 86% for patients with PD [42]. However, Prasuhn et al. showed that the area under curve is only 0.56 in a DTI study using support vector machine [43].

Our study hypothesized that each subtype of PD would be associated with a different extent of cortical involvement. Therefore, a comprehensive analysis of all potentially involved regions in the whole brain would be required for an accurate, imaging-based differential diagnosis. Our approach is advantageous because image normalization enabled the inclusion of multiple regions of interest throughout the brain. Machine learning algorithm then identified the involved brain regions from parcellated whole brain. Finally, our algorithm was applied for the differential diagnosis and validated by including a relatively large cohort of 625 subjects and with an independent blind validation procedure.

The fluctuation for the quantification of diffusion occurred by variations in scanners and imaging protocols [44]. Therefore, the potential interaction among the 3 acquisition protocols and 4 participant groups with different disease status was examined. The result showed that the effect from 410 in 1392 features was statistically significant, which suggested a potential interaction might exist between the acquisition protocols and the disease status from part of the features. The contribution from such interaction in a single feature is difficult to quantify. It is even more complicated to assess the effect on the final diagnostic performance as it propagated through the combination of multiple features. This effect certainly raised a new interest that would be worthy of further exploration in the future. However, our result showed diagnostic performance of 82.3 ± 3.9% (for idiopathic PD) and 67.2 ± 3.8% (for MSA) in the F1 score. The effect of such variations on diagnostic performance can be reduced because the diversity of the training data was increased. These results further suggest that quantitative DTI measurements can be applied in a more general clinical setting.

### *Limitations*

The present study possesses several limitations of note, including the lack of histological evidence, which represents the gold standard for the confirmation of parkinsonian disorders. However, all diagnoses for the patients were further supported with the reduced striatal uptake of TRODAT scans.

Secondly, we adopted a data-driven approach for our analysis, making it difficult to establish a causal relationship between the identified regions and each disease. To mitigate the bias from age and gender, both the cross- and blind validations was repeated extensively throughout the dataset for 5 times. The results showed that the diagnostic performance varied between 67.2 ± 3.8% and 87.4 ± 1.5%. Future studies may help to provide insight into the functional brain changes of the newly identified regions as associated with parkinsonism and with a group of healthy controls who are age and gender matched to the patients.

Thirdly, MSA and PSP were relative rare Parkinson-plus syndromes and difficult to enroll a large number of participants as HC and idiopathic PD, which might cause unbalanced diagnostic probability [45]. The analysis showed that no significant difference in the classification between cross- and blind validations in most procedures by diffusion-related features. This observation might suggest successful training. Significant reduction of F1 score was only noticed in two: PD in the single-step procedure (cross-validation: 76.9 ± 3.0%; blind, 70.5 ± 2.1%) and PSP in the two-step procedure (cross-validation: 51.4 ± 9.3%; blind, 42.2 ± 3.9%). This significant reduction in F1 might be attributed to potential model overfitting. Three-step procedure might reduce the difference caused by the number of participants (i.e. PD/MSA/PSP: 286/52/69; PD/Parkinson-plus syndromes: 286/121) to improve the unbalanced diagnostic probability during classification [45]. The classification results showed similar diagnostic performance between training and blind dataset.

Fourthly, features located in cerebellum might play an important role in distinguishing Parkinsonism plus syndromes

Please cite this article as: Tsai C-C et al., Diffusion tensor imaging for the differential diagnosis of Parkinsonism by machine learning, Biomedical Journal, https://doi.org/10.1016/j.bj.2022.05.006

from PD. Cerebellar abnormality is a critical exclusion criterion for diagnosing Parkinson's disease from Parkinson-plus syndromes [46]. Cerebellar atrophy was reported in patients with idiopathic PD [47], multiple system atrophy [48], and progressive supranuclear palsy [49], which might indicate structural changes in cerebellum. Unfortunately, the cerebellar regions were not included in the atlas as used by FreeSurfer [21]. It would be interesting to investigate the diagnostic performance by using an atlas containing this region in the future study, which might ultimately help to improve when using combined features.

## Conclusions

In conclusion, multivariate analysis using diffusion tensor imaging derived features can improve the performance of differential diagnosis between patients with Parkinson's disease and Parkinson-plus syndromes even when under different imaging conditions.

## Funding statement

This work was financially supported by the Ministry of Science and Technology Taiwan (grant MOST 109-2221-E-182-009-MY3, MOST 109-2314-B-182-021-MY3); the Healthy Aging Research Center (Grant EMRPD1M0451, EMRPD1M0431); and the Chang Gung Memorial Hospital (grants CMRPG2J0141, CMRPG2J0142, CMRPD1L0141).

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

The presents work was supported by the Imaging Core Laboratory of the Institute for Radiological Research and the Center for Advanced Molecular Imaging and Translation. The authors thank the Neuroscience Research Center (Chang Gung Memorial Hospital) and the Healthy Aging Research Center (Chang Gung University) for their invaluable support.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bj.2022.05.006.

## REFERENCES

[1] Lang AE, Lozano AM. Parkinson's disease. First of two parts. N Engl J Med 1998;339:1044–53.
[2] Lang AE, Lozano AM. Parkinson's disease. Second of two parts. N Engl J Med 1998;339:1130–43.
[3] Williams DR, de Silva R, Paviour DC, Pittman A, Watt HC, Kilford L, et al. Characteristics of two distinct clinical phenotypes in pathologically proven progressive supranuclear palsy: Richardson's syndrome and PSP-parkinsonism. Brain 2005;128:1247–58.
[4] Weingarten CP, Sundman MH, Hickey P, Chen NK. Neuroimaging of Parkinson's disease: expanding views. Neurosci Biobehav Rev 2015;59:16–52.
[5] Focke NK, Helms G, Scheewe S, Pantel PM, Bachmann CG, Dechent P, et al. Individual voxel-based subtype prediction can differentiate progressive supranuclear palsy from idiopathic Parkinson syndrome and healthy controls. Hum Brain Mapp 2011;32:1905–15.
[6] Karagulle Kendi AT, Lehericy S, Luciana M, Ugurbil K, Tuite P. Altered diffusion in the frontal lobe in Parkinson disease. AJNR Am J Neuroradiol 2008;29:501–5.
[7] Zhang K, Yu C, Zhang Y, Wu X, Zhu C, Chan P, et al. Voxel-based analysis of diffusion tensor indices in the brain in patients with Parkinson's disease. Eur J Radiol 2011;77:269–73.
[8] Cykowski MD, Coon EA, Powell SZ, Jenkins SM, Benarroch EE, Low PA, et al. Expanding the spectrum of neuronal pathology in multiple system atrophy. Brain 2015;138:2293–309.
[9] Sitburana O, Ondo WG. Brain magnetic resonance imaging (MRI) in Parkinsonian disorders. Parkinsonism Relat Disord 2009;15:165–74.
[10] Prodoehl J, Li H, Planetta PJ, Goetz CG, Shannon KM, Tangonan R, et al. Diffusion tensor imaging of Parkinson's disease, atypical Parkinsonism, and essential tremor. Mov Disord 2013;28:1816–22.
[11] Planetta PJ, Ofori E, Pasternak O, Burciu RG, Shukla P, DeSimone JC, et al. Free-water imaging in Parkinson's disease and atypical parkinsonism. Brain 2016;139:495–508.
[12] Höglinger GU, Respondek G, Stamelou M, Kurz C, Josephs KA, Lang AE, et al. Clinical diagnosis of progressive supranuclear palsy: the movement disorder society criteria. Mov Disord 2017;32:853–64.
[13] Gilman S, Wenning GK, Low PA, Brooks DJ, Mathias CJ, Trojanowski JQ, et al. Second consensus statement on the diagnosis of multiple system atrophy. Neurology 2008;71:670–6.
[14] Martínez-Martín P, Gil-Nagel A, Gracia LM, Gómez JB, Martinez-Sarries J, Bermejo F. Unified Parkinson's disease rating scale characteristics and structure. Mov Disord 1994;9:76–83.
[15] Goetz CG, Poewe W, Rascol O, Sampaio C, Stebbins GT, Counsell C, et al. Movement disorder society task force report on the Hoehn and Yahr staging scale: status and recommendations. Mov Disord 2004;19:1020–8.
[16] Tsai CC, Lin YC, Ng SH, Chen YL, Cheng JS, Lu CS, et al. A method for the prediction of clinical outcome using diffusion magnetic resonance imaging: application on Parkinson's disease. J Clin Med 2020;9:647.
[17] Tabesh A, Jensen JH, Ardekani BA, Helpern JA. Estimation of tensors and tensor-derived measures in diffusional kurtosis imaging. Magn Reson Med 2011;65:823–36.
[18] Mazziotta JC, Toga AW, Evans AC, Fox PT, Lancaster JL. Digital brain atlases. Trends Neurosci 1995;18:210–1.
[19] Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, et al. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. Neuroimage 2002;15:273–89.
[20] Fischl B, Salat DH, Busa E, Albert M, Dieterich M, Haselgrove C, et al. Whole brain segmentation: automated

labeling of neuroanatomical structures in the human brain. Neuron 2002;33:341–55.
[21] Desikan RS, Segonne F, Fischl B, Quinn BT, Dickerson BC, Blacker D, et al. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. Neuroimage 2006;31:968–80.
[22] Kim HY. Statistical notes for clinical researchers: two-way analysis of variance (ANOVA)-exploring possible interaction between factors. Restor Dent Endod 2014;39:143–7.
[23] McFarland NR, Hess CW. Recognizing atypical parkinsonisms: "red flags" and therapeutic approaches. Semin Neurol 2017;37:215–27.
[24] Postuma RB, Berg D, Stern M, Poewe W, Olanow CW, Oertel W, et al. MDS clinical diagnostic criteria for Parkinson's disease. Mov Disord 2015;30:1591–601.
[25] Hua J, Xiong Z, Lowey J, Suh E, Dougherty ER. Optimal number of features as a function of sample size for various classification rules. Bioinformatics 2005;21:1509–15.
[26] Austin PC, Steyerberg EW. The number of subjects per variable required in linear regression analyses. J Clin Epidemiol 2015;68:627–36.
[27] Vittinghoff E, McCulloch CE. Relaxing the rule of ten events per variable in logistic and Cox regression. Am J Epidemiol 2007;165:710–8.
[28] Akoglu H. User's guide to correlation coefficients. Turk J Emerg Med 2018;18:91–3.
[29] Poulsen J, French A. Discriminant function analysis, http://userwww.sfsu.edu/~efc/classes/biol710/discrim/discrim.pdf/; 2008 [Retrieved 4 March 2008].
[30] Lu CS, Ng SH, Weng YH, Cheng JS, Lin WY, Wai YY, et al. Alterations of diffusion tensor MRI parameters in the brains of patients with Parkinson's disease compared with normal brains: possible diagnostic use. Eur Radiol 2016;26:3978–88.
[31] Wang JJ, Lin WY, Lu CS, Weng YH, Ng SH, Wang CH, et al. Parkinson disease: diagnostic utility of diffusion kurtosis imaging. Radiology 2011;261:210–7.
[32] Seppi K, Schocke MF, Prennschuetz-Schuetzenau K, Mair KJ, Esterhammer R, Kremser C, et al. Topography of putaminal degeneration in multiple system atrophy: a diffusion magnetic resonance study. Mov Disord 2006;21:847–52.
[33] Barbagallo G, Sierra-Pena M, Nemmi F, Traon AP, Meissner WG, Rascol O, et al. Multimodal MRI assessment of nigro-striatal pathway in multiple system atrophy and Parkinson disease. Mov Disord 2016;31:325–34.
[34] Schrag A, Ben-Shlomo Y, Quinn NP. Prevalence of progressive supranuclear palsy and multiple system atrophy: a cross-sectional study. Lancet 1999;354:1771–5.
[35] DeLong MR, Wichmann T. Circuits and circuit disorders of the basal ganglia. Arch Neurol 2007;64:20–4.
[36] Wang J, Wai Y, Lin WY, Ng S, Wang CH, Hsieh R, et al. Microstructural changes in patients with progressive supranuclear palsy: a diffusion tensor imaging study. J Magn Reson Imaging 2010;32:69–75.
[37] Ito M, Watanabe H, Kawai Y, Atsuta N, Tanaka F, Naganawa S, et al. Usefulness of combined fractional anisotropy and apparent diffusion coefficient values for detection of involvement in multiple system atrophy. J Neurol Neurosurg Psychiatry 2007;78:722–8.
[38] Wolf U, Rapoport MJ, Schweizer TA. Evaluating the affective component of the cerebellar cognitive affective syndrome. J Neuropsychiatry Clin Neurosci 2009;21:245–53.
[39] Wu T, Hallett M. The cerebellum in Parkinson's disease. Brain 2013;136:696–709.
[40] Quattrone A, Nicoletti G, Messina D, Fera F, Condino F, Pugliese P, et al. 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.
[41] Haller S, Badoud S, Nguyen D, Garibotto V, Lovblad KO, Burkhard PR. Individual detection of patients with Parkinson disease using support vector machine analysis of diffusion tensor imaging data: initial results. AJNR Am J Neuroradiol 2012;33:2123–8.
[42] Du G, Lewis MM, Kanekar S, Sterling NW, He L, Kong L, et al. Combined diffusion tensor imaging and apparent transverse relaxation rate differentiate Parkinson disease and atypical Parkinsonism. AJNR Am J Neuroradiol 2017;38:966–72.
[43] Prasuhn J, Heldmann M, Munte TF, Bruggemann N. A machine learning-based classification approach on Parkinson's disease diffusion tensor imaging datasets. Neurol Res Pract 2020;2:46.
[44] Chen YL, Lin YJ, Lin SH, Tsai CC, Lin YC, Cheng JS, et al. The effect of spatial resolution on the reproducibility of diffusion imaging when controlled signal to noise ratio. Biomed J 2019;42:268–76.
[45] Dal Pozzolo A, Caelen O, Johnson RA, Bontempi G. Calibrating probability with undersampling for unbalanced classification. In: IEEE symposium series on computational intelligence. IEEE; 2015. p. 159–66.
[46] Marsili L, Rizzo G, Colosimo C. Diagnostic criteria for Parkinson's disease: from James Parkinson to the concept of prodromal disease. Front Neurol 2018;9:156.
[47] O'Callaghan C, Hornberger M, Balsters JH, Halliday GM, Lewis SJ, Shine JM. Cerebellar atrophy in Parkinson's disease and its implication for network connectivity. Brain 2016;139:845–55.
[48] Dickson DW. Parkinson's disease and parkinsonism: neuropathology. Cold Spring Harb Perspect Med 2012;2:a009258.
[49] Koga S, Josephs KA, Ogaki K, Labbe C, Uitti RJ, Graff-Radford N, et al. Cerebellar ataxia in progressive supranuclear palsy: an autopsy study of PSP-C. Mov Disord 2016;31:653–62.

Please cite this article as: Tsai C-C et al., Diffusion tensor imaging for the differential diagnosis of Parkinsonism by machine learning, Biomedical Journal, https://doi.org/10.1016/j.bj.2022.05.006